

A novel image-processing based method for the automatic detection, extraction and characterization of marine mammal tonal calls

ANTONIO SÁNCHEZ-GARCÍA¹, PATRICIO MUÑOZ-ESPARZA¹ AND JOSÉ LUIS SANCHO-GOMEZ²

¹Sociedad Anónima de Electrónica Submarina (SAES), Carretera de la Algameca s/n, 30205-Cartagena (Spain), ²Universidad Politécnica de Cartagena, Departamento de Tecnologías de la Información y las Comunicaciones, Campus Muralla del Mar s/n, 30202-Cartagena (Spain)

A novel, automatic method for the detection, extraction and characterization of marine mammal tonal calls is presented. Signals are automatically detected from the spectrogram, isolated using region-based segmentation, extracted and finally characterized by means of a fixed number of radial basis function (RBF) coefficients. A total of sixteen RBF coefficients are sufficient to accurately capture the time–frequency information contained in the calls. These coefficients can be later used to classify signals based on their characteristics. New specific functions for contour extraction and cross-resolution have been developed. The performance of the method has been extensively tested using simulated signals and a set of recordings covering a significant range of situations that can be encountered at sea.

Keywords: detection, extraction, characterization, marine mammal call, region-based segmentation, radial basis functions, contour extraction, cross resolution

Submitted 8 September 2008; accepted 26 March 2009; first published online 9 July 2009

INTRODUCTION

Marine mammal monitoring by a variety of techniques is becoming increasingly important for the detection, localization and identification of specific species of marine mammals. It permits the compilation of information that helps researchers to learn about marine mammal habits and communication behaviour. Additionally, knowledge of the presence of marine mammals in a certain area facilitates the protection of these animals by preventing harmful human activities, which can potentially disturb or damage these animals (Simmonds & Lopez-Jurado, 1991; Richardson *et al.*, 1995; Gordon *et al.*, 2004). Mitigation measures are being progressively incorporated in international regulations such as the US Marine Mammal Protection Act (MMPA) and the EU Habitat Directive (Pavan, 2007).

Different techniques are currently being used to monitor the presence of marine mammals in a specific location. The classic approach is to use visual surveys to spot marine mammals at the surface. The effectiveness of this method is limited: apart from the necessity that animals remain at the surface, it is also strongly dependent on weather conditions and the state of the sea, and it only can be performed in daylight hours. Acoustic monitoring appears to be an effective complement to visual surveys when marine mammals dive underwater (Mellinger & Barlow, 2003). This technique

overcomes the restrictions of the visual method and is effective over a significant detection range. It is restricted to the detection of vocalizing animals, but vocalization is an inherent characteristic of marine mammals that use acoustics for communication, navigation, socializing and echolocation (Richardson *et al.*, 1995). Techniques based on radar, infrared light and thermal imagery are presently under research for improved detection of marine mammals at the surface (Theriault, 2005; Pavan, 2007).

The existence of large quantities of data together with the need for accurate analysis to be able to recognize marine mammal sounds makes automatic call detection a very attractive option. Furthermore, automatic systems are less subjective than humans.

Marine mammals are vocalizing animals that inhabit all the seas and oceans of the world. Their vocal repertoire is broadly composed of three categories: echolocation clicks, burst-pulse sounds and whistles (Popper, 1980). Echolocation clicks are short-duration, broadband sounds that range in frequency between a few Hz and over 150 kHz (Au, 1993). They are used for prey localization and navigation and are usually emitted in trains of up to several hundred clicks per second. When the inter-click interval is so short that individual clicks cannot be distinguished by humans and acquire a tonal quality they are called burst-pulse sounds. It is thought that these signals are mainly used for communication. Whistles are narrowband frequency modulated signals with durations between a few milliseconds and several seconds (Tyack & Clark, 2000; Bazua-Duran & Au, 2002). They range in frequency mostly between 2 and 30 kHz (Lammers

Corresponding author:

A. Sánchez-García

Email: a.sanchez@electronica-submarina.com

et al., 2003; Oswald *et al.*, 2004). They are believed to be associated with social communication and it has been suggested that they carry information that allows individuals to identify themselves to one another (Caldwell *et al.*, 1990; Janik & Slater, 1998).

Whistles are the vocalizations of interest for this study. The performance of the method was tested on dolphin whistle recordings. Nevertheless, the generic characteristics of the method permit its application to the rest of marine mammal tonal sounds.

A pattern recognition system suitable for automatically processing sounds can be generically split into three stages: pre-processing, feature extraction and classification. In the pre-processing stage the calls of interest are detected and extracted from the background noise in which they are embedded. The feature extraction stage reduces the dimensionality of the problem by extracting a set of features that characterize the call. These features are used as the input elements for the classification stage.

Different approaches have been presented in the literature to completely or partially perform the stages of this recognition system. They have been specifically applied to cricket and frog calls (Brandes *et al.*, 2004), snapping shrimps (Learned & Willisky, 1995), bird vocalizations (Chen & Maher, 2006; Trifa *et al.*, 2008) and marine mammal vocalizations (Buck & Tyack, 1993; Janik & Slater, 1998; Mellinger & Clark, 2000; Oswald *et al.*, 2004; Brown & Miller, 2006). In the vast majority of the cases the spectrogram is the input data format of reference on which to apply the detection process (Janik, 1999; Norris *et al.*, 1999; Brandes *et al.*, 2004; Halkias & Ellis, 2006). Other studies are based on the wavelet transform (Learned & Willisky, 1995; Ghosh *et al.*, 1992) and the Hilbert–Huang Transform (Adam, 2006).

The extraction of the tonal signals is sometimes carried out through visual inspections by a skilled operator. In other cases this task is performed in an automatic or semi-automatic way, usually based on following the ridges of the spectrogram peaks. This provides the instantaneous time–frequency evolution of the call or frequency contour. This frequency contour is later used to obtain a group of features characterizing the call (Datta & Sturtivant, 2002; Gillespie, 2004; Brandes *et al.*, 2004; Oswald *et al.*, 2004, 2007). Other approaches for the feature extraction process are based on image processing techniques (Van Ijsselmuide & Beerens, 2004; Sánchez-García *et al.*, 2008), autocorrelation of frequency spectra (Deecke *et al.*, 1999; Searby & Jouventin, 2005), selected and grouped FFT frames from a histogram of the band-limited energy in these frames (Rickwood & Taylor, 2008), Mel-Frequency Cepstral coefficients (MFCC) and linear prediction coefficients (LPC) (Trifa *et al.*, 2008), short-time measurements of duty cycle and peak frequency (Murray *et al.*, 1998) and Cepstral feature vectors (Roch *et al.*, 2007).

A wide variety of classification methods have been proposed in the literature. Among them neural networks is one of the most used techniques (Learned & Willisky, 1995; Murray *et al.*, 1998; Deecke *et al.*, 1999; Deecke & Janik, 2006), due to its generalization capability and its high approximation ability. Other widely used techniques are multivariate discriminant analysis (Gillespie, 2004; Oswald *et al.*, 2004, 2007) and hidden Markov models (HMM) (Datta & Sturtivant, 2002; Rickwood & Taylor, 2008; Trifa *et al.*, 2008). Classification based on a Bayesian classifier (Brandes *et al.*, 2004), Gaussian mixture models (GMM) (Datta &

Sturtivant, 2002) and cluster analysis (McCowan & Reiss, 2001; Van Ijsselmuide & Beerens, 2004) have also been reported.

The present study covers the two first stages, pre-processing and feature extraction, of the recognition system by combining three important research fields: digital signal processing and digital image processing in the pre-processing stage and neural networks in the feature extraction stage. The method performs directly from the input of the waveform signals up to obtaining the feature vector characterizing each detected call. It additionally undertakes the challenging task of handling overlapping calls. The classification stage is not addressed in the present work. This stage would permit, apart from assigning individual whistles to specific species or related group of signals, to discriminate the whistles from other underwater signals with similar characteristics as are the sonar transmissions.

The performance of the method is evaluated by means of simulated signals and a set of recordings, totalling four hours, covering a significant spectrum of combinations of calls, noise and reverberation.

The following sections of the paper are distributed as follows: initially the method is described in detail through its breakdown in processing blocks. Next, the results of its application to simulated waveforms and recorded waveforms are presented and discussed. The paper finishes with the conclusions and a discussion of planned future work.

DESCRIPTION OF THE ALGORITHM

Method description

The spectrogram is a two-dimensional matrix representing sound intensity as a function of frequency and time. This fact allows for the possibility to apply the wide set of already available image processing functions to the spectrogram output, or to invent new ones specifically designed for the particular problem at hand. Neural networks processing is also an effective and well-proven technique for the characterization and classification of data sets that can be separated in regions. In this paper, both image processing and neural networks are integrated with signal processing to build a system addressed to automatically detect, extract and characterize marine mammal calls. The main processing blocks of the proposed system are presented in Figure 1.

A detailed description of each block in Figure 1 follows:

- **Spectrogram.** The input waveform was sampled at 44.1 kHz, thus the analysis band extends up to 22 kHz, which covers the range of fundamental frequencies of the dolphin whistle in the majority of cases. The FFT is applied to blocks of 512 elements with a 50% overlap (256 new samples and 256 overlapped ones). These parameters result in a frequency resolution of 86.1 Hz and a time resolution of 5.8 ms. Results of the present study show that these parameters are effective for dealing with dolphin whistles, which are characterized by a smooth frequency modulation.
- **Normalization and image adaptation.** The normalization process is applied to each element, $x_{i,j}$ (frequency–time cell) of the spectrogram matrix. An observation window

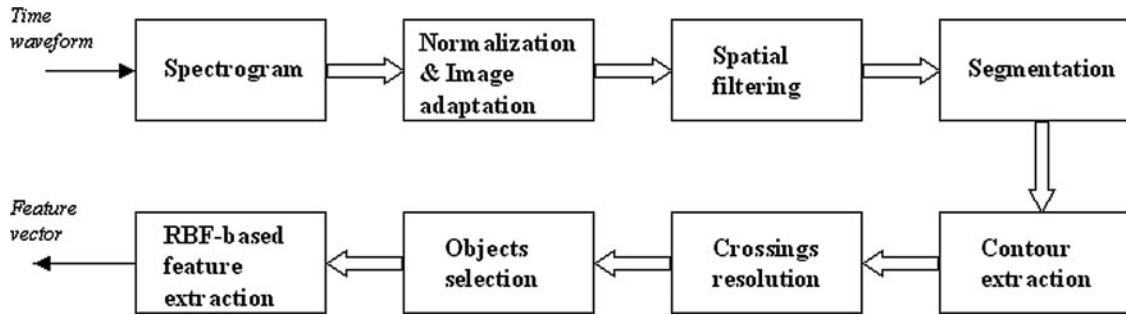


Fig. 1. Block diagram of the proposed method for detection, extraction and characterization of marine mammal calls.

centred on each spectrogram cell is used to statistically estimate the background noise level for this cell. After a comparative study between different normalization methods and different observation windows, based in a trade-off between performance, complexity and execution time, a trimmed-mean normalizer operating on a window of 32 frequency bins has been selected. The background noise estimation, $\mu_{i,j}$ for each window is obtained through a two-step process. First, the 32 elements in the window are sorted and the elements between the percentiles 0% and 90% are selected. Then the mean value of the vector of sorted elements is computed. This value is used as the background noise estimation of the spectrogram cell. The normalized value, $v_{i,j}$ for each element of the spectrogram matrix is finally obtained by:

$$v_{i,j} = \frac{x_{i,j} - \mu_{i,j}}{\mu_{i,j}} \quad (1)$$

We call the matrix resulting from the normalization process the *normalized matrix*.

