.

Automatic recognition of fin and blue whale calls for real-time monitoring in the St. Lawrence

Xavier Mouy and WWAMohammed Bahoura and WWAYvan Simard and WWA

Citation: The Journal of the Acoustical Society of America **126**, 2918 (2009); doi: 10.1121/1.3257588
View online: http://dx.doi.org/10.1121/1.3257588
View Table of Contents: http://asa.scitation.org/toc/jas/126/6
Published by the Acoustical Society of America

Automatic recognition of fin and blue whale calls for real-time monitoring in the St. Lawrence

Xavier Mouy[a)]
Marine Sciences Institute, University of Québec at Rimouski, 310 Allée des Ursulines, Rimouski, Québec G5L-3A1, Canada

Mohammed Bahoura
Department of Mathematics, Informatics and Engineering, University of Québec at Rimouski, 300 Allée des Ursulines, Rimouski, Québec G5L-3A1, Canada

Yvan Simard
*Marine Sciences Institute, University of Québec at Rimouski, 310 Allée des Ursulines, Rimouski, Québec G5L-3A1, Canada and Maurice Lamontagne Institute, Fisheries and Oceans Canada, 850 Route de la Mer, Mont-Joli, Québec G5H-3Z4, Canada

(Received 20 April 2009; revised 30 September 2009; accepted 6 October 2009)

Monitoring blue and fin whales summering in the St. Lawrence Estuary with passive acoustics requires call recognition algorithms that can cope with the heavy shipping noise of the St. Lawrence Seaway and with multipath propagation characteristics that generate overlapping copies of the calls. In this paper, the performance of three time-frequency methods aiming at such automatic detection and classification is tested on more than 2000 calls and compared at several levels of signal-to-noise ratio using typical recordings collected in this area. For all methods, image processing techniques are used to reduce the noise in the spectrogram. The first approach consists in matching the spectrogram with binary time-frequency templates of the calls (coincidence of spectrograms). The second approach is based on the extraction of the frequency contours of the calls and their classification using dynamic time warping (DTW) and the vector quantization (VQ) algorithms. The coincidence of spectrograms was the fastest method and performed better for blue whale A and B calls. VQ detected more 20 Hz fin whale calls but with a higher false alarm rate. DTW and VQ outperformed for the more variable blue whale D calls.
 [DOI: 10.1121/1.3257588]

PACS number(s): 43.30.Sf, 43.60.Bf [WWA] Pages: 2918–2928

I. Introduction

The Atlantic blue whale (*Balaenoptera musculus*) and fin whale (*Blaenoptera physalus*) are listed as endangered and of special concern, respectively, according to the Canadian species at risk act (Canada, 2008). The Northwest Atlantic population sizes are estimated in the low hundreds for the blue whale and between 3500 and 6300 for the fin whale (Perry *et al.*, 1999; Sears and Calambokidis, 2002). Among the threats for the survival of these species are collisions with ships (Laist *et al.*, 2001; Jensen and Silber, 2004) and the increasing low-frequency anthropogenic noise in the oceans (Andrew *et al.*, 2002; McDonald *et al.*, 2006), which is possibly affecting mysticetes, known to produce and probably perceive low-frequency sounds (Richardson *et al.*, 1995; Croll *et al.*, 2001; NRC, 2003; Simard *et al.*, 2008; Au and Hastings, 2008).

The St. Lawrence Estuary is an intensively used habitat of the North Atlantic fin and blue whale populations. This traditional feeding ground (Simard and Lavoie, 1999) is an important whale watching area (Hoyt, 2001). It is also part of an important North American seaway connecting the Great Lakes and the Atlantic. The impact that the anthropogenic activities might have on the survival of fin and blue whales summering in this area is not well known. Basic information on the long-term spatial and temporal distribution of the whales in the region is lacking but is a matter of concern. The development of efficient methods to continuously monitor the whales over the whole basin would be a contribution to fill this gap.

The emergence of passive acoustics monitoring (PAM) technology offers the possibility of addressing this challenging time-space distribution of whales over their extensive life domains in the oceans (Mellinger *et al.*, 2007). The manual detection of the vocalizations on recordings performed aurally and by manual inspection of spectrograms is a long and laborious task leading to an inconstant bias in the analysis dependent on the experience and the degree of fatigue of the operator. Therefore, the development of efficient and robust automatic detection and classification methods is required. Their performance is dependent on the complexity and diversity of the sounds to recognize and the acoustic properties of the environment such as the characteristics of the background noise and sound propagation in the area. Matched filters appear to be efficient for detecting stereotypical signals in low Gaussian noise conditions (Stafford *et al.*, 1998; Mellinger and Clark, 2000). The correlation of spectrograms

[a)]Author to whom correspondence should be addressed. Electronic mail: xavier.mouy@gmail.com

with a time-frequency kernel representing the call has been used on right whale (Mellinger, 2004; Munger *et al.*, 2005), and North Pacific blue whale calls (Mellinger and Clark, 2000; Wiggins *et al.*, 2005). These methods are well adapted for calls presenting low variability in duration and frequency. Chirplet transform shows an interesting potential for classifying blue whale calls (Bahoura and Simard, 2008). Artificial neural networks (ANNs) and hidden Markov modeling (HMM) show good recognition results on elephants (Clemins *et al.*, 2005), birds (Kogan and Margoliash, 1998), and whale calls (Potter *et al.*, 1994; Mellinger and Clark, 2000; Mellinger, 2004). These two methods have been tested using various attributes for characterizing the calls: Fourier coefficients (Mellinger, 2004), Mel-frequency cepstral coefficients (Clemins *et al.*, 2005), and linear predictive coefficients (Clemins and Johnson, 2006). ANN and HMM can accommodate variability within the calls. The time-frequency contour approach extracts the call shape attributes from the spectrogram (e.g., maximum frequency, minimum frequency, duration, etc.) to feed the classification algorithms. The contours can be extracted using an edge detector (Gillespie, 2004) or through instantaneous frequency tracking algorithms (Sturtivant and Datta, 1995a, 1995b; Brown *et al.*, 2006; Halkias and Ellis, 2006, 2008). Automatic contour extraction by tracking algorithms can be computationally expensive (Brown and Zhang, 1991). Classification methods used so far for this approach include discriminant analysis (Bazúa-Durán and Au, 2004; Gillespie, 2004) and dynamic time warping (DTW) (Buck and Tyack, 1993; Brown *et al.*, 2006). Urazghildiiev and Clark (2006, 2007) proposed another approach involving a generalized likelihood ratio test (GLRT) detector and a finite impulse response (FIR) filter operating on spectrograms, which outperformed the human operator for right whale calls.

The blue whale vocal repertoire in the St. Lawrence is similar to the one found in the North Atlantic. It is composed of three basic calls, the A, B, and D calls. (Edds, 1982; Mellinger and Clark, 2003; Berchok *et al.*, 2006). The A call is an 8-s tonal sound at 18 Hz. The B call is an 11-s down sweep from 18 to 15 Hz [Fig. 1(a)]. Although these two calls usually follow each other in AB phrases repeated every 74 s, it is possible to observe repetitions of A calls or B calls only (Mellinger and Clark, 2003; Berchok *et al.* 2006). The two calls are then separated by an ~5-s gap, but this short silence is sometimes absent. Berchok *et al.* (2006) called these calls the hybrid calls. These latter calls are hereafter referred to as *AB calls*. The D call is less stereotyped than the two previous ones and consists of a down sweep in the ~120–40 Hz frequency range. It sometimes starts with an upsweep forming an arch in the spectrogram. The duration, and the starting and ending frequencies are variable [Fig. 1(c)]. The D calls are not repeated in regular sequences as the A and B calls; however, some time patterns can be observed. As noticed by Berchok *et al.* (2006), the blue whale calls in the St. Lawrence are more variable than the ones reported in the North Atlantic. A lot of them are not continuous but segmented into several fragments. Berchok *et al.* (2006) believed that this segmentation is not due to propagation effects but rather a property of the whale call. The other rare blue whale calls such as the 9 Hz sounds reported by Mellinger and Clark (2003), as well as the *blurps* and *grunts* reported by Berchok *et al.* (2006), were not considered here because of their low interest for monitoring applications.

