Digital Signal Processing 20 (2010) 1256–1263

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Blue whale calls classification using short-time Fourier and wavelet packet transforms and artificial neural network

Mohammed Bahoura [a,*], Yvan Simard [b,c]

[a] Department of Engineering, University of Quebec at Rimouski, 300, allée des Ursulines, Rimouski, Qc, Canada, G5L 3A1
[b] Marine Sciences Institute, University of Quebec at Rimouski, 310, allée des Ursulines, Rimouski, Qc, Canada, G5L 3A1
[c] Maurice Lamontagne Institute, Fisheries and Oceans Canada, 850 Route de la Mer, C.P. 1000, Mont-Joli, Qc, Canada, G5H 3Z4

ARTICLE INFO

Article history:
Available online 27 October 2009

Keywords:
Blue whale calls
Feature extraction
Classification
Pattern recognition
Short-time Fourier transform
Wavelet packet transform
Multilayer perceptron

ABSTRACT

Two new characterization methods based on the short-time Fourier and the wavelet packet transforms are proposed to classify blue whale calls. The vocalizations are divided into short-time overlapping segments before applying these transforms to each segment. Then, the feature vectors are constructed by computing the coefficient energies within two subbands in order to capture the AB phrase and D vocalization characteristics, respectively. Finally, a multilayer perceptron (MLP) is used to classify the vocalization into A, B and D classes. The proposed methods present high classification performance (86.25%) on the tested database.

1. Introduction

The largest animal on earth, the blue whale, can be identified by its signature calls in all oceans [1]. These regularly produced and powerful low-frequency (< 100 Hz) vocalizations can propagate over distances exceeding 100 km in deep basins, often being guided by sound channels defined by the local sound speed profile [2]. Low-frequency sensitive hydrophones can be used to record these calls and the recordings can be analyzed to track the presence of whales over the monitored basin [3]. For long time-series and numerous hydrophones, automated detection is essential for efficient use of these operational PAM (passive acoustic monitoring) techniques for studying marine mammals in their environment. Several automatic detection algorithms can be developed [4,5] and their relative performance tested for different ocean environments, signal to noise conditions and PAM application contexts, including real-time operation. Here we present two new methods to detect blue whale signature calls which are tested from a representative subset of vocalizations recorded in the St. Lawrence Estuary, a blue whale feeding environment which has the particularity of being crossed by a major continental seaway, generating a high level of shipping noise in the vocalization band [6].

Two techniques are generally used to recognize blue whale calls: The matched filter [7,8] and the spectrogram correlation [7,4,9,5]. The matched filter is implemented with a synthetic signal instead an actual sound. The synthetic signal has the advantage of being noise free and of allowing inclusion of features obtained by averaging several sounds instead using a simple sound that may not be representative of all calls. The spectrogram correlation, however, operates on spectrograms and performs two-dimensional correlations using a synthetic time-frequency image. Effectiveness of these methods depends considerably on the synthetic kernel design. Therefore, the first step is to measure the characteristics of a representative set of calls to synthesize the filter kernel [7]. The chirplet transform is an other approach recently introduced to characterize

* Corresponding author.
E-mail address: Mohammed_Bahoura@uqar.qc.ca (M. Bahoura).

1051-2004/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.dsp.2009.10.024

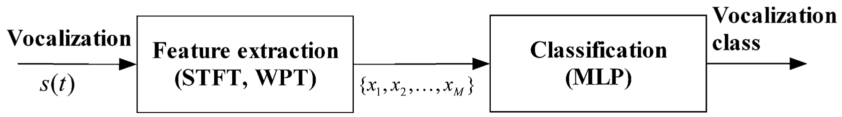

Fig. 1. Bloc diagram of a typical pattern recognition system.

blue whale calls [10]. However, the classification method based on this approach is not fully automated because the choice of the appropriate bandpass filter (13–23 Hz for the AB phrases or 40–80 Hz for the D calls) is manually operated [10].

2. Database

The recordings were collected in the Saguenay-St. Lawrence Marine Park at the head of the 300-m deep Laurentian channel in the Lower St. Lawrence Estuary. The hydrophones were AURAL autonomous hydrophones (Multi-Electronique Inc., Rimouski, Qc, Canada) moored as standard oceanographic moorings at intermediate depths in the water column where a well defined sound channel develops in summer. The 16-bit recording data were acquired at the 2000 Hz optional sampling rate of the AURALs, which includes an appropriate antialiasing low-pass filter (1000 Hz). Details can be found in [6,11,12]. The calls of interest here are the signature calls of North Atlantic blue whales; notably the A and B infrasounds (15–20 Hz), often occurring together in AB phrases, and the audible D-call (35–120 Hz) also known as arch sound [13]. The D call is less stereotyped than A and B [5]. The variation within each call type are generally attributed to propagation effects. Details on the St. Lawrence blue wall call repertoire can be found in [13].