This matrix can be processed as an image, with n elements in the x-axis and m elements in the y-axis containing signal-to-noise ratio (SNR) (rather than absolute intensity) values as a function of time and frequency. From this point we will work with an $n \times m$ dimension image with its elements ranging from 0–255 being referred to as pixels. This image is named *initial image* and is built from the following premises:

- Clipping is performed on elements higher than 255 as this value corresponds to a significantly high SNR.
- A noise reference level is established. It is assumed that values higher than this reference level correspond to valid signals (whistles) and values lower than it correspond to noise. After trials with different methods, the best results come from computing this level by sorting all the elements in the image and selecting the 70% percentile of the resulting vector. The values in the image lower than this reference level are set to 0.
- The elements in the image corresponding to frequency values lower than 800 Hz are set to 0. This is due to the high level of noise in low frequencies. This decision has a very limited impact on the detection process since the vast majority of the dolphin whistles have fundamental frequencies above this value.

An example of the outputs of the spectrogram as well as normalization and image adaptation stages when applied to a simulated image is presented in Figure 2.

- **Spatial filtering.** The application of a Gaussian spatial filter to the initial image has proven to be effective to mitigate the splitting of whistles in several pieces as a consequence of SNR drops. This filter produces a smoothing effect on the image, which is effective for connecting call fragments. If the smoothing kernel is too large it also has the detrimental effect of reducing the image resolution. In this method, after trials with different configurations, a 3×3 smoothed Gaussian kernel with standard deviation of 0.5 has been selected as the best option. This kernel is convolved with each element in the image. The values, $K(m, n)$, of the Gaussian kernel depend on the standard deviation

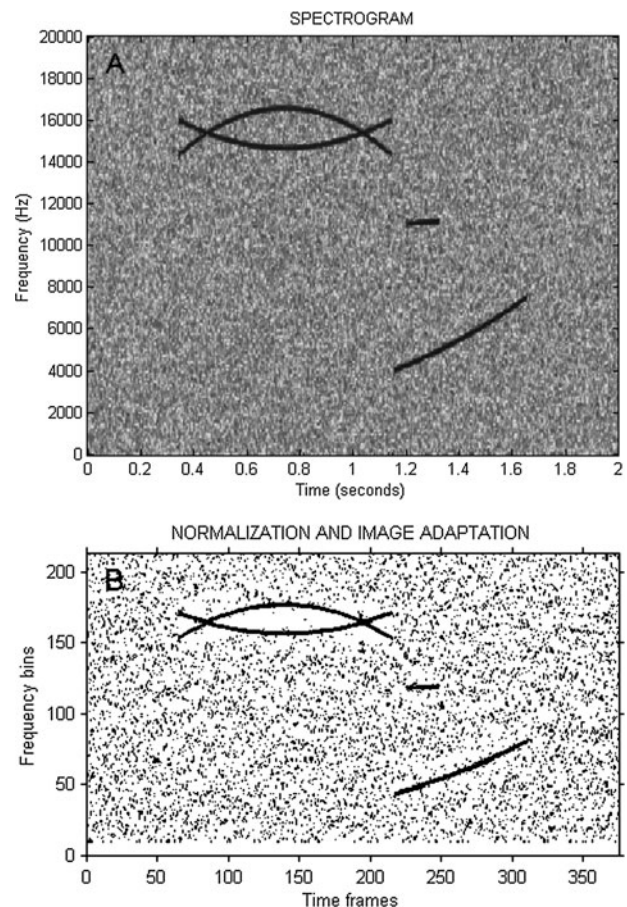


Fig. 2. Example of application of the two first stages of the proposed method to a simulated image containing three whistle-like objects and one spurious small duration signal. (A) spectrogram stage output (B) image adaptation stage output.

selected value in the following way:

$$K(m, n) = ce^{-\frac{(m^2+n^2)}{2\sigma^2}} \quad m, n: 0, 1, \text{ and } 2 \quad (2)$$

where m and n refer to the kernel row and column, taking $m = n = 0$ to be the centre of the kernel. The constant c is computed so that the sum of all the elements in the kernel is unity. For a standard deviation of 0.5, c takes the value of 0.1621. The resulting image is named *filtered image*.

An example of the effect of the application of the Gaussian filter on a synthetic image is presented in Figure 3.

- **Segmentation.** Image segmentation aims to subdivide an image into its constituent regions. Image segmentation algorithms are based on two basic properties of intensity images: discontinuity and similarity. In the former, the approach to partitioning an image is to search for abrupt changes in intensity, because these correspond to edges. In the latter, the approach is to find similar regions according to a set of predefined criteria. For this reason this approach is also known

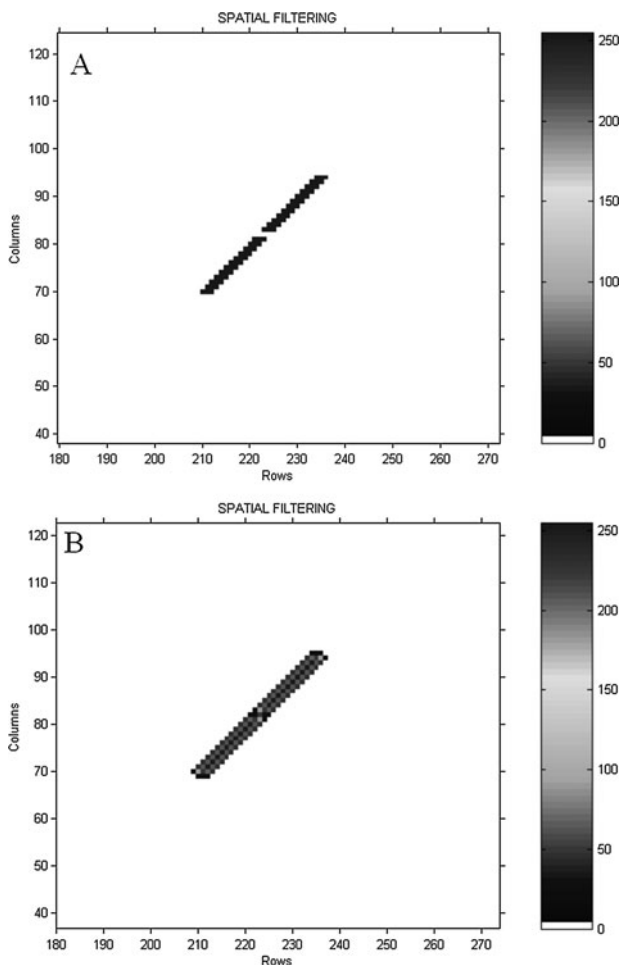


Fig. 3. Effect of the application of a Gaussian spatial filter to a synthetic image. (A) Chirp-like signal split into two pieces; (B) output after the application of a 3×3 spatial Gaussian filter with standard deviation of 0.5. It can be observed as the two separated pieces are joined. A blurring effect is also appreciated in the image.

as region-based segmentation. Some examples of region-based segmentation include: thresholding, region growing and splitting and merging (Gonzalez *et al.*, 2004). In noisy images, this method generally performs better than those based on edges (methods based on discontinuities), where borders are very difficult to detect.

In this study, we work with spectrograms, which can be considered noisy images. Thus a region-based segmentation technique, in particular a region growing procedure, has been selected. Region growing is a procedure that groups pixels or subregions into larger regions based on predefined criteria for growth. The basic approach starts with a set of seed points, and from these points regions grow by appending to each seed neighbouring pixels that have predefined properties similar to the seed (Gonzalez & Woods, 2001). This growing process continues until a stopping criterion is satisfied. The main considerations in region growing are the selection of a set of one or more starting points, the selection of the similarity criteria to add new pixels to the region based on intensity levels or spatial properties, and the formulation of a stopping rule that determines when the growing process has been completed for each region (no more pixels satisfy the similarity criteria).

The proposed region growing algorithm is an adaptation of the method presented in Gonzalez *et al.* (2004) which is based on morphological operations. Separate regions are initially built based on morphological reconstruction. This consists of a morphological transformation involving two images: the marker, which is the starting point for the transformation and is composed of seed points, and the mask, that restricts the transformation and defines the limits of the growing process. Thus, every seed pixel in the marker image grows according to the connectivity criterion evaluated in the neighbourhood around the corresponding pixel in the mask image. If we denote k as the marker image, s as the mask image, h_i as an image of the same dimension as k and s and C is a 3×3 matrix defining the connectivity; the reconstruction of s from k , called $R_s(k)$, is performed through the following iterative process:

- Initialize h_1 to the marker image k .
- Create the matrix C composed of 3×3 ones in the case of 8-connectivity.
- Cycle: $h_{i+1} = (h_i \oplus C) \cap k$ until: $h_{i+1} = h_i$ where \oplus references the morphological operation opening.

In our particular case, the marker image is built by selecting the seeds as the elements in the filtered image that are higher than a fixed threshold. This threshold has been set to a value of 14 after some experimentation. When several markers are connected only one member of the group is selected. The marker image will contain 1 in all the cells where seed points are located and 0 elsewhere, becoming, therefore, a binary image.

The mask image will contain the target regions obtained from the filtered image. These regions are composed of the pixels with a value higher than the previously calculated noise reference level (see the

stage of the Normalization and Image Adaptation). The mask image will contain 1 for all the cells satisfying this criterion and 0 elsewhere, thereby, also becoming a binary image.

The reconstruction process will start from each point in the marker image and regions will be formed by appending the mask image pixels inside the 3×3 connectivity matrix (C) centred in the marker seed pixel. Once the reconstruction process is finished, one-valued pixels separated by single-pixel gaps (zero-valued pixels) are linked in order to prevent the splitting of the whistles as much as possible.

Finally, all the connected components in the resultant binary image after the reconstruction and single-gap removal process are extracted and individually mapped to images with the same dimension as the binary image. From this point, we will work on a set of images containing separate objects corresponding to candidate whistles. These images will be referred to as *segmented images* (see Figure 4 for an example).

- **Contour extraction.** Once the objects have been segmented and mapped to individual images, the next processing stage consists of extracting the frequency contours that correspond to the modulation pattern of the whistle frequency. The objects in the segmented image are thick representations of the whistles. Thus, the objective of the contour extraction is to transform this thick representation into a 2D curve representing the frequency modulation. The pixels in the frequency

contour are obtained from the blocks of contiguous one-valued pixels in each time frame (each column of the segmented image). For each block, the pixel with the maximum value in the initial image is selected. An alternative method consisting of choosing the median values has been also tested but it provides worse global results. Each individual image obtained from the contour extraction process is referred to as *contour image*. It is a binary image where the one-valued pixels reference the location of the extracted frequency contour.

Whistles of many delphinid species are low enough to show one or two harmonics in addition to the fundamental frequency in the 22 kHz bandwidth used for this analysis. The algorithm does not try to associate the possible harmonics to fundamental frequencies and therefore harmonics will be considered as separate whistles.