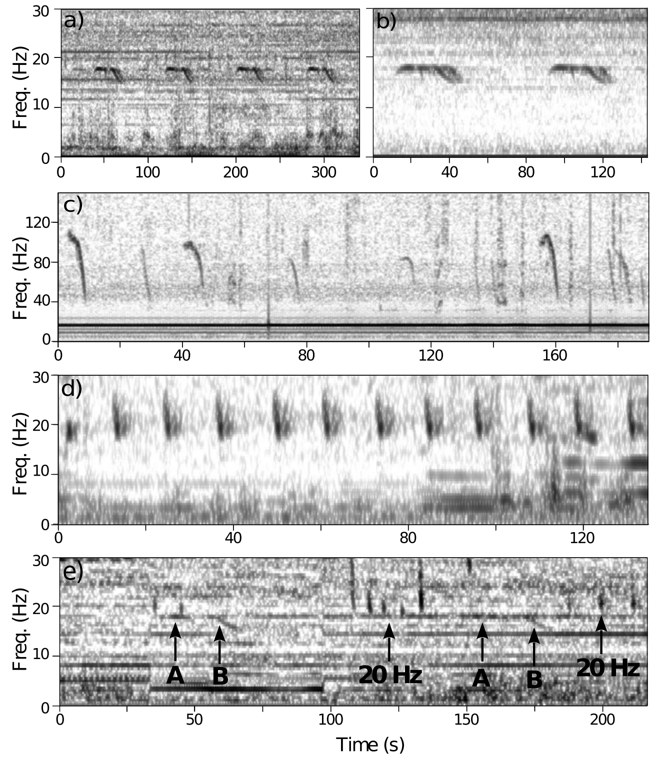

FIG. 1. Blue and fin whales calls. Spectrograms representing (a) 4 blue whale AB phrases, (b) 2 blue whale AB phrases with strong multipath echoes, (c) 11 very variable blue whale D calls, and (d) 12 fin whale 20 Hz calls with multipath echoes. (e) Example of a noisy recording containing blue and fin whale calls.

The 20 Hz pulses are the most frequent fin whale calls recorded in the St. Lawrence (Samaran, 2004). They are similar to the ones reported in the other areas of the world (Schevill *et al.*, 1964; Thompson and Friedl, 1982; Edds, 1988; Thompson *et al.*, 1992), except for the occasional lagging lower frequency backbeat in the call bouts (Samaran, 2004; Simard and Roy, 2008). This 20 Hz call consists of a 1-s downsweep from 23 to 18 Hz [Fig. 1(d)]. The 140 Hz call reported by Edds (1988) and Samaran (2004) and the backbeat are less frequent and are therefore not considered in this paper.

Although several methods were already tested to automatically detect and classify fin and blue whale calls, none were thoroughly tested with the specific calls of the St. Lawrence Estuary and its particular acoustic conditions where intense low-frequency shipping noise in the vocalization band causes most recorded calls to have low signal-to-noise ratios (SNRs) [e.g., Fig. 1(e); Simard *et al.*, 2006a, 2008]. The presence of a sound channel in the upper half of the ~350-m water column, resulting from the summer cold intermediate layer and steep walls at the head of the deep Laurentian Channel, favors multipath propagation and the presence of echoes [e.g., Fig. 1(b); Simard and Roy, 2008]. Consequently, the calls often appear distorted and warped in time, which hinders the detection and classification algorithms.

TABLE I. Time-space coordinates of the acoustic recordings.

Station	Recording period	Latitude (°N)	Longitude (°W)	Hydrophone mean depth (m)	Bottom depth (m)
A	Aug.–Sept. 2003	48.2683	69.4663	125	130
B	Aug.–Sept. 2003	48.2687	69.4636	188	193
C	Sept.–Oct. 2003	48.3544	69.1167	54	88
D	Sept.–Oct. 2003	48.2504	69.2225	53	55
E	Sept.–Oct. 2003	48.1900	69.5925	43	60
F	Aug.–Oct. 2004	48.3783	69.3300	155	239
G	Aug.–Oct. 2004	48.1233	69.5433	126	135

This paper explores the capability of signal processing methods that could be implemented in real-time to detect fin and blue whale calls in an intense shipping noise background, and to cope with the call duration and frequency variability observed in the St. Lawrence. Two approaches based on time-frequency representations are tested. The classical spectrogram correlation method, previously used for blue and fin whale call detection, is adapted to the St. Lawrence acoustic conditions. The second approach, consisting in the extraction and classification of the call's time-frequency contours, tries to improve some weaknesses of this adapted classical approach. This work is part of the real-time monitoring project "Whale On the Web" (Simard *et al.*, 2006b).

II. Materials and Methods

This section describes the acoustic data used for this work and the various steps involved in the two classification approaches that were developed. The steps for the first approach are (a) the calculation of the spectrogram, (b) the noise removal, and (c) the coincidence of spectrograms. The steps for the second approach are (a) the calculation of the spectrogram, (b) the noise removal, (c) the extraction of the call contours, and (d) their classification. Two contour classification methods are investigated, dynamic time warping and vector quantization (VQ).

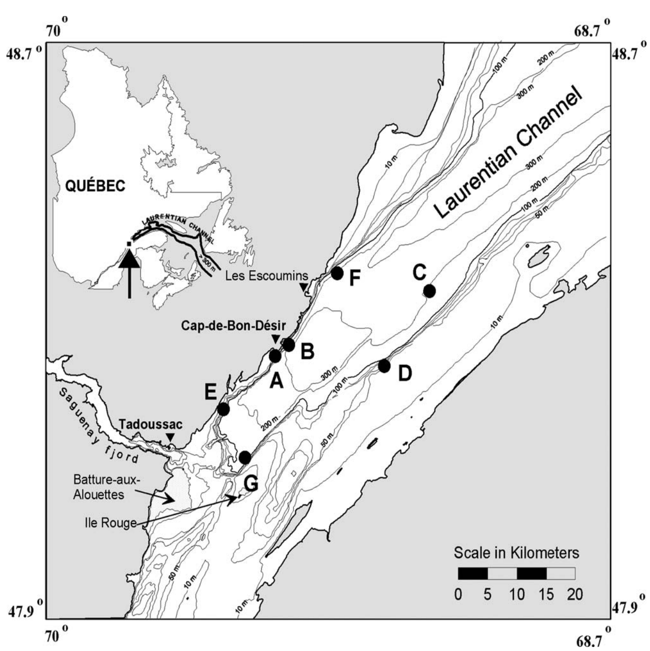

FIG. 2. Map of the study area with bathymetry and locations of the acoustic recording devices.

A. Data collection and data sets

All the data were collected in summer at seven locations in the lower St. Lawrence Estuary (Table I). Locations A and B are part of a hydrophone costal array deployed on the seabed at Cap-de-Bon-Desir in 2003 (Fig. 2; Simard and Roy, 2008). Locations C, D, E, F, and G had aural autonomous hydrophones (Multi-Electronique Inc., Rimouski, QC, Canada) moored around the sound channel depth in 2003 and 2004. The recordings were digitized at 16 bits using a sampling rate of 20 kHz for locations A and B and 2 kHz for the other locations. All recordings were decimated to 2000, 400, or 200 Hz for blue and fin whale call processing. Different sampling rates were used to get optimal time-frequency resolutions for each call type.

Among these recordings, 2472 calls were manually identified, labeled, and time stamped by an experienced observer (XM) through visual inspection of the spectrogram and aural validation when needed. A total of 2063 calls were used to test the performance of the algorithms while the other 409 calls were used for training the algorithms and adjusting their parameters (Table II). These calls were selected to represent the different noise and multipath echo conditions encountered in this part of the St. Lawrence. Calls were carefully chosen to ensure that data for testing and training were recorded at different locations and different times. To test the false alarm rate of the algorithms, 14 h of recordings containing different kinds of typical local noises but no calls were selected. This noise data set was made up of 90 wave files lasting 5–10 min each, collected at different locations of the estuary and containing intense low-frequency shipping noise. Files were usually selected at the closest point of approach of the ship (i.e., center of Lloyd's mirror "V-pattern," Urick, 1983). These recordings corre-

TABLE II. Number of calls of the training and test data sets by call type.