3. Method

A typical pattern recognition system includes two blocs: feature extraction and classification (Fig. 1). These blocs are described in the following subsections.

3.1. Feature extraction

Feature extraction is a process that transforms originally high-dimensional vectors into lower-dimensional vectors. It is a mapping $f : \mathbb{R}^N \to \mathbb{R}^D$, where $D \ll N$. Feature extraction can be considered as a data reduction process that attempts to capture the essential characteristics of the analyzed signal with a small data rate [14].

In this paper, we propose two feature extraction methods based on short-time Fourier transform (STFT) and wavelet packet transform (WPT). These techniques extract a reduced number of features from two subbands corresponding to the frequency ranges of AB phrases and D vocalizations, respectively.

3.1.1. Fourier transform

The short-time Fourier transform (STFT) of a discrete-time signal $s[n]$ is a Fourier transform performed in successive frames:

$$S[m,k] = \sum_n s[n]w[n-m]e^{-j2\pi nk/N} \tag{1}$$

where $w[n]$ is a short-time windowing function of size L, centered at time location m, and N is the number of discrete frequencies ($N \geqslant L$). Usually, N is chosen to be a power-of-2 for using an efficient fast Fourier transform (FFT). Since the Fourier transform is a complex function, we use the power spectrum density (PSD) given by:

$$P_s[m,k] = \frac{1}{N}\left|S[m,k]\right|^2 \tag{2}$$

At the sampling frequency f_S, each windowed segment (frame) is represented by N-points PSD covering the frequency range $[-\frac{f_S}{2}, \frac{f_S}{2}[$. As power spectrum is symmetric, it can be described by only N discrete frequencies $f_k = kf_S/N$, where $0 \leqslant k < N/2$. For a given sampling frequency ($f_S = 250$ Hz), a large frame length (N) improves the frequency resolution at the expense of the temporal resolution. In this work, we used a Hamming window function of length $N = 512$ and a 50% overlap.

The PSD cannot be used directly as feature vector because its too large number of points ($N/2$) and the fact that the calls of interest cover only a part of the spectrum. Thus, the STFT-based characterization method consists in extracting features from two subbands, (15.625–20.996 Hz) and (39.062–85.449 Hz) respectively corresponding to the AB and D calls frequency ranges. The first six components of the feature vector $\mathbf{x}_m$ are obtained by sequentially averaging PSD points between $P_s[m,31]$ and $P_s[m,42]$ by bins of 2 points (Fig. 2). The last six features are similarly obtained for the PSD interval from $P_s[m,79]$ to $P_s[m,174]$ but with bins of 15 points. For a given frame m, the feature vector component $x_m[n]$ is defined by Eq. (3).

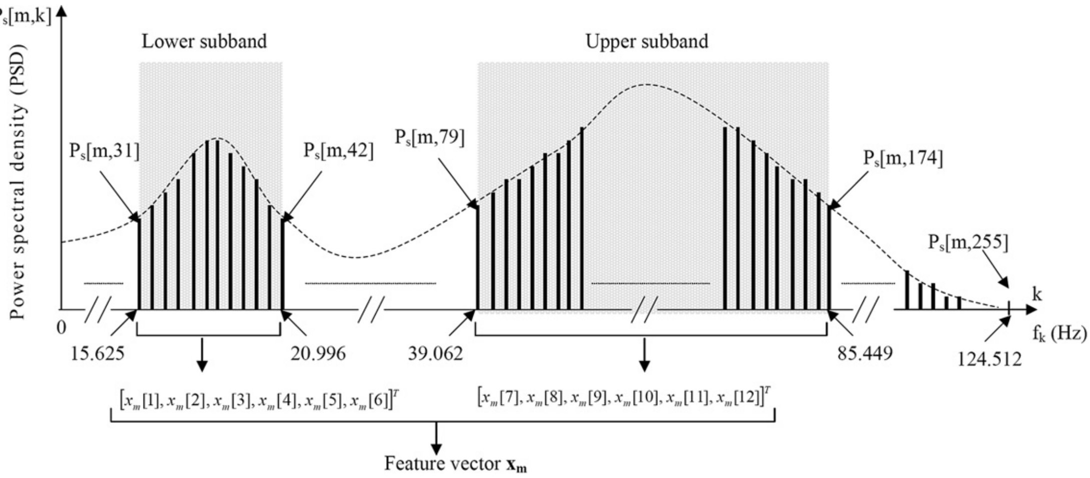

Fig. 2. Example of the feature extraction process using short-time Fourier transform (STFT). For each segment m, the feature vector $\mathbf{x}_m$ is constructed by extracting six components from lower subband (15.625–20.996 Hz) and six components from upper subband (39.062–85.449 Hz). The discrete frequencies are obtained with a sampling frequency of 250 Hz and a frame length of 512 samples.