- **Crossings resolution.** The algorithm that isolates overlapping curves takes as input each contour image. The first step consists of checking if a contour image contains crossing curves or one isolated curve. If there is any column with at least two one-valued pixels, the contour image is considered to contain crossing curves. In the opposite case (every column of the contour image contains no more than one pixel with unity value) it is considered that the image contains one isolated curve. The following step consists of detecting the tentative crossing areas, and

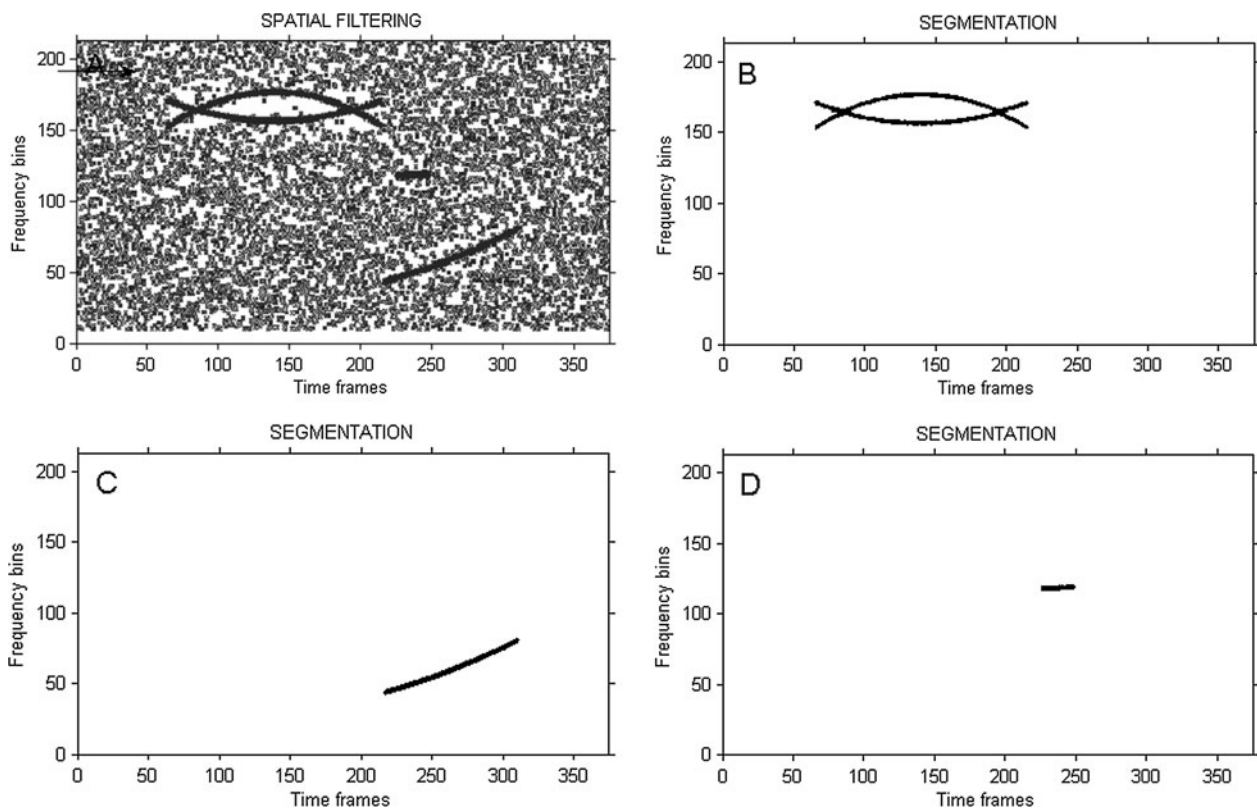


Fig. 4. Example of application of the spatial filtering and segmentation stages to the simulated image shown in Figure 2. (A) Output of the spatial filtering stage that is the input to the segmentation stage; (B–D) the region-growing based segmentation stage provides as outputs three images of the same size as the input image, containing one object each.

facilitating the correct separation of the curves taking part of the crossing. For doing that, the columns from the corresponding *segmented image* (not the contour image) containing a number of frequency bins higher than a predefined threshold (highest density of one-valued pixels) are identified and the corresponding frequency bins in the contour image are set to zero. As a result, the tentative crossing region of two overlapping curves in the input contour image is eliminated (set to zero), leaving the original curves split into segments that will have to be connected properly. After trials with different values, this threshold has been set to 5.

Next, the segments obtained previously are extracted and mapped to individual images. Two auxiliary images are used in this process: image A which registers, as a one-valued pixel, each new pixel identified as part of the curve, and image B which contains all the one-valued pixels from the contour image that have not already been assigned to a segment.

The extraction of the first segment starts from the uppermost left pixel in the auxiliary image B. From this initial point a tracking process is performed that aims to build the curve by adding points from the rest of columns until the ends of segment condition is satisfied. This condition is met in the following cases:

- a) The next two columns do not contain any one-valued pixel.
- b) The Euclidean distance between the candidate pixel and the last pixel registered as part of the segment has to be lower or equal to a predetermined constant. By experimentation a value of 7 has been assigned to this constant.

The tracking algorithm starts by counting the number of one-valued pixels in the next column. If there is not one the next column is checked. If there is only one this will become the candidate pixel that will be added as a new point in the segment if none of the ends of segment conditions are met. When at least two one-valued pixels are found in the following column, the Euclidean distance between each point in the column and the last point registered in the segment is computed and the distance is sorted in a vector. The candidate pixel will be selected based on the difference

between the two lower values in the sorted vector. If this difference is high (equal or higher than a predefined constant, a value of 6 has been chosen by experimentation), a criterion of lower distance is applied and the first pixel in the sorted vector is considered to be the candidate pixel. If the difference is low (lower than the constant) it is considered that these two pixels are close to or making part of a crossing between segments and the candidate pixel is chosen based on the trend. This trend is computed as the ratio between the increase in frequency bins (rows) and the increase in time frames (columns). The trend value of both pixels is computed and the one closer to the trend value of the last point registered in the segment is selected as the candidate pixel. This candidate pixel is added to the segment if none of the end of segment conditions is met. This way of operating has shown to be effective when dealing with the most complex cases of overlapping curves. An example of the process of segments extraction is presented in Figure 5.

Once a segment has been extracted, if there are more than four points left (the minimum length considered for a whistle is of five pixels) in the auxiliary image B the previously described process starts again from its uppermost left pixel. The process of extracting segments finishes when only four or fewer pixels remain unassigned in the auxiliary image B.

The next step consists of linking segments to obtain the individual curves present in a crossing. For this process, criteria of proximity, trend and intensity are evaluated. A rectangular searching area of dimensions 39 (rows) \times 30 (columns) starting on the last point of each segment, which corresponds to the pixel with the highest value of the column, is obtained. This point is referred to as the reference point. If (x, y) denote the coordinates of this point, then the rectangle extends from $x - 19$ to $x + 19$ in rows and from y to $y + 30$ in columns. Inside this searching area, up to the three closest segments to the reference point, in terms of Euclidean distance, are selected. These segments are referred to as candidate segments. Next, a weighting scheme with scores for proximity, trend and intensity is performed. The proximity corresponds to the previously obtained Euclidean distance. The

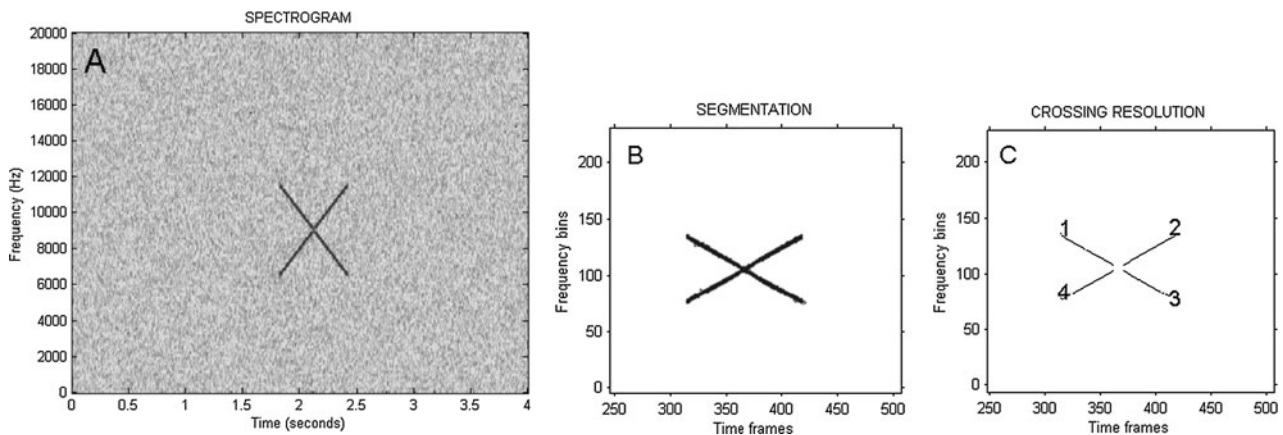


Fig. 5. Example of the process of segments extraction in the crossings resolution stage for a simulated waveform composed of two overlapping linear chirps. (A) Spectrogram of the waveform; (B) output of the segmentation stage; (C) splitting in segments of the object from the contour extraction stage where the columns with highest density of pixels in the image from the segmentation stage are set to 0.

trend is computed as the average of the set of local trends. This local trend, computed as defined previously, is determined for each group of five consecutive points in the curve. The intensity is computed as the average of the intensity of all points along the segment. If the obtained score surpasses an established threshold, the candidate segment is selected to be linked with the current segment. The values of the above parameters have been selected after trials with different combinations of parameter values to get the best results. An example of the process of segments linking is shown in Figure 6.

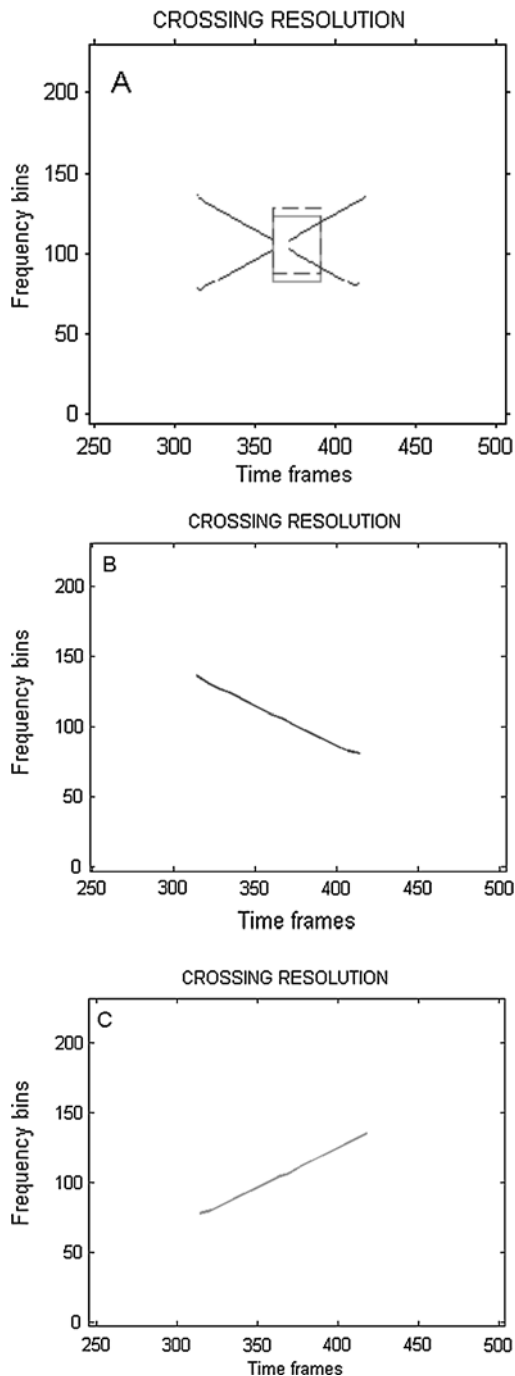


Fig. 6. Example of linking of segments in the crossings resolution stage. (A) Setting of the rectangular searching areas; (B–C) curves resultant of the linking process based on criteria of proximity, trend and intensity.