	A call	B call	D call	20 Hz call	Total
Training	117	91	97	104	409
Test	568	490	510	495	2063

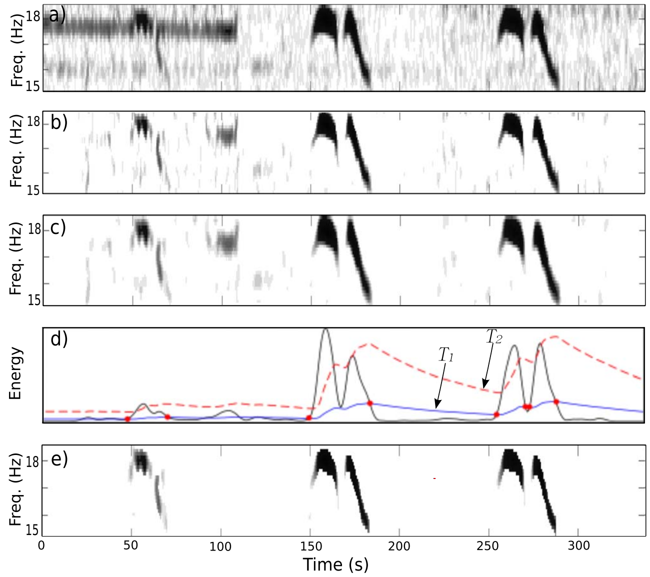

FIG. 3. (Color online) Sequence of steps of the noise reduction process. (a) The original spectrogram representing three blue whale AB phrases, time resolution=0.64 s, frequency resolution=0.13 Hz; (b) the spectrogram after equalization; and (c) Gaussian smoothing. (d) Representation of the energy e (bold line) and the adaptive thresholds T_1 (thin line) and T_2 (dashed line). The dots represent the intersections of e with T_1. (e) Final spectrogram after applying the thresholds.

spond to the most problematic conditions for the detection and classification algorithms.

TABLE III. Values of the parameters used in the first approach (i.e., coincidence of spectrograms).

Parameter	A call	B call	D call	20 Hz call
Sample frequency (Hz)	200	200	200	200
Spectrogram				
Frame size (points)	512	512	128	64
Overlap (%)	87.5	87.5	75	87.5
FFT size (points)	4608	4608	128	192
Frequency resolution (Hz)	0.04	0.04	1.56	1.04
Time resolution (s)	0.32	0.32	0.16	0.04
Noise reduction				
Δt (s)	55	15	55	15
K_x (bins)	9	6	6	5
K_y (bins)	3	3	3	3
δ_1	−0.06	−0.06	−0.1	−0.06
δ_2	0.9	0.6	0.6	1.2
δ_3	0.92	0.85	0.92	0.92
λ	0.99	0.986	0.97	0.994
Coincidence of spectrograms				
f_1 (Hz)	18.1	17.5	80	26
f_2 (Hz)	17.8	15.8	40	20
D_{call} (s)	7	9	3	1
Δf (Hz)	0.3	0.3	4	1
D_{blank} (s)	2	0.7	1	1
T_{CS} (%)	77	75	67	74

B. Noise reduction

The time-frequency representation used short-time Fourier transform for both approaches [Fig. 3(a)]. The three noise reduction steps that were applied to both approaches consist of spectrogram equalization, Gaussian kernel smoothing, and adaptive energy thresholding. The spectrogram resolutions differed for each call type and method. First, the equalization consisted of subtracting the time-averaged spectrogram from the original spectrogram. The time-averaged spectrogram was calculated by applying a moving average of the energy values for each frequency band of the spectrogram matrix (Mellinger, 2004). Length of the moving average window, Δt, is reported in Tables III and IV. This reduced the long constant spectral rays produced by the fundamental rotational frequencies of the ship propulsion machinery (Arveson and Vendittis, 2000) and by mooring vibration [Fig. 3(b)]. Second, the spectrogram was smoothed using a Gaussian kernel (Gillespie, 2004). The spectrogram matrix was convolved with a two-dimensional Gaussian mask [Fig. 3(c)]. The size of the kernel, $K_x \times K_y$ (time $\times$ frequency), varied depending on call type and classification method (Tables III and IV). Non-square kernels smooth the spectrogram in a particular direction, improving the continuity of truncated calls along the time or frequency axis. Third, an adaptive energy threshold was applied to discard the parts of the spectrogram with locally non-significant energy. The energy at each time step n of the spectrogram was smoothed using a moving average to generate a smoothed energy curve $e(n)$ [bold line in Fig. 3(d)]. The moving average was performed with a 5-s window for the A and B calls and a 1-s window for the D and 20 Hz calls. Then, two adaptive thresholds T_1 and T_2 are defined, as in Renevey and Drygajlo, 2001.

$$T_1[n] = E[n] + \delta_1 E[n] \tag{1}$$

and

$$T_2[n] = E[n] + \delta_2 E[n], \tag{2}$$

where δ_1 and δ_2 are constants in the range [0,1], and where $E[n]$ is defined by

$$E[n] = \lambda E[n-1] + (1-\lambda)e[n]. \tag{3}$$

In Eq. (3), λ is the adaptive coefficient. The adaptive threshold T_2 allows detection of events while the crossing points of the two series T_1 and $e(n)$ are used to define the start and the end of these events [dots in Fig. 3(d)]. The spectrogram values outside the call windows are blanked. Finally, the pixels exceeding the adaptive threshold T_3 are retained at each time step

$$T_3[n] = \mu[n] + \delta_3 \sigma[n], \tag{4}$$

where μ and σ are the mean and the standard deviation of the spectrogram energy at the step n and δ_3 is a constant in the range [0,1]. The final result of this noise reduction process is shown in Fig. 3(e). Values of parameters δ_1, δ_2, δ_3, and λ were chosen empirically (Tables III and IV).

C. First approach: Coincidence of spectrograms

The coincidence of spectrograms method consists of finding a specific type of sound represented by a time-

TABLE IV. Values of the parameters used in the second approach (i.e., extraction and classification of call contours).

Parameter	A and B calls	D call	20 Hz call
Sample frequency (Hz)	2000	400	400
	Spectrogram		
Frame size (points)	4096	64	128
Overlap (%)	93	75	87.5
FFT size (points)	12 288	192	384
Frequency resolution (Hz)	0.16	2.08	1.04
Time resolution (s)	0.14	0.04	0.04
Minimum frequency (Hz)	15	30	18
Maximum frequency (Hz)	18.5	115	30
	Noise reduction		
Δt (s)	30	15	35
K_x (bins)	8	11	11
K_y (bins)	4	3	3
δ_1	−0.16	−0.16	−0.16
δ_2	0.7	0.5	0.9
δ_3	0.25	0.6	0.7
λ	0.996	0.992	0.993
	Contour extraction		
T_c	0.2	0.3	0.7
T_{con} (s)	6.5	2.5	0.5
$T_{\min}$ (s)	3.5	1	0.5
$T_{\max}$ (s)	35	9	3
	Dynamic time warping classification		
T_{DTW}	0.2	5	1.3
	Vector quantization classification		
T_{VQ}	1	18	4

frequency kernel (Mellinger and Clark, 1997, 2000). Various operations could be used to match the kernel with the spectrogram. Here the logical AND operation is used because of the very low computational cost of this operation (Tiemann *et al.*, 2004). The tested spectrogram is first binarized by setting the high-energy parts that exceed a threshold to 1 and the rest to 0. Then, the time-frequency kernel representing a candidate call template is defined by a synthetic binary image generated from the call starting frequency f_1, ending frequency f_2, duration D_{call}, frequency span Δf, and silence duration before and after the call D_{blank} (Table III). Finally the matching of the call kernel with the spectrogram is the arithmetic sum of the bits resulting from a logical AND between the binary kernel and spectrogram at each time step. A perfect match corresponds to a coincidence rate of 100%. The call is detected when the coincidence rate exceeds the empirically chosen threshold value T_{CS} (Table III).

D. Second approach: Extraction and classification of call contours

This second approach proceeds by extracting the call contours followed by their classification. Two classification algorithms are tested: DTW and VQ.

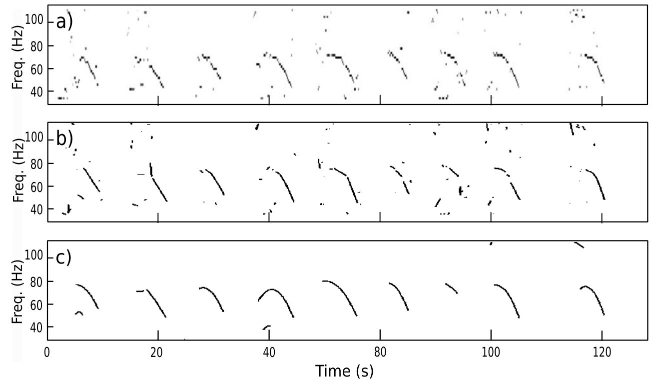

FIG. 4. Example of frequency contour extraction for a series of nine blue whale D calls. (a) Extraction of the local energy maxima, (b) connection of the maxima by time contiguity and frequency proximity, and (c) contours retained after connection of the call segments and duration restriction. The retained contours are approximated by a second-order polynomial.