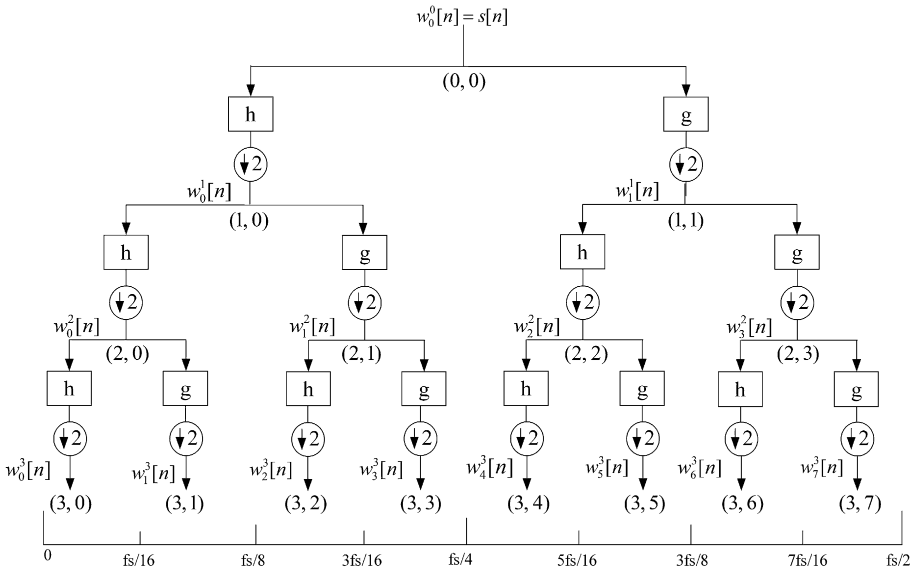

Fig. 3. The 3-level wavelet packet decomposition tree. Each unit consists of a low-pass (h) and high-pass (g) filter pair, followed by a decimation (↓). The frequency band is uniformly divided.

$$x_m[n] = \begin{cases} \frac{1}{2}\sum_{k=29+2n}^{30+2n} P_s[m,k] & n = 1, 2, \ldots, 6 \\ \frac{1}{15}\sum_{k=63+16(n-6)}^{78+16(n-6)} P_s[m,k] & n = 7, 8, \ldots, 12 \end{cases} \tag{3}$$

Hence, a 12-dimensional feature vector $\mathbf{x}_m = [x_{m,1}, x_{m,2}, \ldots, x_{m,12}]^T$ is constructed, where T represents the transpose operation.

3.1.2. Wavelet packet transform

The wavelet packet transform (WPT), proposed by Coifman and Wickerhauser [15], can be considered as an extension of the wavelet transform (see Fig. 3). Unlike the wavelet transform, which is obtained by iterating the low-pass branch, the wavelet packet transform is obtained by iterating both low-pass (h) and high-pass (g) branches at each level j. The wavelet packet decomposition is equivalent to a multichannel filtering where the number of filters and their bandwidths are related to the level j. The wavelet packet coefficients $w_k^{j+1}[n]$ corresponding to even and odd k subbands are given by:

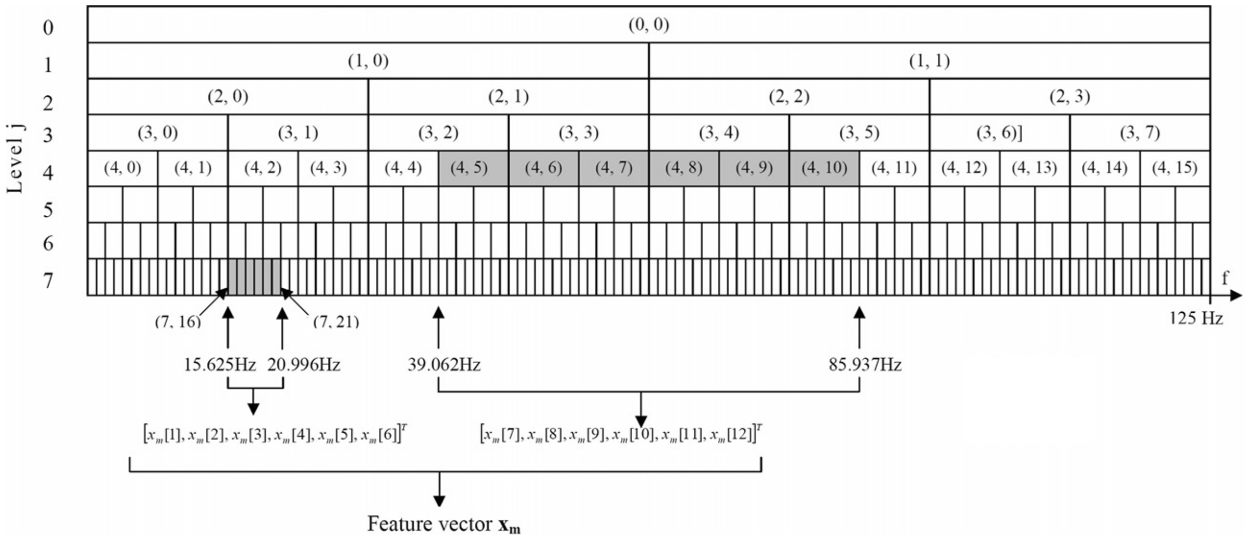

Fig. 4. The wavelet packet decomposition tree at level 7. For each frame, the feature vector **x** is constructed by extracting twelve components extracted from selected subbands corresponding to the calls (shaded areas). The frequency values are obtained with a sampling frequency of 250 Hz and a frame length of 512 samples.