Figure 7 shows the output of the crossing resolution stage when applied to a simulated image.

- **Objects selection.** The individual curves resulting from the crossing resolution stage are taken as the candidate whistles. In order to be considered as whistles they have to meet the established criteria for a minimum number of columns (times frames) and minimum number of rows (frequency bins). These criteria help us to avoid selecting short-duration broadband signals as whistles, as is the case of the clicks, or constant-frequency noises. The drawback of this selection is that whistles with very high or very low frequency modulation could be discarded, but these types of modulation have been rarely appreciated in the available recordings. After experimentation a constant value of 5 has been selected for both the minimum number of columns and the minimum number of rows.

Figure 8 shows the output of the objects selection stage when applied to a simulated image.

- **Radial basis function (RBF)-based feature extraction.** The aim of this processing step is to characterize the whistles by means of a reduced set of parameters, instead of the whole set of time–frequency cells that form the frequency contour of each whistle. This characterization process is especially valuable for a further classification process, which requires that these parameters capture enough information to permit reaching an accurate classification level. In our case, once the whistle selection process is finished, each whistle is registered in a separate binary image and its frequency contour defines a specific curve for this whistle. Radial basis functions neural networks are used to approximate this curve to obtain a set of parameters characterizing the whistle. These RBF networks have shown to be a very effective way of approximating a curve (Park & Sandberg, 1991, 1993). Thus, RBF neural networks are used as a feature extraction tool and, at the end of this processing step, each whistle will remain characterized by a set of 16 coefficients which convey significant information on the curve evolution.

A radial basis function is a real-valued function whose values depend on the distance from the input vector to its centre, μ . Radial basis functions are used to build up approximation functions, $F(x)$, of the form:

$$F(x) = \sum_{i=1}^N \omega_i \Phi(\|x - \mu_i\|) \quad (3)$$

where $F(x)$ is computed as the sum of N radial basis functions with a different centre μ_i and weighed by a coefficient ω_i .

The approximation function in (3) admits a parallel structure as a feed-forward neural network that is called the radial basis function (RBF) network. What is especially remarkable for our intent of characterizing whistles is that this network is able to approximate any smooth non-linear input–output mapping to an arbitrary degree of accuracy, provided that a sufficient

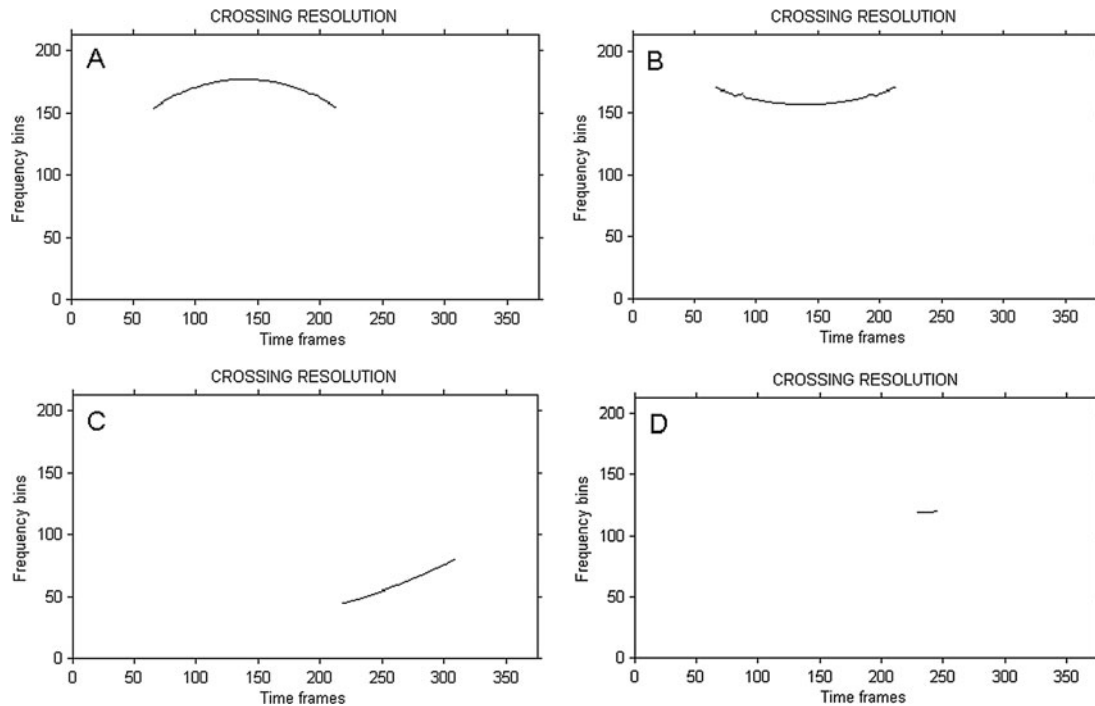


Fig. 7. Example of application of the crossing resolution stage to the simulated image shown in Figure 2. (A–D) Overlapping curves from the contour stage are isolated and mapped to individual images.

number of hidden layer neurons are used (Park & Sandberg, 1991, 1993). This is often referred to as the universal approximation theorem.

Radial basis function networks implement a fixed structure composed of three layers with entirely different roles (Haykin, 1999). The nodes of the input layer receive the data vectors entering the network. The second constitutes the only hidden layer of the network where each hidden unit implements a radially activated function, corresponding to a non-linear transformation from the input space to the output space. The third layer, which is linear, implements a weighted sum of the outputs of the hidden units. A schematic of this network is presented in Figure 9.

The most widely implemented radial basis activation function is the Gaussian kernel, defined as follows:

$$\Phi(\|x - \mu\|) = \exp\left(-\frac{\|x - \mu\|^2}{2\sigma^2}\right) \quad (4)$$

where μ is the centre of the Gaussian and σ^2 is the variance, i.e. a scale factor measuring the width of the Gaussian.

The argument of the activation function of each hidden unit in an RBF network computes the Euclidean norm (distance) between the input vector and the centre of that unit. Gaussian basis functions whose centres are closest to the input data will provide the largest output. Hence, RBF networks are able to model data locally.

Training an RBF network consists of calculating its unknown parameters. This means determining: (a) the number of radial basis functions that correspond to the number of hidden units; (b) the centres and widths

of each radial basis function; and (c) the weights of the output layer (Hu & Hwang, 2002). Different approaches have been proposed to determine these parameters. This method considers a fixed number of radial basis functions in order to further use the output layer weighting coefficients as the feature vector characterizing each whistle. After experimentation, 16 radial basis functions have proven to provide a precise fitting for the vast majority of the analysed whistles characterized by a smooth evolution. More precise fitting can be achieved through increasing the number of radial basis functions by increasing the feature vector size.

The centres are selected to be equally spaced along the segment of the x-axis between the beginning and the end of the curve (Sanner & Slotine, 1994; Matej & Lewitt, 1996). A 50% overlap between consecutive Gaussians is established. This determines the Gaussian width to be twice the distance between consecutive centres.

For the determination of the output layer weights the linear least square method is employed. If we let \mathbf{y} be the desired radial basis function network output vector, \mathbf{w} to the weighting vector and \mathbf{G} be the matrix containing the hidden layer nodes outputs, we will have in vector notation:

$$\mathbf{y} = \mathbf{G}\mathbf{w} \quad (5)$$

with: $\mathbf{y} = (y_1, \dots, y_n)^T$, $\mathbf{w} = (w_1, \dots, w_n)^T$, $\mathbf{G} = (g_{j1}, \dots, g_{jn})^T$ $j: 1, \dots, m$, $m = 16$, $n = 16$.

We can then obtain the vector \mathbf{w} by inverting the matrix \mathbf{G} :

$$\mathbf{w} = \mathbf{G}^{-1}\mathbf{y} \quad (6)$$

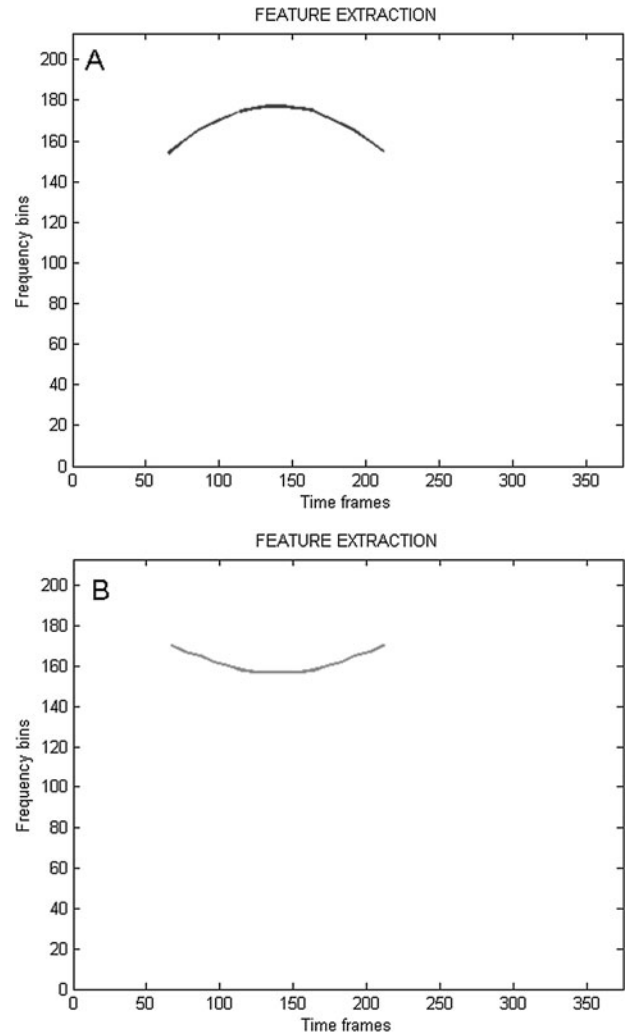
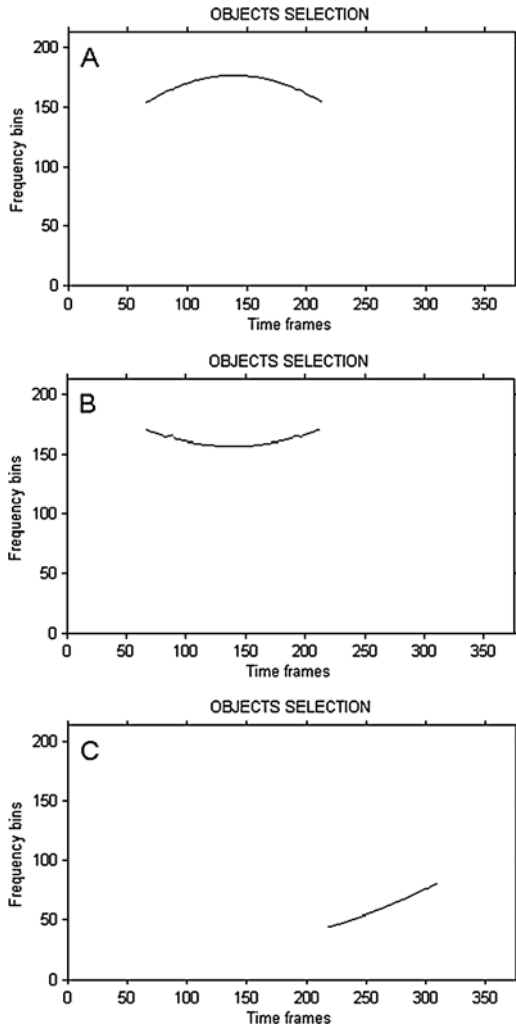


Fig. 8. Example of application of the objects selection stage to the simulated image shown in Figure 2. (A–C) It can be appreciated that the image containing the spurious-like signal has not been considered as a valid whistle.