1. Extraction of the contours

Extraction of time-frequency contours has been used with musical signals (Brown and Zhang, 1991; Brown, 1992), speech (Rabiner *et al.*, 1976; Rabiner, 1977), and odontocete calls (Sturtivant and Datta, 1995a, 1995b; Datta and Sturtivant, 2002; Brown *et al.*, 2006; Halkias and Ellis, 2006). Because real-time implementation is one constraint of the present work, the algorithm has to be optimized for both accurate contour extraction and short computation time. Drawing from Halkias and Ellis, 2006, the algorithm is divided into three steps: identification of the local energy maxima on the spectrogram, the delineation of call segments, and their conditional connection. First, the local energy maxima for each time step of the spectrogram are defined as the spectrum peaks [Fig. 4(a)]. Second, the maxima separated by less than four frequency bins in frequency at adjacent time steps are connected together to form segments [Fig. 4(b)]. To remove the irregularities of the extracted contour segments, their frequency time-series were smoothed with a moving average using a sliding window length of 3 s for the A and B calls, 1 s for the D call, and 0.1 s for the 20 Hz call. Third, the last step consists of connecting some of the isolated segments into more complete call contours. This connection is based on a stochastic model. The probability of connection is evaluated for each close segment pair. Each x_i pair is defined by its starting and ending slopes, α_{i1} and α_{i2}. Two frequency continuity distances β_{i2} and β_{i1} are measured for each possible linear connection; β_{i2} corresponds to the gap in frequency when the end of the first segment is linearly extended to the beginning of the second segment, and β_{i1} corresponds to the gap in frequency measured when the beginning of the second segment is linearly extended to the end of the first segment. The distance β_i is the minimum of these two distances. Two series of observations L_α and L_β representing the parameters α and β calculated on N pairs of connecting segments are extracted from the training database

$$L_\alpha = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{i1} & \alpha_{i2} \\ \vdots & \vdots \\ \alpha_{N1} & \alpha_{N2} \end{bmatrix} \quad \text{and} \quad L_\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_i \\ \vdots \\ \beta_N \end{bmatrix}.$$

Two normal distributions $\Theta_\alpha(\mu_\alpha, \Sigma_\alpha)$ and $\Theta_\beta(\mu_\beta, \sigma_\beta)$ are defined to represent the observations L_α and L_β, where Σ_α, μ_α, σ_β, and μ_β are, respectively, the covariance matrix and mean vector of L_α and the standard deviation and mean of L_β. The connection of an unknown call segment pair x is evaluated by the likelihood $P(x)$ defined by

$$P(x) = P(x|\Theta_\alpha)P(x|\Theta_\beta), \tag{5}$$

with $P(x|\Theta_\alpha)$ and $P(x|\Theta_\beta)$ being, respectively, the likelihood that the connection x can be generated by the models Θ_α and Θ_β. For convenience, the maximum of each of the two distributions is individually normalized to 1. Two segments are connected only if the likelihood $P(x)$ exceeds a threshold value T_c empirically adjusted by the operator (Table IV). Among all resulting valid possibilities, the connection proceeds with the segment that maximizes the likelihood. Once a connection is accepted, the two segments are connected by interpolation using an order 2 polynomial [Fig. 4(c)]. Connection is considered only for segments separated by less than the duration T_{con}. Contours shorter than the duration $T_{\min}$ or longer than the duration $T_{\max}$ are not considered for classification (Table IV).

2. Classification using DTW

The DTW algorithm is used by the speech processing community for isolated word recognition (Rabiner and Juang, 1993). It was applied in bioacoustics for the classification of dolphin sounds (Buck and Tyack, 1993), orca sounds (Brown *et al.*, 2006), and bird sounds (Ito *et al.*, 1996; Anderson *et al.*, 1996). The algorithm consists of classifying an unknown sound T by comparing it to a set R of k reference sounds from a dictionary, obtained from manually identified calls of the training data set. The T and R_k calls are represented by a feature vector composed of the instantaneous frequency f, the speed f', and the acceleration f'' of the call contour at each time step (Fig. 5). The feature vectors were not normalized.

$$T:[T[i]; i = 1,2, \ldots, I],$$

$$R_k:[R_k[j]; j = 1,2, \ldots, J_k],$$

where I and J_k represent the number of time frames of the T and R_k calls, respectively. The comparison of T with each template R_k is performed by calculating the dissimilarity (or distortion) between them based on their feature vectors. During this step, the algorithm has the particularity to take into consideration the possible compressions or extensions in time of the calls. A matrix X_k of dimension $I \times J_k$, is fitted with the Euclidean distances, $d(T[i], R_k[j])$ separating all the vectors $T[i]$ from $R_k[j]$ (Fig. 5). For the DTW algorithm, calculating the similarity between T and R_k consists of finding the shortest monotonic path w, of length C, within the

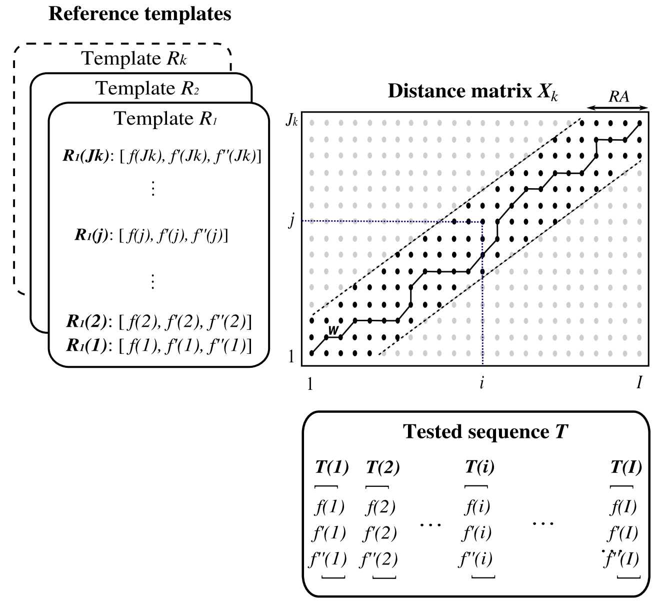

FIG. 5. (Color online) Scheme of the DTW comparison process for an unknown call T and call reference templates R_k. For each template R_k, a matrix X_k representing the Euclidian distances between the feature vectors of T and R_k is calculated. The dissimilarity between the two compared calls is represented by the path w going from the point (1,1) to (I, J_k) of the matrix X_k that minimizes the accumulated distance. The path w is found by calculating recursively the minimum cumulative distance for each point using local constraints. To avoid exaggerated distortions, the path search area is restricted by the two lines parallel to the diagonal of X_k.

matrix X_k, progressing from the point (1,1) to the point (I, J_k),

$$w:[i(c), j(c)]; c = 1,2, \ldots, C,$$

with $i(1)=1$, $j(1)=1$, $i(C)=I$, and $j(C)=J_k$, such that its associated distance, $D_w(T, R_k)$, defined by

$$D_w(T, R_k) = \frac{\sum_{c=1}^{C} d\{T[i(c)], R_k[j(c)]\} g(c)}{N(g)}, \tag{6}$$

is the minimum. In Eq. (6), $g(c)$ is a weighting function and $N(g)$ is a normalization factor (Rabiner and Juang, 1993). Instead of calculating all the possible paths, the DTW algorithm recursively calculates the minimum cumulative distance for each point using local constraints (Vintsyuk, 1968; Myers *et al.*, 1980; Ney, 1984). The local constraint used in this study limits the connections to the right, the top, and the top right diagonal from the current point. To avoid exaggerated distortions and reduce computing time, a global constraint is applied by a restriction to the area defined by two lines parallel to the diagonal of the matrix X_k. The positions of these two lines are defined by a parameter RA such as $|i - j| \leq RA$ (Sakoe and Chiba, 1978) (Fig. 5). The parameter RA was empirically chosen to be 5 for all the whale call types. The unknown vocalization is classified as the reference template of the dictionary for which the distance $D_w(T, R_k)$ is the minimum.

$$T \triangleq \arg\min_k D_w(T, R_k). \tag{7}$$

The dictionary is composed of the blue whale A, B, D, and AB calls and fin whale 20 Hz call. Noise templates were not included in the dictionary since noise is highly variable and cannot be extensively represented by a limited number of time-frequency contours. Therefore, extracted contours corresponding to noise events were classified as such when the minimum distance $D_w(T,R_k)$ exceeded the threshold T_{DTW}, defined empirically (Table IV).