$$w_{2p}^{j+1}[n] = \sum_m h[m-2n] w_p^j[m] \quad \text{even } k \tag{4}$$

$$w_{2p+1}^{j+1}[n] = \sum_m g[m-2n] w_p^j[m] \quad \text{odd } k \tag{5}$$

where $p = 0, \ldots, 2^j - 1$ refers to the subband of level j.

For a given level j, the WPT decomposes the input signal $s[n]$ of length N into 2^j subbands corresponding to the set of wavelet coefficients.

$$w_k^j[n] = \text{WPT}\{s[n], j\} \tag{6}$$

where $w_k^j[n]$ defines the nth coefficient of the kth subband, where $n = 0, \ldots, \frac{N}{2^j} - 1$ and $k = 0, \ldots, 2^j - 1$. In fact, n is the time index and k is the frequency index. For a given sampling frequency ($f_S = 250$ Hz), the time and frequency resolutions depend on the frame length N and the level j. For comparing this characterization method with the previous one, the WPT is computed for $j = 7$ using rectangular window function of length $N = 512$ with a 50% overlap. The choice of the mother wavelet depends on the application. By testing several wavelets (Battle–Lemarie, Beylkin, Coiflet, Daubechies and Symmlet) of various orders, the best classification rate is obtained with Battle–Lemarie wavelet of order 5.

The WPT-based characterization consists to extract features from the same subbands as previously. As shown in Fig. 4, the first six components of the feature vector are obtained by computing the wavelet coefficient energies of nodes located between $(7, 16)$ and $(7, 21)$, while the last six ones are obtained from nodes located between $(4, 5)$ and $(4, 10)$. For a given frame m, the feature vector $x_m[k]$ is defined by Eq. (7).

$$x_m[k] = \begin{cases} \frac{1}{4} \sum_{n=0}^{3} |w_{k+15}^7[n]|^2 & k = 1, 2, \ldots, 6 \\ \frac{1}{32} \sum_{n=0}^{31} |w_{k-2}^4[n]|^2 & k = 7, 8, \ldots, 12 \end{cases} \tag{7}$$

Then a 12-dimensional feature vector $\mathbf{x}_m = [x_{m,1}, x_{m,2}, \ldots, x_{m,12}]^T$ is constructed. Therefore, a given vocalization $s[n]$ is characterized by a sequence of feature vectors $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M\}$, where M is the number of overlapped frames.

Figs. 5, 6 and 7 present the spectrogram and the feature vectors sequence based on STFT and WPT for a typical A, B and D vocalizations. The time-domain waveforms of these calls aren't presented because they don't provide pertinent information. It can be seen that both STFT and WPT characterization methods can be used to easily separate A and B calls from the D call, but it is more difficult to discriminate between the A and B classes. It can be also seen that D call has a lower energy than the A and B calls has expected from [13].

3.2. Classification

The multi-layer perceptron (MLP) is the most popular and the simplest artificial neural network (ANN) for solving pattern classification problems. This network consists of a set of neuron (also called nodes or units) that are arranged in layers: an input layer that acts as a data buffer, one or more hidden layers, and an output layer. The hidden layers enable the network

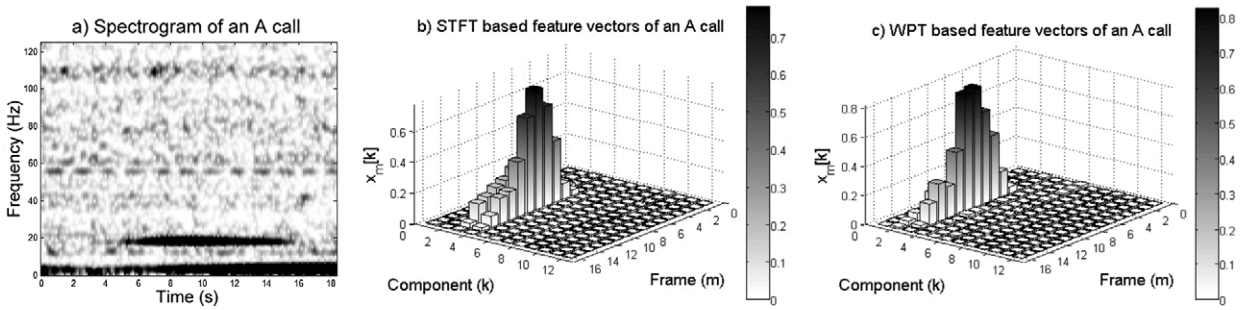

Fig. 5. (a) Spectrogram of an A call, (b) feature vectors based on STFT, and (c) feature vectors based WPT. The feature vectors are computed using a frame length of 512 samples with 50% overlap.