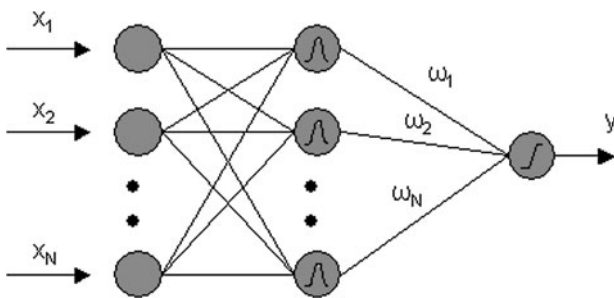


Fig. 9. Three-layer structure of a RBF neural network where x_1 to x_N reference the data vectors entering the nodes of the input layer, w_1 to w_N correspond to the weights applied to the outputs of the hidden layer nodes and y references the neural network output vector.

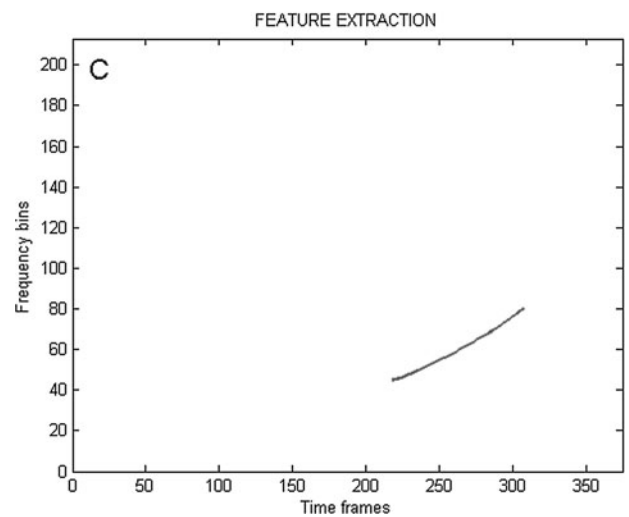


Fig. 10. Example of application of the feature extraction stage to the simulated image shown in Figure 2. (A–C) Each image is built linking the 16 neural network outputs.

To prevent possible singularities the pseudo-inverse matrix is calculated instead of the inverse matrix:

$$w = (G^T G)^{-1} G^T y \tag{7}$$

After calculating the weighting coefficients vector, w , the function approximating the whistle contour is obtained from (5).

Figure 10 presents an example of application of the feature extraction stage to a simulated image.

TEST OF THE ALGORITHM

In order to improve the efficiency of this method it has been initially tested with synthetic and simulated signals. This testing process has permitted us to refine the method by dealing with artificial environments of incremental difficulty. Items not providing the expected degree of accuracy were logged and an alternative solution was posed and evaluated with the goal of finding more accurate approaches. This same refinement process was applied when testing with recorded signals. In this latter case a wide range of sea conditions have been evaluated in order to gain an overview of the performance of the call detection method.

Application to synthetic and simulated signals

Synthetic signals have been built by direct assignation of values to the pixels in an image. In this way curves with specific shapes, widths and lengths have been initially implemented and used to test the efficiency of the method. This has permitted us to test particular situations including curves with several branches, curves with drops of level, and crossings between curves.

After finishing tests with synthetically generated signals, the next step in the processing chain refinement was based on building signals that are more similar to real signals. Simulated signals have been created using specific mathematical functions (linear sweep, logarithmic sweep, etc.) embedded in a background of noise. This way of operating has let us test all the stages in the processing chain and particularly the effect of the variation of the signal-to-noise ratio in the process of detection and characterization.

In order to have a graphical reference of the processing performed at each of the analysis stages, their outputs have been obtained from the processing of one waveform, four seconds long, containing a set of simulated signals embedded in a background of Gaussian noise. Apart from the noise, the waveform contains two crossing quadratic swept-frequency (chirp) signals, starting both at 1.2 s, in the high range of frequencies (above 14 kHz), one linear chirp, starting at 2.0 s, crossing with one quadratic chirp, starting at 1.8 s, in the low range of frequencies (below 8 kHz) and one logarithmic chirp, starting at 3.15 s, alone in the middle range of frequencies (centred on 11 kHz). Also two spurious small duration signals, starting at 1.0 s and 2.4 s, have been incorporated into the waveform. These outputs are presented in Figure 11.

From the analysis of the images in Figure 11, we can observe that the filtering stage results in image blurring (Figure 11 C). The segmentation stage provides continuous objects including crossing curves (Figure 11 D₂, D₃). The remaining stages work with individual objects, each in its own independent image. The contour extraction stage provides a total of five objects of one pixel width, which are separated into individual curves such that crossings are detected. A total of seven independent curves are extracted after solving the two crosses between curves. The object selection stage selects the curves matching the conditions based on minimum variation in time frames (pixels in the horizontal axis) and frequency bins (pixels in the vertical axis). A total of five curves are selected. Finally, each curve is approximated by means of a radial basis function network

and drawn linking these points. Each curve remains characterized by the set of sixteen weighting coefficients used to compute the RBF network output.

APPLICATION TO SIGNALS RECORDED AT SEA

In order to evaluate the effectiveness of the proposed method on signals recorded at sea, a total of four hours of recordings have been tested. Some of these recordings have been provided by CIBRA (Centro Interdisciplinare di Bioacustica e Ricerche Ambientali – Università di Pavia) which contain whistles from bottlenose dolphins, common dolphins and striped dolphins, and the rest come from the 3rd International Workshop on Detection and Classification of Marine Mammals (compiled by NUWC (NATO Undersea Warfare Centre) and WHOI (Woods Hole Oceanographic Institute)) which contain Risso's dolphin whistles.

The calls in these recordings present a wide range of variation in terms of signal-to-noise ratio, density of whistles, presence of interfering noise and overlapping between whistles. In order to have a reference of the main characteristics being evaluated, the recordings have been generically assigned to three categories depending on their degree of complexity: low, medium and high. Low complexity recordings are characterized by the presence of isolated whistles with medium to high SNR. Medium complexity recordings include a moderate number of crosses between whistles, and whistles with low, medium or high SNR. The remaining recordings contain high levels of overlap between whistles and/or presence of a significant number of whistles with very low SNR (normally accompanied by a high degree of discontinuity) and are categorized as having high complexity. An example of these categories is presented in the Figure 12.

Although the main focus of the present paper is to show a qualitative analysis of the effectiveness and performance of the proposed method when applied to signals recorded at sea, in order to have a quantitative insight, two kinds of measurements have been carried out on the available recordings. First, the number of false positives (FP) has been accounted for on recording fragments not containing whistles. In the second one the percentage of correctly extracted whistles and the percentage of correctly resolved crosses between whistles have been logged.

The number of FP is highly dependent on the criteria used, in the selection stage, for the minimum length and the minimum frequency modulation required for a contour to be classified as whistle. To make an estimate of this number, five minutes of recording fragments containing only noise have been selected. With the chosen value of five for both ranges, a total number of eight FP were counted. If four had been used instead of five, the number of FP would have increased to twenty-three. Conversely a range value of six for bins and time frames would have reduced the number to four. Generically, higher values for these range values prevent the selection of valid whistles from an aggregation of noise because it excludes short duration signals and signals with very slow and very high frequency rates.

To optimize the percentage of correctly detected and extracted whistles, eight minutes of recording segments of low and medium complexity have been used for their

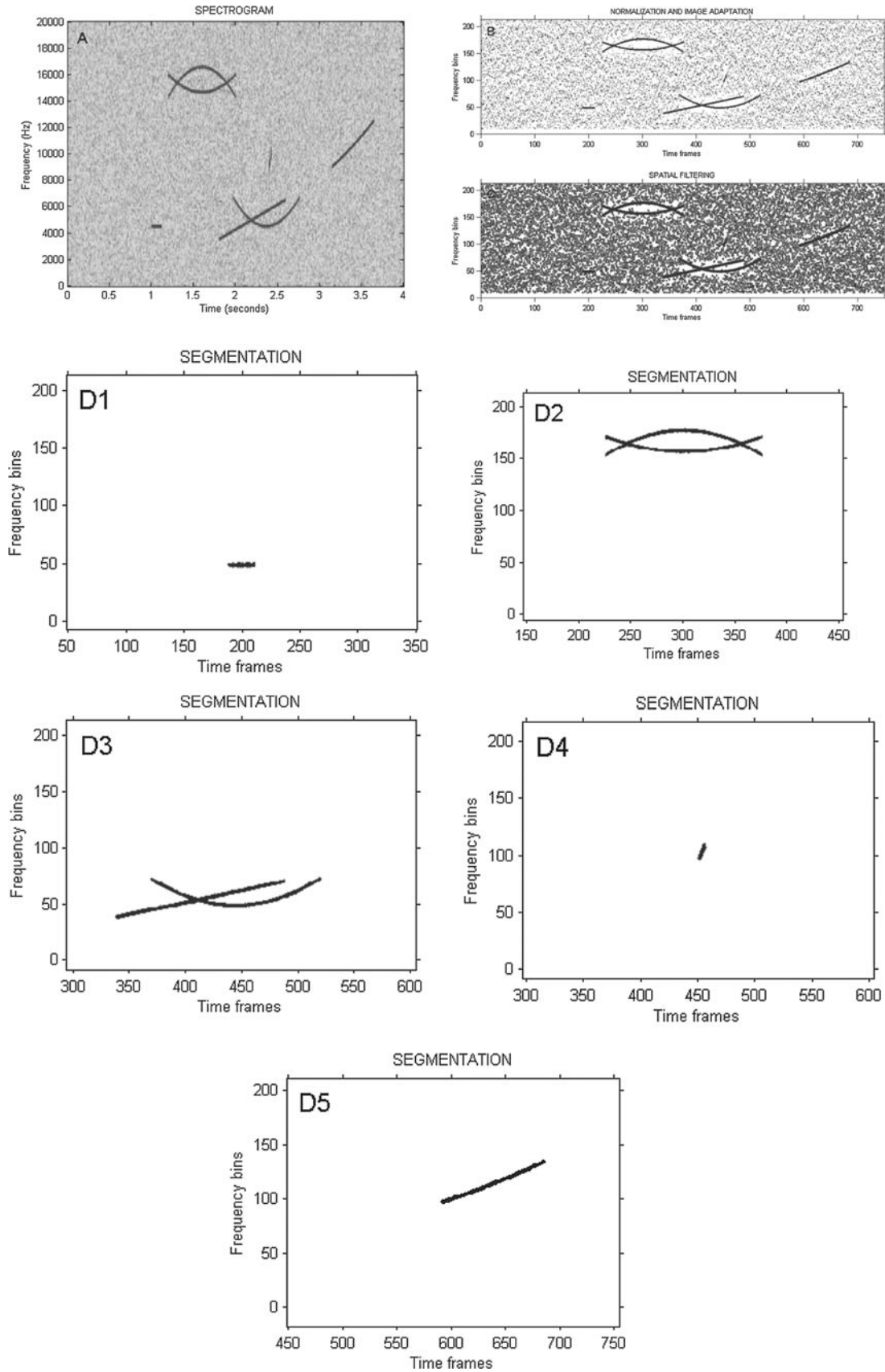


Fig. 11. Outputs for each stage in the proposed method for a simulated waveform containing five simulated modulated signals and two spurious small duration signals. (A) Spectrogram of the waveform; (B) output after normalization and image adaptation; (C) output after the application of the spatial filter; (D1–D5) output after region-growing based segmentation. From this stage the detected objects are individually managed; (E1–E5) output after contour extraction; (F1–F7) output after crossing resolution; (G1–G5) output after objects selection (some items have been discarded); (H1–H5) output after the feature extraction stage, where the curves have been built from the RBF neural network outputs and remain characterized by their vector of weighting coefficients.