The choice of the reference templates is known to be a critical phase of this contour classification method. Subjective choices of the reference templates by an operator usually do not perform well. Here we used a semi-supervised selection of the reference templates (Rabiner and Juang, 1993). For each call type, the operator first randomly assigns a call from the training data set as the reference template. Then, each contour of the training data set is compared to the chosen reference templates using the DTW algorithm. If the contour is properly classified it is then properly represented by the chosen template. If its classification is wrong, the tested contour is added to the call template dictionary. This routine is repeated for all calls of the training data set. This procedure removes the subjectivity and minimizes the redundancy of the dictionary. The final dictionary used in this study was composed of 12 A calls, 17 B calls, 7 AB phrases, 17 D calls, and 2 20 Hz pulses.

3. Classification using VQ

VQ is a clustering/classification algorithm that was used for speaker identification (Pan *et al.*, 1985; Soong *et al.*, 1985), respiratory sound classification (Bahoura and Pelletier, 2003), and hand writing recognition (Camastra and Vinciarelli, 2001). It is comprised of two steps: the training and the classification. During the training step, each contour extracted from the training database is represented by a set of features. Here, these features are its minimal frequency, maximum frequency, duration, and frequency shift. Thereby, all the calls are characterized by a set of four-dimensional vectors that can be represented in a four-dimensional feature space. The objective of the training is to represent the cloud of scattered points for each call class, by a small set, C_k (also called codebook) of local centroids, c_{ki}. To do so, the Linde–Buzo–Gray (LBG) algorithm is used (Linde *et al.*, 1980). In contrast to the well known *k -means* algorithm, the initialization of the LBG algorithm is not performed randomly. During the classification step, the set of features from an unknown call contour T is extracted and represented in the feature space. Then, the distance $D(T,C_k)$ separating T from each codebook C_k is calculated by

$$D(T,C_k) = \frac{1}{M}\sum_{i=1}^{M} d(T,c_{ki}), \tag{8}$$

where M is the number of centroids in the codebook C_k, and $d(T,c_{ki})$ is the Euclidean distance separating T from the centroids c_{ki}. The call T is classified as the class for which the distance $D(T,C_k)$ is minimum.

$$T \triangleq \arg\min_k D(T,C_k). \tag{9}$$

Codebooks were created using all the vocalizations of the training data set. The same 5 call classes used for the DTW method were considered and each codebook was represented by 16 centroids ($M=16$). To avoid classifying noise, the tested time-frequency contour is discarded if its distance $D(T,C_k)$ exceeds the empirically determined threshold T_{VQ} (Table IV).

E. Performance evaluation

The false negatives and false positives were counted for each method by running the algorithms on the test data sets. The false negatives correspond to the calls in the test data set that are not detected or that are considered as noise by the classification process. For a given call class, the false negative rate is defined as the ratio of the number of missed calls to the total number of calls in that class. The false positives correspond to periods of noise classified as calls (false alarms). It is calculated by testing each method on the noise data set. The number of false alarms divided by the duration of the noise data set (14 h) gives an average number of false positives per hour. The performance results are presented for several SNR intervals. The SNR calculation is inspired by Mellinger, 2004 and Mellinger and Clark, 2006. The call power is defined on the spectrogram by the average power of the call in the 15–18 Hz frequency range for the A and B calls, 18–26 Hz for the 20 Hz pulses, and 30–100 Hz for the D calls. The noise power is calculated by the average power before and after the call, in the same frequency band for periods equal to the call durations. This calculation is done on the original spectrogram (i.e., without noise reduction). The performances are reported separately for SNR intervals <0 dB, [0, 5 dB], [5, 10 dB], and >10 dB.

To quantify the relative processing time of each method, a real-time index I_r was defined as

$$I_r = \frac{T_p}{T_r}, \tag{10}$$

where T_p is the duration of the detection and classification process, and T_r is the recording duration. The index I_r is <1 if the processing is faster than the real-time and ≥ 1 otherwise. Processing was done with an Intel Pentium® 4 processor working at 3.2 GHz with 1 Gbyte of random access memory (RAM). The computer was running the Microsoft Windows XP® Professional operating system and the processing was performed in MATLAB© (MathWorks, Inc., Natick, MA). The real-time index was computed for each recording of the test data set and analyzed by call type. The A and B calls were processed simultaneously on the same spectrogram and their computation time could not be separated.

III. Results

A. Performance of classification

The test data set represents an accumulated recording of 17 h of A calls, 17 h of B calls, 7 h of D calls, and 3 h of 20 Hz calls. Their respective SNR distribution is given in the first column of Fig. 6. The A and B calls with a SNR between 0 and 10 dB were the most frequent in the data set. The 20 Hz pulses were equally distributed in the SNR intervals >0 dB. The majority of D calls had a SNR of 0–5 dB.

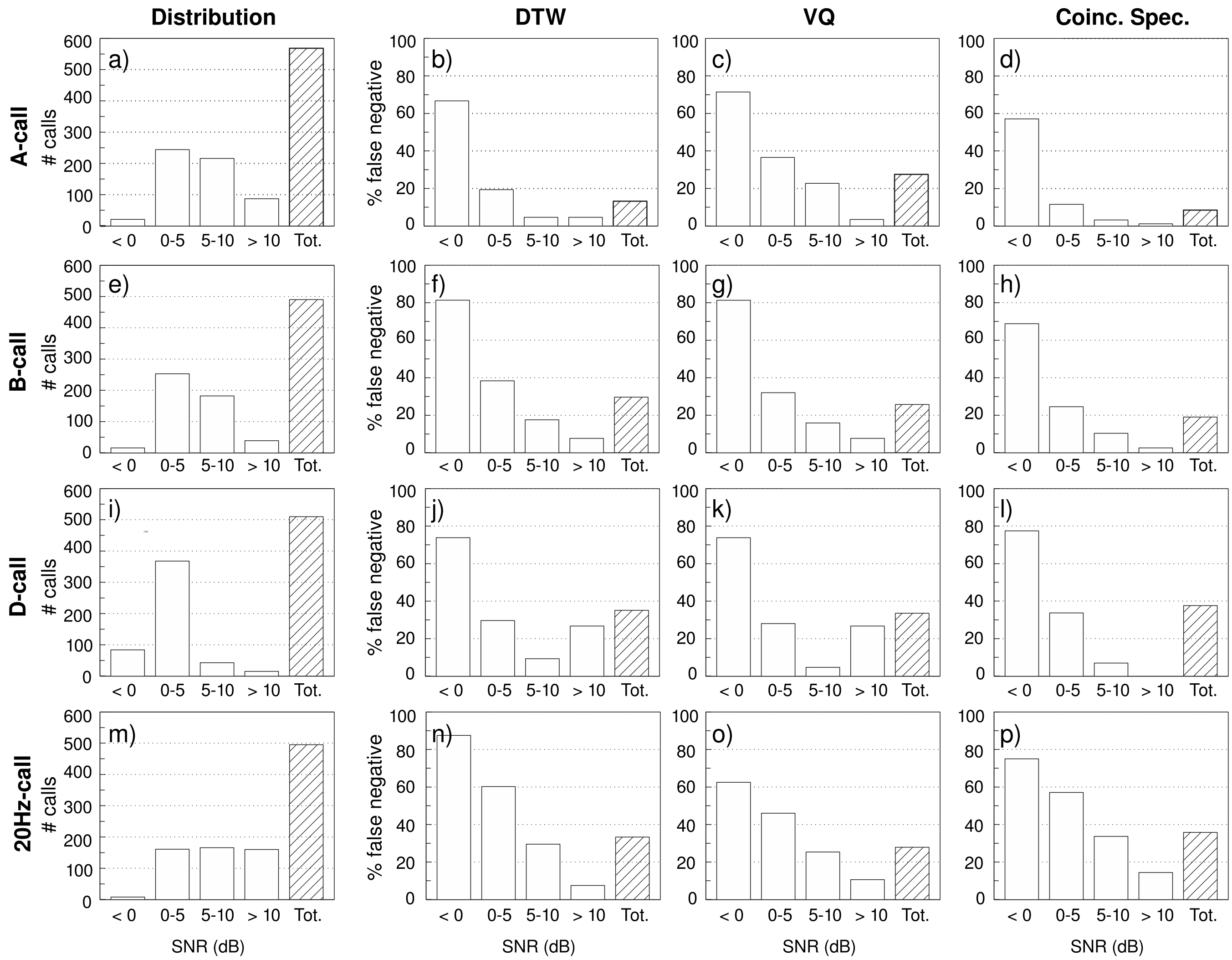

FIG. 6. Signal-to-noise ratio distributions of the (a) A calls, (e) B calls, (i) D calls, and (m) 20 Hz calls used to test the performance of the algorithms. False negative rates for the frequency contour [(b), (f), (j), and (n)] DTW and [(c), (g), (k), and (o)] VQ methods, and [(d), (h), (l), and (p)] the coincidence of spectrograms method, separately for four signal-to-noise ratio intervals and for all calls (hatched bars).