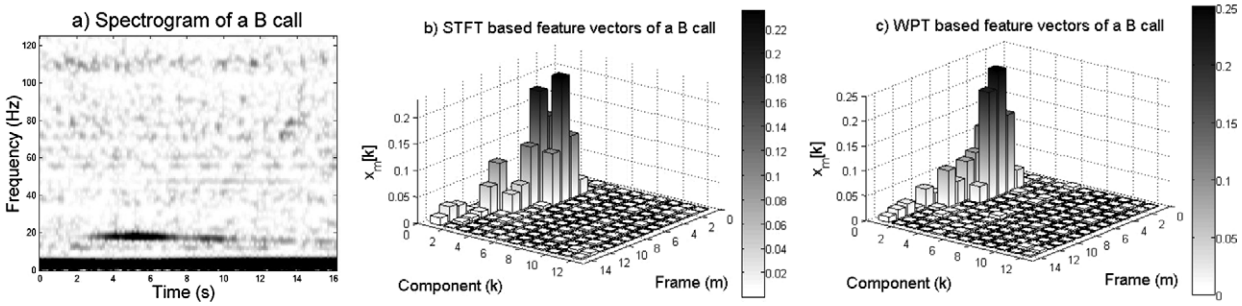

Fig. 6. As in Fig. 5 but for a B call.

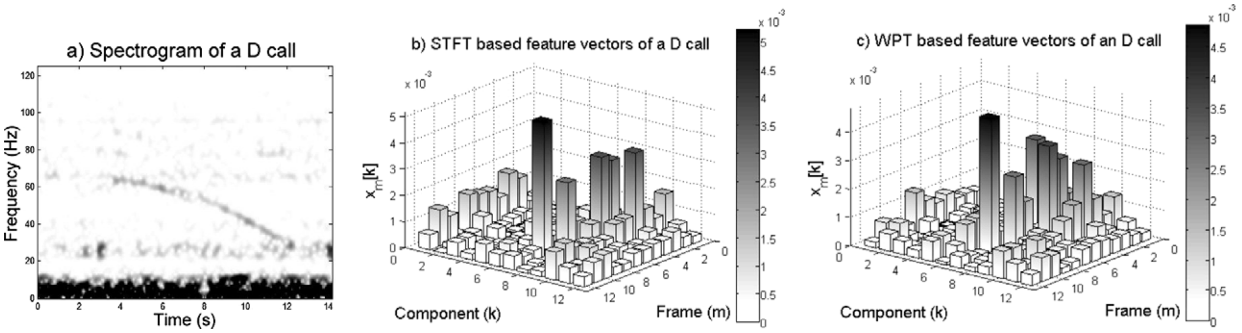

Fig. 7. As in Fig. 5 but for a D call.

to process data that are not linearly separable. It has been demonstrated that an MLP with only one hidden layer can serve as a universal estimator if sufficient neurons are used [16,17]. Signals pass into the input layer neurons, progress forward through the hidden layers and finally emerge from the output layer [18]. Each hidden and output neuron is a processor element that first computes a weighted sum of the output of all neurons in the previous layer, then adds or subtracts a threshold value (bias), and finally passes it through an activation function. Fig. 8 represents an example of an MLP network characterized by D inputs, one hidden layer of N neurons, and K outputs. The output of the hidden neuron j is given by Eq. (8).

$$z_j = f_h\left(\sum_{i=0}^{D} w_{j,i}^{h} x_i\right) \quad j = 1, \ldots, N \tag{8}$$

where $w_{j,i}$ is the connection weight that joins neuron i from the input layer to neuron j of the hidden layer, f_h is the activation function, and $x_0 = 1$ is the bias input of the hidden layer. In the same manner, the output of each node j, in hidden layer, is given by Eq. (9).

$$y_j = f_o\left(\sum_{i=0}^{N} w_{j,i}^{o} z_i\right) \quad j = 1, \ldots, K \tag{9}$$