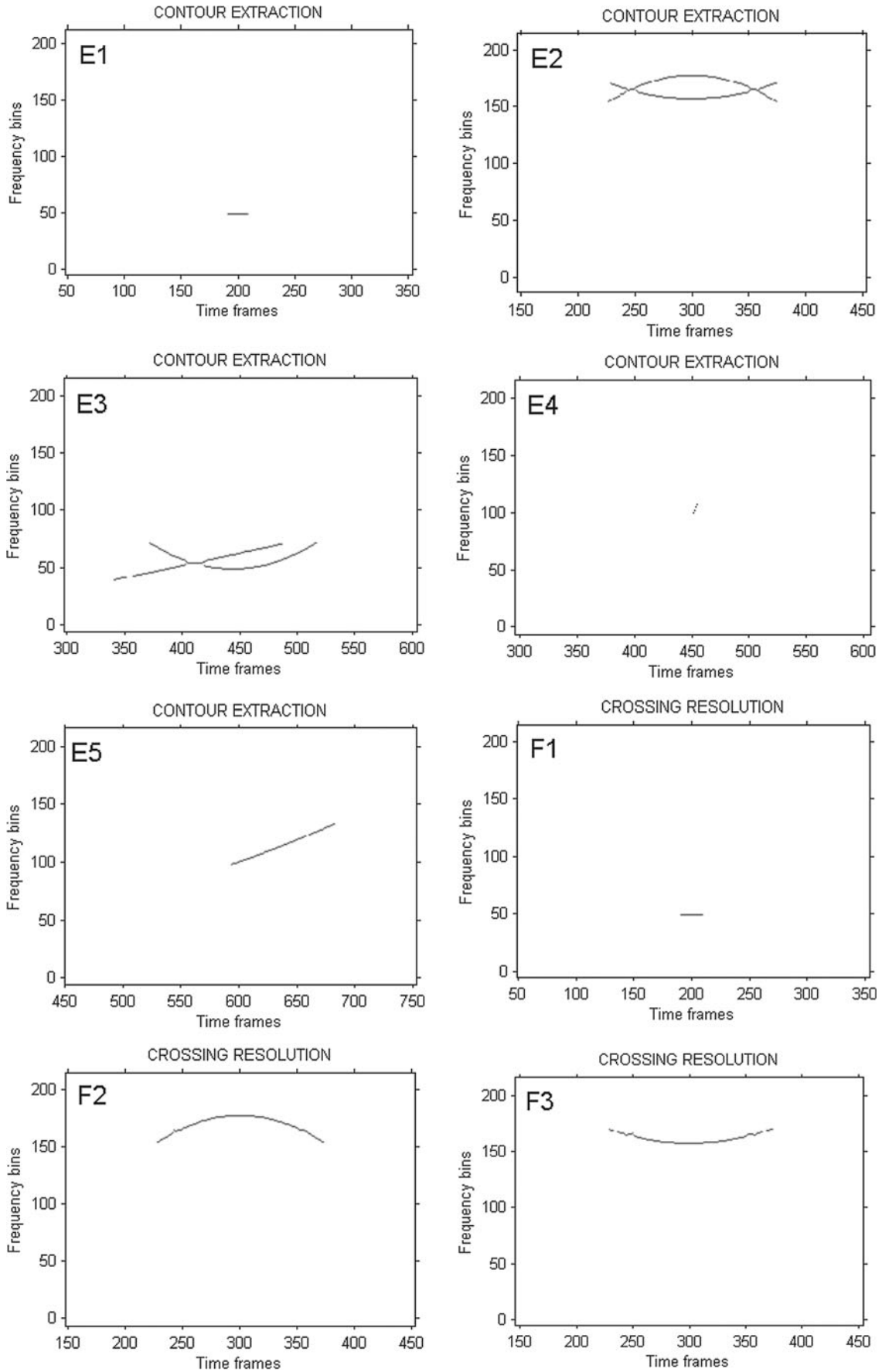


Fig. 11. Continued.

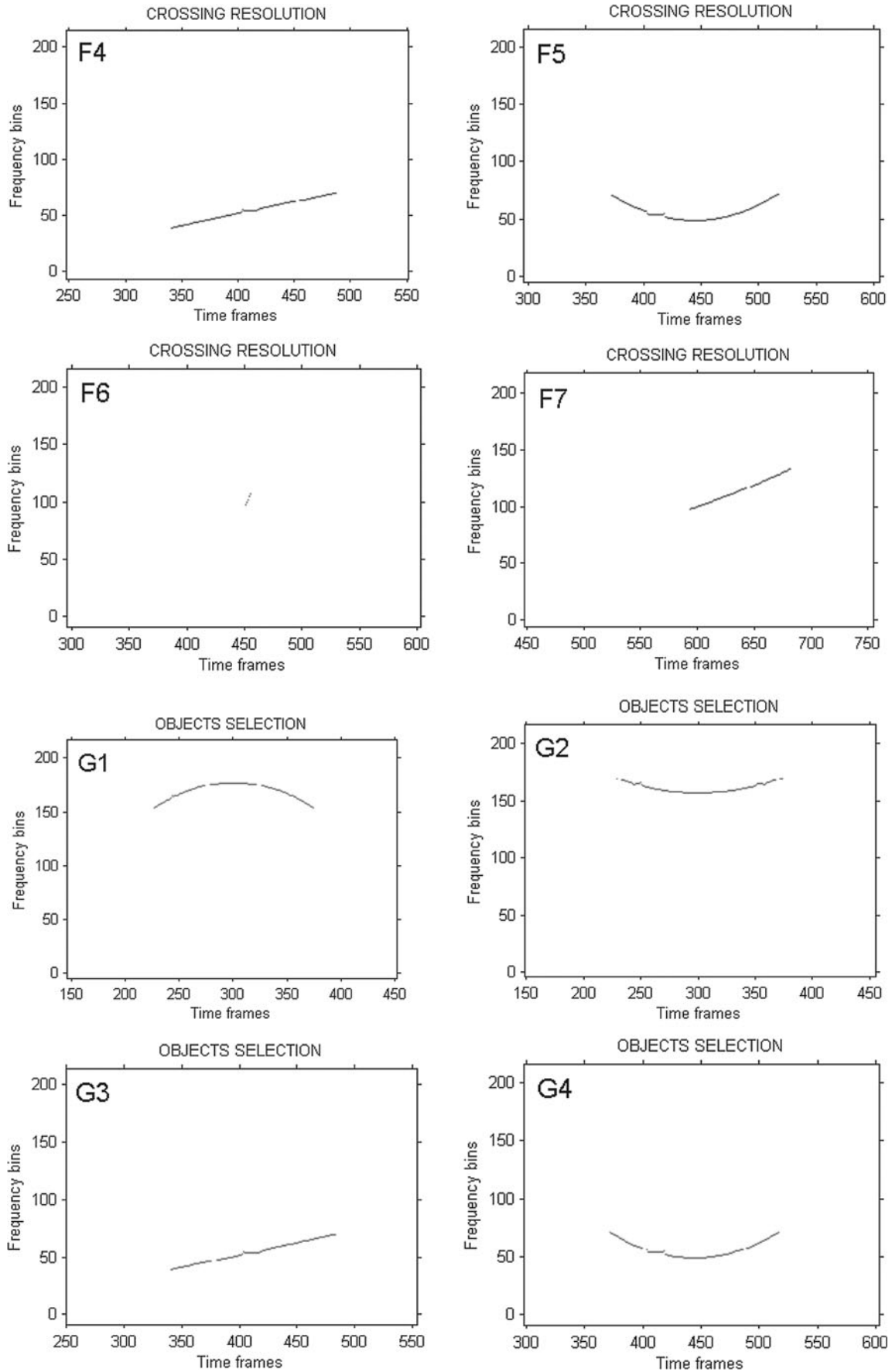


Fig. 11. Continued.