False negative rates for A and B calls decreased when the SNR increased for all methods as expected [Figs. 6(b)–6(d) and 6(f)–6(h)]. The coincidence of spectrograms had the smallest false negative rates with overall values of 8.5% for the A calls and 19% for the B calls. When tested on the noise data set, the three methods had comparable false detection rates for these two call types; coincidence of spectrograms (3 and 2 false alarms per hour for calls A and B, respectively) and DTW (4 and 1 false alarms per hour for calls A and B, respectively) having less false detections than VQ (Table V). For the D calls, VQ had a lower average false negative rate (33.5%) than the other methods [Figs. 6(j)–6(l)]. DTW and VQ showed unexpected increases in false negative rates (26.7%) at SNR > 10 dB. These two methods resulted in 11 and 12 false detections per hour, respectively, compared to 22 for the coincidence of spectrograms method (Table V). For the 20 Hz pulses all three methods presented a gradual decrease in false negative rate with increasing SNR [Figs. 6(n)–6(p)]. Lowest global false negative rates were obtained by VQ and DTW. These methods had, however, a very high number of false alarms per hour (63 with DTW and 74 with VQ, Table V).

TABLE V. Average number of false positives per hour for the three methods tested with the noise data set.

Method	A call	B call	D call	20 Hz call
DTW	4	1	11	63
VQ	4	3	12	74
Coincidence of spectrograms	3	2	22	23

B. Processing time

The coincidence of spectrograms method had the lowest computation time index of the three methods with average values of 0.0083, 0.0039, and 0.013 for the A and B calls, D calls, and 20 Hz calls, respectively [Fig. 7(a)]. The processing time of the VQ method was the same as DTW for the AB and 20 Hz calls with average real-time index value smaller than 0.1. However, VQ real-time index for the D call (average of 0.28) was significantly faster than DTW (average of 0.44) [Figs. 7(b) and 7(c)].

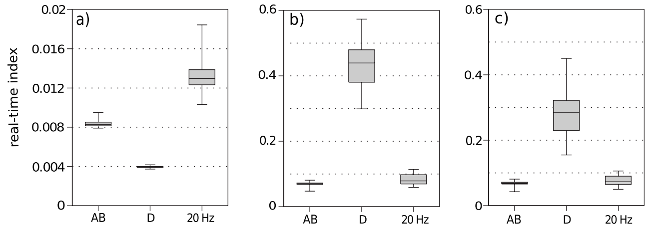

FIG. 7. Box–Whisker plots of the real-time index computation time of the (a) coincidence of spectrograms, (b) DTW, and (c) VQ methods for each call type. Note the different scale in (a).

IV. DISCUSSION

The coincidence of spectrogram method showed the best performance for the A and B calls. For these calls, the different methods differed mainly by their false negative rates. These missed calls were generally resulting from the contour extraction step. Some call contours were truncated into several non-connecting contour segments. The smallest segments were ignored because of the time constraint parameter but the other ones were presented to the classification step and they were classified as noise because of their excessively fragmentary representation of the actual call contour. For the 20 Hz calls, the contour approach, especially VQ, performed better for the false negative criteria, but produced three times more false alarms than the coincidence of spectrogram method. The majority of these false alarms were caused by the low-frequency impulsive noises usually coming from nearby transiting merchant ships that were not filtered out at the noise reduction step. The contour extraction approach best performed for the recognition of D calls. The fixed template used by the coincidence of the spectrograms appeared inappropriate to cope with the variations in duration, frequency sweep rate, and frequency band of the D call. The coincidence of spectrogram method was also more vulnerable to false alarms due to occasional strumming noise in this frequency band from the hydrophone mooring system. As expected, the false negative rates generally increased with decreasing SNRs, except for the D call with VQ and DTW for SNRs$>$10 dB. This exception appeared to result from the distortions of the calls caused by multipath propagation generating overlapping echoes of the calls that modified the extracted classification features. The low number of D calls in this SNR interval (15 calls) does not really allow a definitive conclusion about the performance of the different algorithms at high SNRs.

Computation costs depend on call frequency range (85 Hz for D calls versus 3.5 Hz for A and B calls) for time-frequency contour extraction, and on spectrogram resolution for the coincidence of spectrogram. DTW requires more computation than VQ and its cost is function of the number of call templates used. Recording sampling frequency differs according to call types and methods (Tables III and IV) and therefore also influences the computation times.

The classical method of coincidence of spectrograms was adapted to the noisy acoustic condition of the St. Lawrence by adding an initial step of noise reduction that was critical for successful classification. The simplicity of the logical AND operation makes this algorithm very fast, which fulfills the real-time requirements of this study. Choosing the optimal call templates empirically can be laborious; thus performing statistical analysis of the acoustic parameters of the calls to produce the kernel is recommended. The additional computation time for the contour approach is variable and depends on the complexity of the recordings' content. However, the calls considered in this paper have a fairly simple structure without harmonics, so the computing time remains relatively small. The training steps to determine the parameters for connecting the contour segment and building the dictionary of call models were easy and robust since they were semi-supervised on a representative call database. This study shows that this method has a promising potential for the recognition of marine mammal calls. Its performance is still slightly lower than the classical approach for the stereotyped calls; however, several modifications could be implemented to improve it. To reduce DTW computation time, some clustering techniques can be used to minimize the number of reference templates to consider for the same call type. Representing all calls with the same time-frequency scheme independently of their duration and frequency band would also improve all methods, which presently required computing spectrograms with different resolutions. It may be worth investigating other time-frequency representation techniques such as the Wigner–Ville transform (Caimi and Hassan, 2000). The number of parameters involved in the methods can be cumbersome to manually tune to optimize the classification results. Statistical optimization methods such as the *design of experiment* (Lochner and Matar, 1990) would be a simpler and more accurate way of finding the optimal set of parameter to maximize the algorithm performance. The algorithm could incorporate an additional step to identify and filter out or exploit the frequent delayed copies of the calls from reverberations. Use of propagation models and matched field techniques (e.g., Thode *et al.*, 2000) is among possible methods. This approach is computer intensive and attached to a particular environment.

What is the effect of the observed detection/classification results for PAM application in the St. Lawrence Estuary given fin and blue whale call occurrence frequency? These animals are known to produce very regular repeated calls (Mellinger and Clark, 2003; Samaran, 2004; Berchok *et al.*, 2006), although nonregular patterns are sometimes observed (Oleson *et al.*, 2007). The A and B calls are usually repeated every ~74 s (Mellinger and Clark, 2003). If a blue whale is calling for 1 h without interruptions, it would produce about 51 A and B calls. Referring to Table V and Fig. 6, the DTW classification method would be expected to detect 48 real A calls with a SNR above 5 dB, and 4 false alarms. The same reasoning is followed to assess the detections expected by the different methods and for the other calls (Table VI). By assuming a swimming speed between 1 and 5 $\mathrm{m\ s^{-1}}$ (Goldbogen *et al.*, 2006), all methods correctly detect enough calls for accurately tracking the animals producing A and B calls with a SNR higher than 5 dB. However, for the 20 Hz calls, only the coincidence of spectrograms appears to be suitable at low SNRs. The false alarms could be rejected using a tracking algorithm taking advantage of detection at

TABLE VI. Number of A, B, and 20 Hz calls expected and detected per recording hour, assuming that the animal calls regularly without interruptions, for the three tested algorithms.