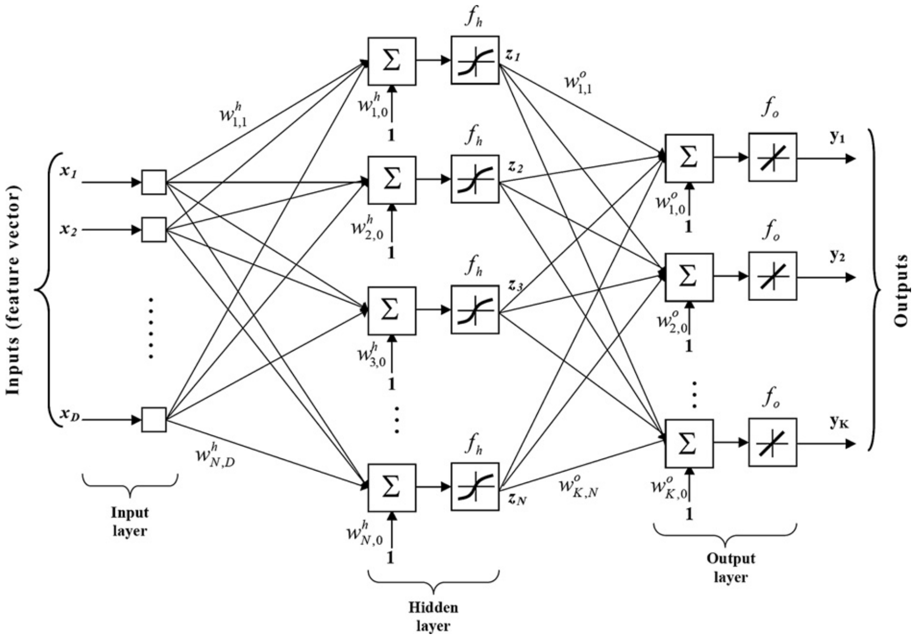

Fig. 8. A typical example of multi-layer perception (MLP). This network is characterized by D inputs, one hidden of N nodes, and K outputs. $w^h_{j,i}$ and $w^o_{j,i}$ are the connection weights of the hidden layer and the output layer respectively.

where f_o is the transfer function, $w^o_{j,i}$ is the connection weight, and $z_0 = 1$ is the bias input of the output layer. The activation functions of the hidden neurons are typically hyperbolic tangent or logistic sigmoid. The output neurons activation function can be linear or non-linear depending on the task performed by the network: a function approximation or a classification, respectively [19]. In this study, we use a *hyperbolic tangent* function $f_h(x) = (1 - e^{-2x})/(1 + e^{-2x})$ for the hidden layer and a *logistic sigmoid* function $f_o(x) = (1 + e^{-x})^{-1}$ for the output layer.

The connection weights $\mathbf{w} = \{w^h_{j,i}, w^o_{j,i}\}$ are determined in the training phase using a set of inputs for which the desired outputs are known $\{(\mathbf{x_1}, \mathbf{d_1}), (\mathbf{x_2}, \mathbf{d_2}), \ldots, (\mathbf{x_p}, \mathbf{d_p})\}$. In the pattern classification problems, the desired or target outputs are vectors of K elements, where K is the number of the reference classes. For each desired vector, $\mathbf{d_i}$, one element corresponding to the presented input pattern, $\mathbf{x_i}$, is set to 1 and the others are set to 0. To accomplish the training task, the backpropagation (BP) algorithm is commonly used [20]. More details can be found in [14].

The training procedure should be repeated until an acceptable error rate is achieved or until a certain number of iterations are completed using training examples. In this experiment, the learning rate, the target error (MSE) and the number of iterations are fixed to 0.01, 0.001 and 300, respectively. As the connections weights $w_{j,i}$ are initialized with random values, the learning and testing process is repeated 50 times in order to take the average value.

To classify an unknown vocalization, the observation sequence $\mathbf{X} = \{\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_M}\}$, corresponding to the feature vectors obtained from M adjacent segments of this vocalization, is passed through the network yielding the actual output sequence $\mathbf{Y} = \{\mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_M}\}$. For each output k, the network provides a set of M values $\{y_{i,k}\}$, $i = 1, \ldots, M$. To classify the entire vocalization, the mean values of the individual outputs are commonly used. The unknown call is identified as the reference call that corresponds to the largest mean value of the individual outputs:

$$\hat{k} = \arg \max_{1 \leqslant k \leqslant K} \{\bar{y}_k\} \tag{10}$$

where the mean values $\bar{y}_k$ are computed over the M segments.

$$\bar{y}_k = \frac{1}{M} \sum_{i=1}^{M} y_{i,k} \quad k = 1, \ldots, K \tag{11}$$

In the pattern classification problems, the size of the feature vector determines the number of neurons in the input layer, while the number of reference classes determines the number of neurons in the output layer. The number of hidden neurons has to be determined by means of a trial and error procedure [19]. A balance must be found between a few number neurons that can lead to underfitting and too many neurons can contribute to overfitting [21]. In this study, the model with 25 neurons in hidden layer gave the best performance.

Table 1
Optimized values of the parameters affecting classification.

Feature extraction			Classification	
Method	STFT	WPT	Method	MLP
Window type	Hamming	Rectangular	Nodes in hidden layer	25
Frame overlap	50%	50%	Learning rate	0.01
Feature vector length	12	12	Target error (MSE)	0.001
Type of wavelet	–	Battle–Lemarie	Number of iterations	300
			Learning algorithm	Resilient BP

Table 2
Performances obtained with STFT and WPT based classifiers.