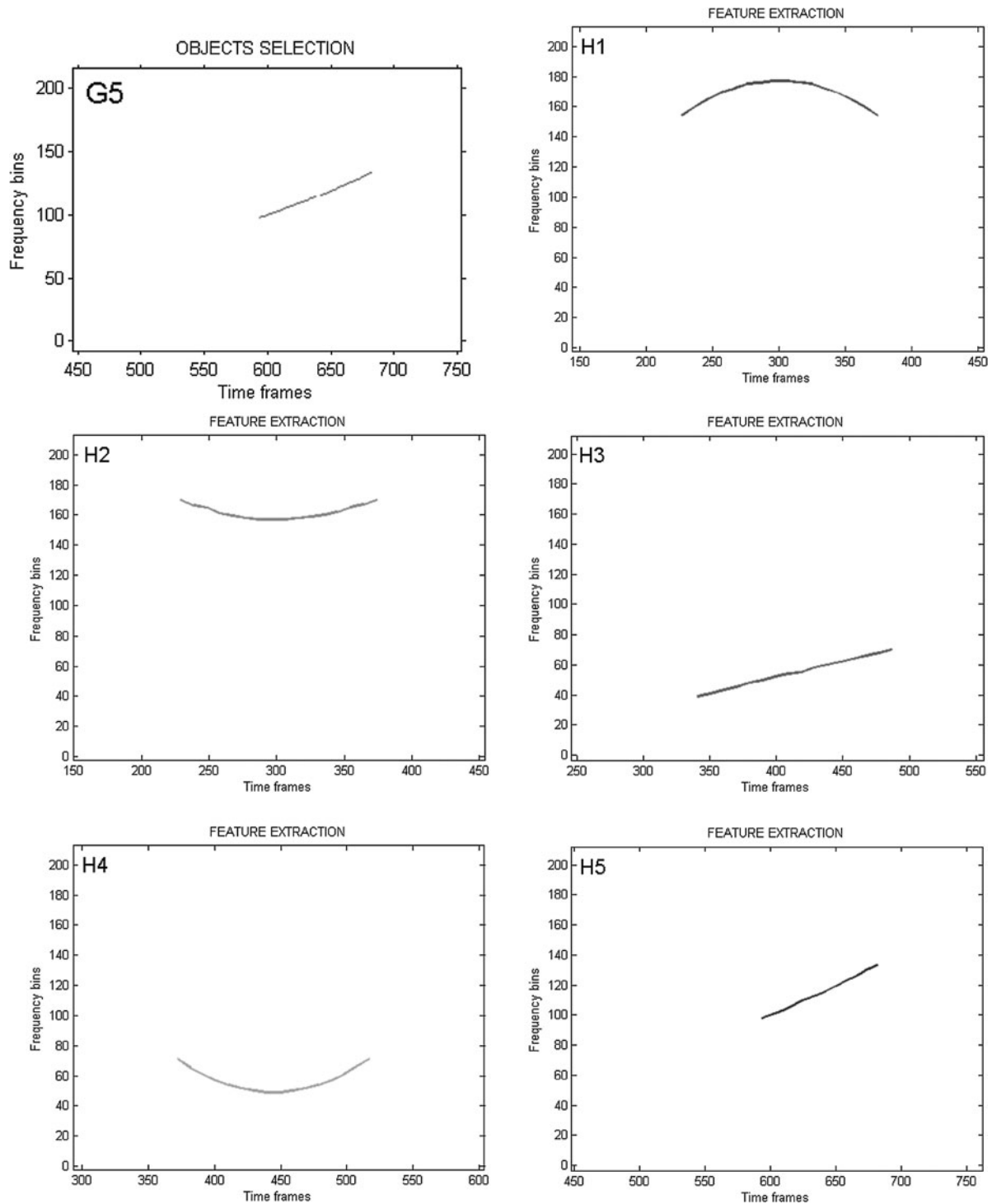


Fig. 11. Continued.

computation. The number of whistles present in the recording was manually logged by the first author and used as reference. A whistle was scored as correct when a significant part of it (at least 50%) has been correctly extracted and characterized. A total number of 737 whistles were manually identified in the analysed recordings. From these, the number of whistles considered correctly extracted totals 656, which correspond to a percentage of 89.0%. The undetected whistles correspond, for the most part, to ones of small duration and/or very low or

very high frequency modulation that were initially detected but finally not considered as whistles in the selection process. The rest correspond to whistles that were not completely extracted due primarily to the presence of missing whistle pieces or very low SNR.

Additionally, to test the effectiveness of the method when dealing with overlapping signals, the percentage of correctly detected crosses were separately computed. In the analysed recording a total of 58 crosses between whistles were manually

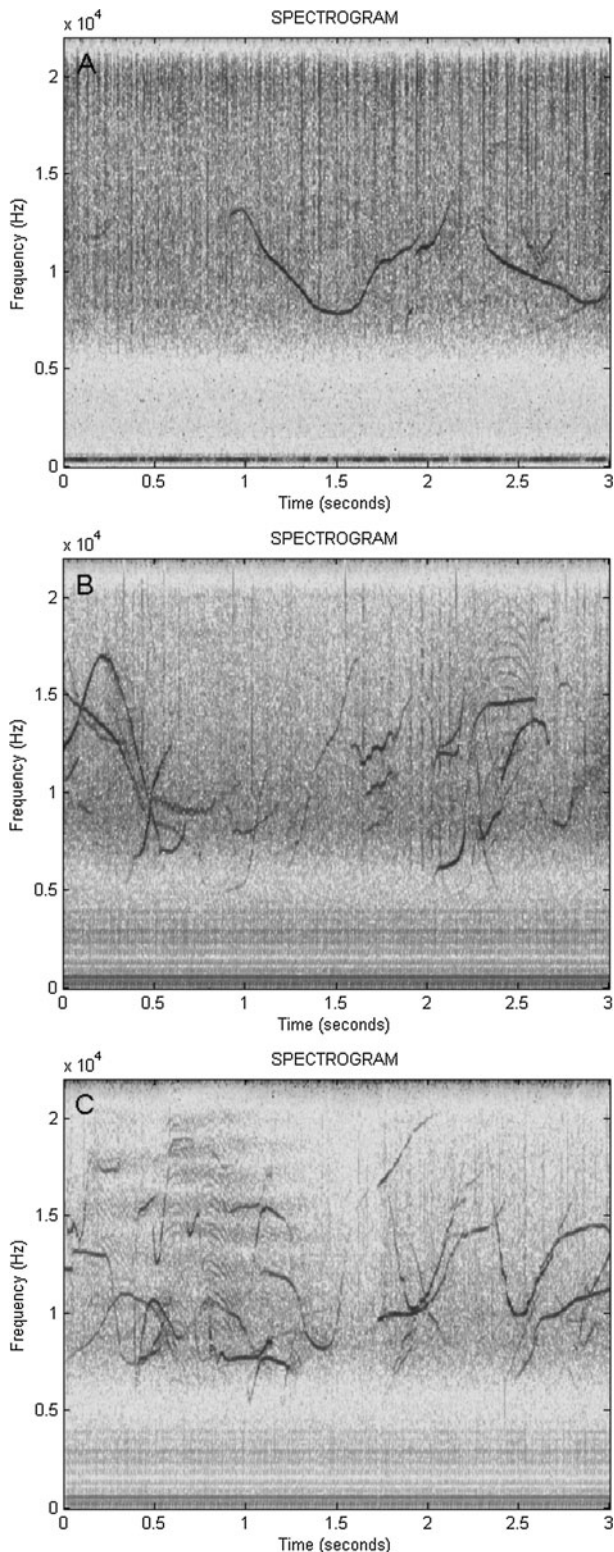


Fig. 12. Examples of pieces of recording. (A) Low complexity; (B) medium complexity; (C) high complexity.

identified. From this a total number of 46 were correctly resolved, which means that the whistles involved in a crossing were correctly extracted and characterized. This corresponds to a percentage of 79.3%. These results from the recorded signals are summarized in Table 1.

To further determine the characteristics of the method when dealing with recorded signals, it was next tested on

Table 1. Results on eight minutes recording analysis.

Number of manually identified whistles	737
Number and percentage of correctly extracted whistles	656 (89.0%)
Number of manually identified crosses between whistles	58
Number and percentage of correctly resolved crosses	46 (79.3%)

pieces of recordings of low and medium complexity. In particular the following aspects were addressed:

- The quality of the characterized signals, in the sense of their resemblance to the whistles identified in the spectrogram output.
- The influence of the presence of clicks.
- The influence of the presence of bands of noise.
- The ability in the process of handling overlapping signals.

Figure 13 includes the outputs of several stages of the processing chain for two pieces of recordings. The first corresponds to a three seconds fragment that can be qualified as low complexity. The waveform contains six independent whistles and the presence of clicks. The proposed method is able to correctly extract and characterize all the whistles included in the waveform and remains unaffected by the presence of clicks. The second fragment of recording, of five seconds length, can be qualified as medium complexity and contains isolated and overlapping whistles, clicks and horizontal bands of noise. An analysis of the output for this second recording yields the following conclusions:

- The presence of clicks has no significant influence on the performance of the method.
- The horizontal bands of noise do not generate false positives.
- Three crosses between whistles can be appreciated in the spectrogram. Two of them (corresponding to the objects labelled 18, 19 and 20) are correctly solved. The remaining (corresponding to the objects labelled with the numbers 5 and 6) is not correctly solved.
- The vast majority of the isolated whistles present in the recording are correctly extracted. When the SNR is significantly low along fragments of a whistle it can be split into more than one segment, as is the case of the objects labelled with the numbers 2 and 4.
- Three whistles are considered missed. They are located in the spectrogram at 2.0 seconds, 10 kHz; 3.0 seconds 8 kHz and 3.4 seconds, 15 kHz. The first one is of very low SNR and the third one has been only partially extracted (labelled with the number 17).
- Two false positive extractions have been also counted. They correspond to labels 10 and 21.
- The objects built from the radial basis function network outputs present a high level of resemblance with the whistles identified in the spectrogram.

An average time of 23.2 seconds has been measured for processing four-second duration recording segments, using the Matlab software and a standard PC equipped with Intel Pentium-4 Processor, CPU 3.0 GHz, 1.0 GB of RAM. More powerful computers and compilation-based software (such as C) will reduce this processing time, approximating it to real-time processing.

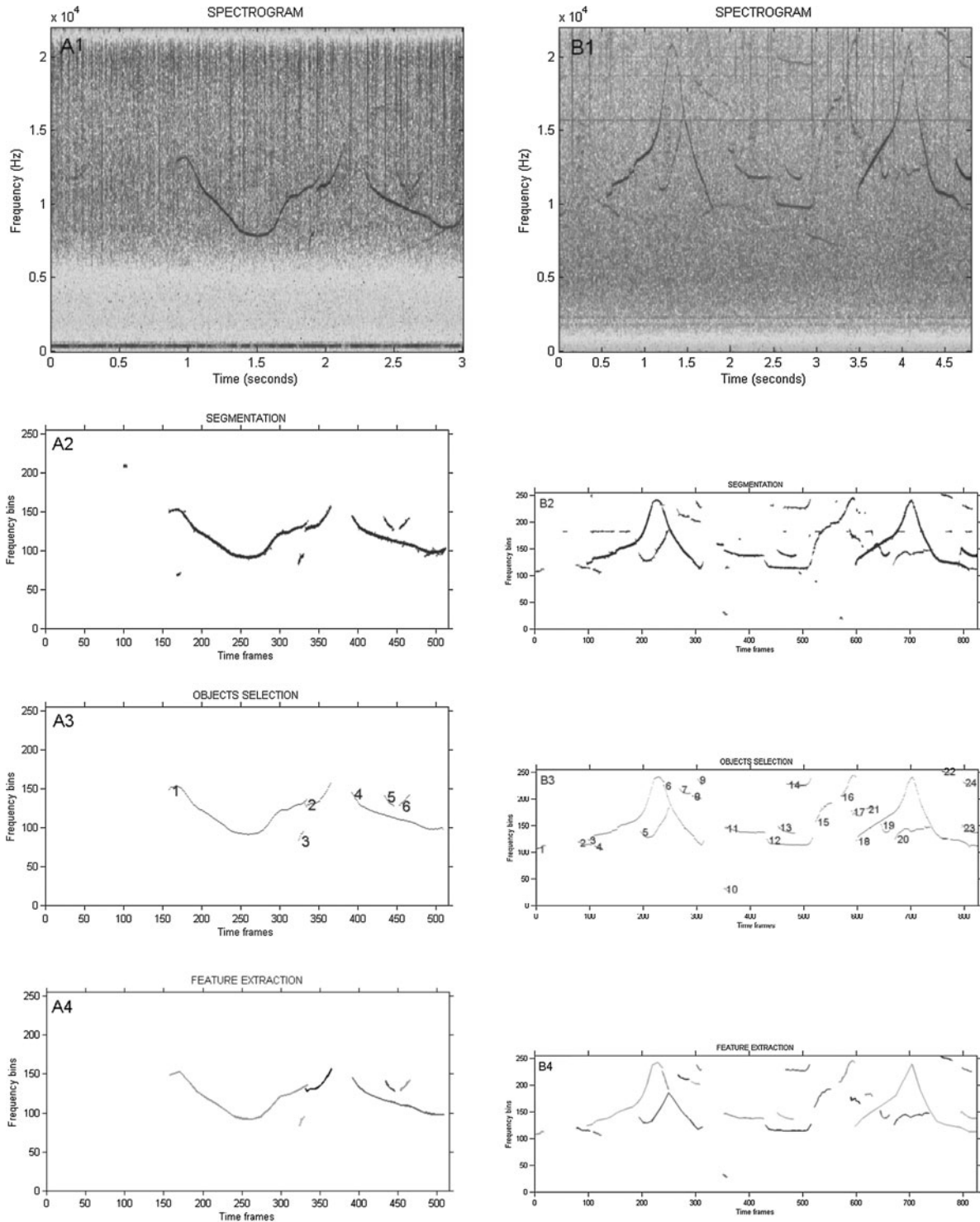


Fig. 13. Example of application of the proposed method to segments of recorded signals. The output of the stages: spectrogram, segmentation, objects selection and feature extraction are presented for: (A1 – A4) a segment of recording three seconds long containing clicks and six non-overlapping whistles; and (B1 – B4) a segment of recording five seconds long containing isolated and overlapping whistles, clicks and interfering horizontal bands of noise. The selected whistles from the objects selection stage are numbered.