Call type	Method	No. of expected calls	No. of correct detections SNR (dB) <0	0–5	5–10	>10	No. of false alarms
A call	DTW	51	17	41	48	48	4
	VQ	51	14	32	39	49	4
	Coincidence of spectrograms	51	21	45	49	50	3
B call	DTW	51	9	31	42	47	1
	VQ	51	9	34	42	47	3
	Coincidence of spectrograms	51	15	38	45	49	2
20 Hz call	DTW	360	45	143	253	333	63
	VQ	360	135	194	269	322	74
	Coincidence of spectrograms	360	90	154	238	308	23

different locations by several instruments differently affected by noise from transiting ships (Seebaruth, 2006; Simard *et al.*, 2008). It is to be noticed that the values in Table V assume a uniform distribution of the false alarms in time, which is not the case in the real world. Because the D calls are not repeated on predictable patterns, it is not possible to follow the same reasoning for this call type.

For assessing the actual density of calls per unit of time or trying to assess the whale density from the calls, the performance of the call detection algorithm under different noise conditions must be taken into account. The relation of the missed detection rate with the SNR level observed with the different methods could then serve to correct the "apparent call density" measured.

ACKNOWLEDGMENTS

This work has been founded by the FQRNT Québec research grant to M.B. and Y.S., Fisheries and Oceans Canada, and NSERC Discovery and Ship time grants to Y.S. We thank the scientific teams and the crews of the RV Coriolis II, CCGS Isle Rouge, and Cap d'Espoir for their generous contribution to the work at sea. Thanks to Professor Jean Rouat (Sherbrooke University) for his very constructive comments on this project. Many thanks to the two anonymous reviewers for their contribution in improving the clarity and rigorousness of this paper. Thanks to Terry Deveau for refining the English.

Anderson, S. E., Dave, A. S., and Margoliash, D. (**1996**). "Template-based automatic recognition of birdsong syllables from continuous recordings," J. Acoust. Soc. Am. **100**, 1209–1219.

Andrew, R. K., Howe, B. M., Mercer, J. A., and Dzieciuch, M. A. (**2002**). "Ocean ambient sound: Comparing the 1960s with the 1990s for a receiver off the California coast," ARLO **3**, 65–70.

Arveson, P. T., and Vendittis, D. J. (**2000**). "Radiated noise characteristics of a modern cargo ship," J. Acoust. Soc. Am. **107**, 118–129.

Au, W. L., and Hastings, M. C. (**2008**). *Principles of Marine Bioacoustics* (Springer, New York).

Bahoura, M., and Pelletier, C. (**2003**). "New parameters for respiratory sound classification," Canadian Conference on Electrical and Computer Engineering, 2003. IEEE CCECE 2003, Vol. **3**, pp. 1457–1460.

Bahoura, M., and Simard, Y. (**2008**). "Chirplet transform applied to simulated and real blue whale (Balaenoptera musculus) calls," in *Image and Signal Processing*, Lecture Notes in Computer Science Vol. **5099** (Springer, Berlin), pp. 296–303.

Bazúa-Durán, C., and Au, W. L. (**2004**). "Geographic variations in the whistles of spinner dolphins (*Stenella longirostris*) of the Main Hawaiian Islands," J. Acoust. Soc. Am. **116**, 3757–3769.

Berchok, C. L., Bradley, D. L., and Gabrielson, T. B. (**2006**). "St. Lawrence blue whale vocalizations revisited: Characterization of calls detected from 1998 to 2001," J. Acoust. Soc. Am. **120**, 2340–2354.

Brown, J. C., and Zhang, B. (**1991**). "Musical frequency tracking using the methods of conventional and 'narrowed' autocorrelation," J. Acoust. Soc. Am. **89**, 2346–2354.

Brown, J. C. (**1992**). "Musical fundamental frequency tracking using a pattern recognition method," J. Acoust. Soc. Am. **92**, 1394–1402.

Brown, J. C., Hodgins-Davis, A., and Miller, P. J. O. (**2006**). "Classification of vocalizations of killer whales using dynamic time warping," J. Acoust. Soc. Am. **119**, EL34–EL40.

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.

Canada (**2008**). "Species at risk public registry," http://www.registrelep.gc.ca/ (Last viewed 4/20/2009).

Caimi, F. M., and Hassan, G. A. (**2000**). "Pattern classification approach to underwater acoustic communications based on the Wigner-Ville distribution," Proc. SPIE **4045**, 167–174.

Camastra, F., and Vinciarelli, A. (**2001**). "Cursive character recognition by learning vector quantization," Pattern Recogn. Lett. **22**, 625–629.

Clemins, P. J., Johnson, M. T., Leong, K. M., and Savage, A. (**2005**). "Automatic classification and speaker identification of African elephant (*Loxodonta africana*) vocalizations," J. Acoust. Soc. Am. **117**, 956–963.

Clemins, P. J., and Johnson, M. T. (**2006**). "Generalized perceptual linear prediction features for animal vocalization analysis," J. Acoust. Soc. Am. **120**, 527–534.

Croll, D. A., Clark, C. W., Calambokidis, J., Ellison, W. T., and Tershy, B. R. (**2001**). "Effect of anthropogenic low-frequency noise on the foraging ecology of Balaenoptera whales," Anim. Conserv. **4**, 13–27.

Datta, S., and Sturtivant, C. R. (**2002**). "Dolphin whistle classification for determining group identities," Signal Process. **82**, 251–258.

Edds, P. (**1982**). "Vocalizations of the blue whale, *Balaenoptera musculus*, in the St. Lawrence River," J. Mammal. **63**, 345–347.

Edds, P. (**1988**). "Characteristics of finback (*Balaenoptera physalus*) vocalizations in the St. Lawrence estuary," Bioacoustics **1**, 131–149.

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

Goldbogen, J. A., Calambokidis, J., Shadwick, R. E., Oleson, E. M., McDonald, M. A., and Hildebrand, J. A. (**2006**). "Kinematics of foraging dives and lunge-feeding in fin whales," J. Exp. Biol. **209**, 1231–1244.

Halkias, X. C., and Ellis, D. P. (**2006**). "Call detection and extraction using Bayesian inference," Appl. Acoust. **67**, 1164–1174.

Halkias, X. C., and Ellis, D. P. (**2008**). "A comparison of pitch extraction methodologies for dolphin vocalization," Can. Acoust. **36**, 74–80.

Hoyt, E. (**2001**). *Whale Watching 2001: Worldwide Tourism Numbers, Expenditures and Expanding Socioeconomic Benefits* (International Fund for Animal Welfare, Yarmouth Port, MA).

Ito, K., Mori, K., and Iwasaki, S. (**1996**). "Application of dynamic programming matching to classification of budgerigar contact calls," J. Acoust. Soc. Am. **100**, 3947–3956.

Jensen, A., and Silber, G. K. (**2004**). "Large whale ship strike database," NOAA Technical Memorandum NMFS-OPR, U.S. Department of Commerce, Silver Spring, MA.

Kogan, J. A., and Margoliash, D. (**1998**). "Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study," J. Acoust. Soc. Am. **103**, 2185–2196.

Laist, D. W., Knowlton, A. R., Mead, J. G., Collet, A. S., and Podesta, M. (**2001**). "Collision between ships and whales," Marine Mammal Sci. **17**, 35–75.

Linde, Y., Buzo, A., and Gray, R. (**1980**). "An algorithm for vector quantizer design," IEEE Trans. Commun. **28**, 84–95.