Vocalization	Performance (%)	
	STFT/MLP	WPT/MLP
A	80.48	80.32
B	82.34	78.06
D	95.94	94.28
Total	86.25	84.22

3.3. Evaluation criteria

For each vocalization class, the classification performance is evaluated by the ratio of the number of the correctly classified vocalization (NCCV) to the number of tested vocalizations (NTV).

$$\text{Performance}\,(\%) = \frac{\text{NCCV}}{\text{NTV}} \times 100 \tag{12}$$

The STFT and WPT characterization methods are evaluated using a neural network classifier (MLP). These techniques are implemented in Matlab using WaveLab [22] and neural network [21] toolboxes.

4. Results and discussion

Blue whale vocalizations were extracted manually and categorized into the three classes (A, B and D) by visualizing their spectrogramme using *Adobe Audition* software [23]. Each class of the used database contains 100 calls. Due to the limited size of this test dataset, the *"k-fold cross-validation"* method is employed to evaluate the classification performance. The principle consists in dividing each class into 10 groups, each time we use 9 groups for training and the last one for the classification test. As shown on examples of Figs. 5–7, these vocalizations are usually corrupted with various kinds of noises.

After normalization, the signals were analyzed to extract the feature vectors that are used as inputs of the neural network (MLP) classifiers. The performance of these classifiers are affected by various parameters related to the characterization methods (window type, frame overlap, type of wavelet) and the architecture of the neural network (number of neurons in hidden layer, the learning rate, target error and number of iterations). Optimal values of these parameters were determined by repeated experimentations. We have tested different weighting windows (blackman, hamming, hanning and rectangular), overlap rates (25%, 50% and 75%), wavelets (Battle–Lemarie, Beylkin, Coiflet, Daubechies and Symmlet) of various orders, and number of neurons in hidden layer. Also, four learning algorithms were investigated, namely, adaptive learning rate BP (GDA), resilient BP (RP), scaled conjugate gradient (SCG), and Levenberg–Marquardt (LM) algorithms. The resilient BP algorithm was selected because it was the fastest for this classification problem. The best performances were obtained by the values reported in Table 1.

One of the problems that occurs during neural network training is called overfitting [21]. This situation occurs when the error on the training set is driven to a very small value, but when new data is presented to the network the error becomes large. The network has then memorized the training examples, but has not learned to generalize to new situations. One method for improving network generalization is to use a network that is just large enough to provide an adequate fit [21]. After only 50 iterations in the learning phase, the neural network provides interesting results. The classification performance continued to improve by increasing the number of iterations, and became almost constant after 300 iterations.

Using the parameter values of Table 1, the performance of the resulting classification methods obtained with the database are reported in Table 2. These results represent the averaged values of 50 repeated tests. It can be seen that STFT and WPT based methods give comparable global performances (86.25% and 84.22%, respectively). Table 3 gives the confusion matrix of the true class versus assigned (predicted) class for STFT and WPT based methods, which indicates the high performance of these methods to separate D class from A and B (95.94% and 94.28%, respectively), and their difficulties to discriminate A and B classes (error of about 17%).

Table 3
Confusion matrix obtained with STFT and WPT based classifiers.

True class	Assigned class (STFT/MLP)			Assigned class (WPT/MLP)		
	A	B	D	A	B	D
A	80.48	16.94	2.58	80.32	16.60	3.02
B	15.20	82.34	2.46	18.20	78.06	3.74
D	0.12	3.94	95.94	0.14	5.58	94.28

5. Conclusion

In this paper, we proposed a method based on the short-time Fourier transform and wavelet packet transform combined with neural network (MLP) to classify blue whale calls. Unlike existing methods based on matched filter or spectrogram correlation to realize this task, which require the adjustment of various parameters related to the synthetic kernel, the proposed method does not need the adjustment of any parameters.

As possible improvements, the combination of this classifier with the one based on the chirplet transform will be explored in the future works. Also, an automatic system that permits to efficiently detect and isolate vocalizations will be integrated into the classification system.

References

[1] M.A. McDonald, S.L. Mesnick, J.A. Hildebrand, Biogeographic characterisation of blue whale song worldwide: Using song to identify populations, J. Cetacean Res. Manage. 8 (2006) 55–65.
[2] A. Sirovic, J.A. Hildebrand, S.M. Wiggins, Blue and fin whale call source levels and propagation range in the southern ocean, J. Acoust. Soc. Am. 122 (2007) 1208–1215.
[3] W.A. Watkins, M.A. Daher, G.M. Reppucci, J.E. George, D.L. Martin, N.A. DiMarzio, D.P. Gannonm, Seasonality and distribution of whale calls in the north pacific, Oceanography 13 (2000) 62–67.
[4] D.K. Mellinger, C.W. Clark, Recognizing transient low-frequency whale sounds by spectrogram correlation, J. Acoust. Soc. Am. 107 (2) (2000) 3518–3529.
[5] X. Mouy, Détection et identification automatique en temps-réel des vocalises de rorqual bleu (*Balaenoptera musculus*) et de rorqual commun (*Balaenoptera physalus*) dans le Saint-Laurent, Master's thesis, Université du Québec à Rimouski, Rimouski, Qc, Canada, 2007.
[6] Y. Simard, N. Roy, C. Gervaise, Shipping noise and whales: World tallest ocean liner vs largest animal on earth, in: OCEANS'06, MTS/IEEE, Boston, MA, USA, 2006, pp. 1–6.
[7] D. Mellinger, C.W. Clark, Method for automatic detection of mysticete sounds, Mar. Fresh. Behav. Physio. 29 (1997) 63–181.
[8] K.M. Stafford, C.G. Fox, D.S. Clark, Long-range acoustic detection and localization of blue whale calls in the northeast pacific ocean, J. Acoust. Soc. Am. 104 (6) (1998) 3616–3625.
[9] D.K. Mellinger, A comparison of methods for detecting right whale calls, Canadian Acoustics 32 (2) (2004) 55–65.
[10] M. Bahoura, Y. Simard, Chirplet transform applied to simulated and real blue whale (*Balaenoptera musculus*) calls, in: A. Elmoataz, O. Lezoray, F. Nouboud, D. Mammass (Eds.), Image and Signal Processing, in: Lecture Notes in Comput. Sci., vol. 5099, 2008, pp. 296–303.
[11] Y. Simard, M. Bahoura, N. Roy, Acoustic detection and localization of whales in Bay of Fundy and St. Lawrence estuary critical habitats, Canadian Acoustics 32 (2) (June 2004) 107–116.
[12] Y. Simard, N. Roy, Detection and localization of blue and fin whales from large-aperture autonomous hydrophone arrays: A case study from the St. Lawrence estuary, Canadian Acoustics 36 (1) (2008) 104–110.
[13] C.L. Berchok, D.L. Bradley, T.B. Gabrielson, St. Lawrence blue whale vocalizations revisited: Characterization of calls detected from 1998 to 2001, J. Acoust. Soc. Am. 120 (2006) 2340–2354.
[14] M. Bahoura, Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes, Comput. Biol. Med. 39 (9) (2009) 824–843.
[15] R.R. Coifman, M.V. Wickerhauser, Best-adapted wave packet bases, Preprint, Yale University, 1990.
[16] G. Cybenko, Approximation by superpositions of a sigmoidal function, Math. Control Signals Syst. 2 (4) (1989) 303–314.
[17] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, Neural Netw. 2 (5) (1989) 359–366.
[18] Q. Chen, Y.W. Chan, K. Worden, Structural fault diagnosis and isolation using neural networks based on response-only data, Comput. Struct. 81 (22–23) (2003) 2165.
[19] D. Cherubini, A. Fanni, A. Montisci, P. Testoni, A fast algorithm for inversion of MLP networks in design problems, COMPEL – Int. J. Comput. Math. Electr. Electron. Eng. 24 (3) (2005) 906–920.
[20] S. Haykin, Neural Networks: A Comprehensive Foundation, 2nd edition, Prentice Hall, Upper Saddle River, NJ, 1999.
[21] M. Beale, H. Demuth, M. Hagan, Neural Network Toolbox User's Guide, The MathWorks, Natick, MA, 2007.
[22] WaveLab 850, Stanford Statistics Department, Stanford University, http://www-stat.stanford.edu/~wavelab/, 2009.
[23] Adobe audition, http://www.adobe.com/products/audition/.

Mohammed Bahoura received the B.Sc. degree in Electronics from the University of Science and Technology of Algiers, Algeria, in 1990, the M.Sc. degree (D.E.A.) in Instrumentation and Control from the University of Rouen, France, in 1994 and the Ph.D. in Electrical Engineering, from the same University, in 1999. He was a Postdoctoral Fellow at the University of Quebec at Chicoutimi, Canada, from 1999 to 2001. Since 2001, he has been with the University of Quebec at Rimouski, where he is a Professor of Electrical Engineering. His research interests focus on digital signal processing and pattern recognition.

Yvan Simard completed a Ph.D. in biological oceanography at Laval University, Qc, in 1985. After a Postdoct in ocean ecology at the Institute of Ocean Sciences, BC, he joined Fisheries and Oceans Canada (DFO) at the Maurice Lamontagne Institute, Qc, in 1986, where he is conducting research in oceanography and hydroacoustics applied to ecosystem functioning and marine mammal studies. Since 2001, he is directing the DFO research chair in underwater acoustics applied to ecosystem and marine mammal research at the Marine Sciences Institute, University of Quebec at Rimouski (ISMER-UQAR). The author wishes to thank his research sponsors: DFO, NSERC, FQRNT, and ISMER-UQAR.