Lochner, R. H., and Matar, J. E. (**1990**). *Designing for Quality: An Intro-*

duction to the Best of Taguchi and Western Methods of Statistical Experimental Design (Quality Resources, New York).
McDonald, M. A., Hildebrand, J. A., and Wiggins, S. M. (**2006**). "Increases in deep ocean ambient noise in the Northeast Pacific west of San Nicolas Island, California," J. Acoust. Soc. Am. **120**, 711–718.
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. (**1997**). "Methods for automatic detection of mysticete sounds," Mar. Freshwater Behav. Physiol. **29**, 163–181.
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., and Clark, C. W. (**2003**). "Blue whale (*Balaenoptera musculus*) sounds from the North Atlantic," J. Acoust. Soc. Am. **114**, 1108–1119.
Mellinger, D. K., and Clark, C. W. (**2006**). "MobySound: A reference archive for studying automatic recognition of marine mammal sounds," Appl. Acoust. **67**, 1226–1242.
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," Oceanogr. **20**, 36–45.
Munger, L., Mellinger, D., Wiggins, S., Moore, S., and Hildebrand, J. (**2005**). "Performance of spectrogram correlation in detecting right whale calls in long-term recordings from the Bering Sea," Can. Acoust. **33**, 25–34.
Myers, C., Rabiner, L., and Rosenberg, A. (**1980**). "Performance tradeoffs in dynamic time warping algorithms for isolated word recognition," IEEE Trans. Acoust., Speech, Signal Process. **28**, 623–635.
Ney, H. (**1984**). "The use of a one-stage dynamic programming algorithm for connected word recognition," IEEE Trans. Acoust., Speech, Signal Process. **32**, 263–271.
NRC (**2003**). *Ocean Noise and Marine Mammals* (The National Academy Press, Washington, DC).
Oleson, E. M., Calambokidis, J., Burgess, W. C., McDonald, M. A., LeDuc, C. A., and Hildebrand, J. A. (**2007**). "Behavioral context of call production by eastern North Pacific blue whales," Mar. Ecol.: Prog. Ser. **330**, 269–284.
Pan, K.-C., Soong, F., and Rabiner, L. (**1985**). "A vector-quantization-based preprocessor for speaker-independent isolated word recognition," IEEE Trans. Acoust., Speech, Signal Process. **33**, 546–560.
Perry, S. L., DeMaster, D. P., and Silber, G. K. (**1999**). "The great whales: History and status of six species listed as endangered under the U.S. Endangered Species Act of 1973," Mar. Fish. Rev. **61**, 1–82.
Potter, J. R., Mellinger, D. K., and Clark, C. W. (**1994**). "Marine mammal call discrimination using artificial neural networks," J. Acoust. Soc. Am. **96**, 1255–1262.
Rabiner, L., Cheng, M., Rosenberg, A., and McGonegal, C. (**1976**). "A comparative performance study of several pitch detection algorithms," IEEE Trans. Acoust., Speech, Signal Process. **24**, 399–418.
Rabiner, L. (**1977**). "On the use of autocorrelation analysis for pitch detection," IEEE Trans. Acoust., Speech, Signal Process. **25**, 24–33.
Rabiner, L., and Juang, B. (**1993**). *Fundamentals of Speech Recognition* (Prentice-Hall, Englewood Cliffs, NJ).
Renevey, P., and Drygajlo, A. (**2001**). "Entropy based voice activity detection in very noisy conditions," in Proceedings of the 7th European Conference on Speech Communication and Technology EUROSPEECH '2001, Aalborg, Denmark, pp. 1887–1890.
Richardson, W. J., Greene, C. J., Malme, C., and Thomson, D. (**1995**). *Marine Mammals and Noise* (Academic, San Diego, CA).
Samaran, F. (**2004**). "Détectabilité des vocalisations de rorquals communs (*Balaenoptera physalus*) à partir d'une station côtière dans la voie maritime de l'estuaire du Saint-Laurent [Detectability of fin whale calls (*Balaenoptera physalus*) from a coastal array in the St. Lawrence Estuary]," MSc thesis, University of Québec in Rimouski, QC, Canada.
Sakoe, H., and Chiba, S. (**1978**). "Dynamic programming algorithm optimization for spoken word recognition," IEEE Trans. Acoust., Speech, Signal Process. **26**, 43–49.
Schevill, W., Watkins, W., and Backus, R. (**1964**). "The 20-cycle signal and Balaenoptera (fin whales)," in *Marine Bioacoustics*, edited by W. N. Tavolga, Vol. **1** (Pergamon, Oxford), pp. 147–152.
Sears, R., and Calambokidis, J. (**2002**). "Update COSEWIC status report on the blue whale Balaenoptera musculus in Canada," COSEWIC assessment and update status report on the blue whale *Balaenoptera musculus* in Canada, Committee on the Status of Endangered Wildlife in Canada, Ottawa, Canada, http://www.registrelep.gc.ca/virtual_sara/files/cosewic/sr%5Fblue%5Fwhale%5Fe%2Epdf (Last viewed 4/20/2009).
Seebaruth, D. (**2006**). "Algorithmes de localisation de baleines pour le Saint-Laurent (Localization algorithms for the whales of the St. Lawrence)," MSc thesis, Departement de Génie Électrique et de Génie Informatique (GEGI), University of Sherbrooke, QC, Canada.
Simard, Y., Bahoura, M., Park, C. W., Rouat, J., Sirois, M., Mouy, X., Seebaruth, D., Roy, N., and Lepage, R. (**2006b**)."Development and experimentation of a satellite buoy network for real-time acoustic localization of whales in the St. Lawrence," in *OCEANS'06 MTS/IEEE*, Boston, MA (IEEE, Piscataway, NJ), pp. 1–6.
Simard, Y., and Lavoie, D. (**1999**). "The rich krill aggregation of the Saguenay–St. Lawrence Marine Park: Hydroacoustic and geostatistical biomass estimates, structure, variability, and significance for whales," Can. J. Fish. Aquat. Sci. **56**, 1182–1197.
Simard, Y., and Roy, N. (**2008**). "Detection and localization of blue and fin whales from large-aperture autonomous hydrophone arrays: A case study from the St. Lawrence estuary," Can. Acoust. **36**, 104–110.
Simard, Y., Roy, N., and Gervaise, C. (**2006a**). "Shipping noise and whales: World tallest ocean liner vs largest animal on earth," in *OCEANS'06 MTS/IEEE*, Boston, MA (IEEE, Piscataway, NJ), pp. 1–6.
Simard, Y., Roy, N., and Gervaise, C. (**2008**). "Passive acoustic detection and localization of whales: Effects of shipping noise in Saguenay–St. Lawrence Marine Park," J. Acoust. Soc. Am. **123**, 4109–4117.
Soong, F., Rosenberg, A., Rabiner, L., and Juang, B. (**1985**). "A vector quantization approach to speaker recognition," IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '85, Vol. **10**, pp. 387–390.
Stafford, K. M., Fox, C. G., and Clark, D. S. (**1998**). "Long-range acoustic detection and localization of blue whale calls in the northeast Pacific Ocean," J. Acoust. Soc. Am. **104**, 3616–3625.
Sturtivant, C. R., and Datta, S. (**1995a**). "Techniques to isolate dolphin whistles and other tonal sounds from background noise," Acoust. Lett. **18**, 189–193.
Sturtivant, C. R., and Datta, S. (**1995b**). "The isolation from background noise and characterization of bottlenose dolphin (Tursiops truncatus) whistles," J. Acoust. Soc. India **23**, 199–205.
Thode, A. M., D'Spain, G. L., and Kuperman, W. A. (**2000**). "Matched field processing, geoacoustic inversion, and source signature recovery of blue whale vocalizations," J. Acoust. Soc. Am. **107**, 278–286.
Tiemann, C. O., Porter, M. B., and Fraser, L. N. (**2004**). "Localization of marine mammals near Hawaii using an acoustic propagation model," J. Acoust. Soc. Am. **115**, 2834–2843.
Thompson, P. O., and Friedl, W. A. (**1982**). "A long term study of low frequency sound from several species of whales off Oahu, Hawaii," Cetology **45**, 1–19.
Thompson, P. O., Findley, L. T., and Vidal, O. (**1992**). "20-Hz pulses and other vocalizations of fin whales, *Balaenoptera physalus*, in the Gulf of California, Mexico," J. Acoust. Soc. Am. **92**, 3051–3057.
Urazghildiiev, I. R., and Clark, C. W. (**2006**). "Acoustic detection of North Atlantic right whale contact calls using the generalized likelihood ratio test," J. Acoust. Soc. Am. **120**, 1956–1963.
Urazghildiiev, I. R., and Clark, C. W. (**2007**). "Acoustic detection of North Atlantic right whale contact calls using spectrogram-based statistics," J. Acoust. Soc. Am. **122**, 769–776.
Urick, R. J. (**1983**). *Principles of Underwater Sound*, 3rd ed. (McGraw-Hill, New York).
Vintsyuk, T. K. (**1968**). "Speech discrimination by dynamic programming," Kibernetika **4**, 81–88.
Wiggins, S. M., Oleson, E. M., McDonald, M. A., and Hildebrand, J. A. (**2005**). "Blue whale (*Balaenoptera musculus*) diel call patterns offshore of Southern California," Aquat. Mamm. **31**, 161–168.