DISCUSSION

The proposed method provides an efficient tool to automatically extract and characterize the whistles present in environments of low and medium complexity. Evidently, the method can also be applied to more complex environments with a lower percentage of success. The effectiveness and

performance of this method have been evaluated in realistic environments including different levels and types of interfering noises and the presence of clicks. Moreover, not only isolated whistles with high SNR have been used to test the method, but rather emphasis has been put on working with a wide variety of whistles including ones with low SNR and whistles split in several fragments. Also, one of the most

difficult issues in the process of detection and extraction of whistles is the handling of overlapping signals, which has been undertaken in the present study with an appreciable level of success.

The method is intended to automate whistle extraction and characterization by reducing the operator intervention. Parameters at the different stages of the processing chain have been tuned based on the analysed recordings. These parameters can be tuned to deal with specific environments. In particular the range of the suppressed low frequencies can be extended or reduced to adapt the processing to environments with high level of noise at low frequencies or when trying to detect low frequency calls. Also the minimum range of frequency bins and/or time frame selected to allow a curve to correspond to a whistle, can be modified to adapt to particular situations. Besides, the number of used radial basis functions, which determine the size of the feature vector and the quality of the matching with the extracted curve, can be specifically tuned.

The characteristics of the implemented normalization process on the spectrogram matrix minimize the effect of interfering clicks and interfering noises. Also, the accuracy of the whistles contours built from the radial basis function network outputs permit each whistle curve to be robustly characterized by means of the output layer weights of the RBF network.

The characteristics of the method mean that it can be generically applied to other calls characterized by a smooth frequency modulation, such as whale calls. Also the method has been built with the principle of modularity in mind in order to facilitate the improvements of specific stages of the processing chain. This greatly reduces the effect on the remaining stages.

The main contributions of the presented method can be summarized as follows:

- Automates the detection, extraction and characterization of whistles.
- Integrates techniques of digital signal processing, digital image processing and neural network processing.
- Extracts objects (whistles) based on a new image processing segmentation algorithm.
- Works with recording from different marine environment conditions in terms of SNR, presence of interfering noise and presence of interfering broadband calls.
- Extends its application to calls from other animals having similar characteristics in frequency evolution with time.
- Addresses the challenging case of overlapping whistles.
- Characterizes robustly the extracted whistles with a reduced set of elements by means of radial basis functions.

CONCLUSIONS

An automated method for the detection, extraction and characterization of marine mammal whistles has been presented. The performance of the method was tested on simulated signals and recordings made at sea encompassing a wide range of environmental conditions, including the presence of clicks, different levels of interfering noise and presence of isolated and overlapping whistles with a wide range of SNR. The method can be also applied to other marine mammals emitting calls with smooth frequency modulation. It combines techniques of signal processing, image processing and neural network processing and shows promising results for automating the extraction and characterization of marine mammals calls.

Significant percentages of correctly extracted whistles have been attained in environments of low and medium complexity. Also the challenging issue of whistle overlapping has been undertaken with significant success. The method follows the modularity principle in its design in order to simplify the process of future modifications and improvements.

Radial basis function (RBF) neural networks, which provide a fast and accurate means of approximating non-linear functions based on observed data, have been used to accurately characterize the extracted whistle curves. The feature vector for each whistle has been built from the weighting coefficients of the output layer of the RBF network.

Present and future efforts centre on refining the algorithm in charge of handling the crossing among whistles and to link pieces of whistles split in several parts due to drops in the SNR along the whistle. In future work we plan to tackle the classification of the characterized signals.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the help of the organizers of the 3rd International Workshop on Detection and Classification of Marine Mammals by permitting public access to the data sets used in the Workshop and of CIBRA (Centro Interdisciplinare di Bioacustica e Ricerche Ambientali–Università di Pavia) and especially of Gianni Pavan for providing data for this research. This work is partially supported by Ministerio de Educación y Ciencia under grant TEC2006-13338/TCM.

REFERENCES

- Adam O.** (2006) Advantages of the Hilbert Huang transform for marine mammals signal analysis. *Journal of the Acoustical Society of America* 120, 2965–2973.
- Au W.W.L.** (1993) *The sonar of dolphins*. 1st edition. New York: Springer-Verlag.
- Bazua-Duran C. and Au W.W.L.** (2002) The whistles of Hawaiian spinner dolphin. *Journal of the Acoustical Society of America* 112, 3064–3072.
- Brandes T.S., Naskrecki P. and Figueroa H.K.** (2004) Using image processing to detect and classify narrow-band cricket and frog calls. *Journal of the Acoustical Society of America* 120, 2950–2957.
- Brown J.C. and Miller P.J.O.** (2006) Automatic classification of killer whale vocalizations using dynamic time warping. *Journal of the Acoustical Society of America* 122, 1201–1207.
- Buck J.R. and Tyack P.L.** (1993) A quantitative measure of similarity for *Tursiops truncatus* signature whistles. *Journal of the Acoustical Society of America* 94, 2497–2506.
- Caldwell M.C., Caldwell D.K. and Tyack P.L.** (1990) A review of the signature whistle hypothesis for the Atlantic bottlenose dolphin. In Leatherwood S. and Reeves R.R. (eds) *The bottlenose dolphin*. San Diego, CA: Academic Press, Inc, pp. 199–234.
- Chen Z. and Maher R.C.** (2006) Semi-automatic classification of bird vocalization using spectral peak tracks. *Journal of the Acoustical Society of America* 120, 2974–2984.
- Datta S. and Sturtivant C.** (2002) Dolphin whistle classification for determining group identities. *Signal Processing* 82, 251–258.

- Deecke V.B. and Janik V.M. (2006) Automated categorization of bioacoustic signals. Avoiding perceptual pitfalls. *Journal of the Acoustical Society of America* 119, 645–653.
- Deecke V.B., Ford J.K.B. and Spong P. (1999) Quantifying complex patterns of bioacoustic variation: use of a neural network to compare killer whale (*Orcinus orca*) dialects. *Journal of the Acoustical Society of America* 105, 2499–2507.
- Ghosh J., Deuser L.M. and Beek S.D. (1992) A neural network based hybrid system for detection, characterization and classification of short-duration oceanic signals. *IEEE Journal of Oceanic Engineering* 17, 351–363.
- Gillespie D. (2004) Detection and classification of right whale calls using an 'edge' detector operating on a smoothed spectrogram. *Canadian Acoustics* 32, 39–47.
- Gonzalez R.C. and Woods R.E. (2001) *Digital image processing*. 2nd edition. Upper Saddle River, NJ: Prentice-Hall.
- Gonzalez R.C., Woods R.E. and Eddins S.L. (2004) *Digital image processing using MATLAB*. 1st edition. Upper Saddle River, NJ: Prentice-Hall.
- Gordon J.C.D., Gillespie D., Potter J., Frantzis A., Simmonds M., Swift R. and Thompson D. (2004) A review of the effects of seismic survey on marine mammals. *Marine Technology Society Journal* 37, 14–34.
- Halkias C.H. and Ellis D.P.W. (2006) Call detection and extraction using Bayesian inference. *Applied Acoustics* 67, 1164–1174.
- Haykin S. (1999) *Neural networks. A comprehensive foundation*. 2nd edition. Upper Saddle River, NJ: Prentice-Hall.
- Hu Y.H. and Hwang J.N. (eds) (2002) *Handbook of neural network signal processing*. 1st edition. Boca Raton, FL: CRC Press.
- Janik V.M. (1999) Pitfalls in the categorization of behaviour: a comparison of dolphin whistle classification methods. *Animal Behaviour* 57, 133–143.
- Janik V.M. and Slater P.J.B. (1998) Context specific use suggests that bottlenose dolphin signature whistles are cohesion calls. *Animal Behaviour* 56, 829–838.
- Lammers M.O., Au W.W.L. and Herzing D.L. (2003) The broadband social acoustic signalling behaviour of spinner and spotted dolphins. *Journal of the Acoustical Society of America* 114, 1629–1639.
- Learned R.E. and Willsky A.S. (1995) A wavelet packet approach to transient signal classification. *Applied and Computational Harmonic Analysis* 2, 265–278.
- Matej S. and Lewitt R.M. (1996) Practical considerations for 3-D image reconstruction using spherically symmetric volume elements. *IEEE Transactions on Medical Imaging* 15, 68–78.
- McCowan B. and Reiss D. (2001) The fallacy of 'signature whistles' in bottlenose dolphins: a comparative perspective of 'signature information' in animal vocalizations. *Animal Behaviour* 62, 1151–1162.
- Mellinger D.K. and Clark C.W. (2000) Recognizing transient low-frequency whale sounds by spectrogram correlation. *Journal of the Acoustical Society of America* 107, 3518–3529.
- Mellinger D.K. and Barlow J. (2003) *Future directions for acoustic marine mammal surveys: stock assessment and habitat use*. NOAA OOR Special Report. NOAA/PMEL Contribution 2557. 37 pp.
- Murray S.O., Mercado E. and Roitblat H. (1998) The neural network classification of false killer whale (*Pseudorca crassidens*) vocalizations. *Journal of the Acoustical Society of America* 104, 3626–3633.
- Norris T.F., McDonald M. and Barlow J. (1999) Acoustic detection of singing humpback whales (*Megaptera novaeangliae*) in the eastern North Pacific during their northbound migration. *Journal of the Acoustical Society of America* 106, 506–514.
- Oswald J.N., Rankin S. and Barlow J. (2004) The effect of recording and analysis bandwidth on acoustic identification of delphinid species. *Journal of the Acoustical Society of America* 116, 3178–3185.
- Oswald J.N., Rankin S., Barlow J. and Lammers M.O. (2007) A tool for real-time acoustic species identification of delphinid whistles. *Journal of the Acoustical Society of America* 122, 587–595.
- Park J. and Sandberg I.W. (1991) Universal approximation using radial-basis-function networks. *Neural Computation* 3, 246–257.
- Park J. and Sandberg I.W. (1993) Approximation and radial-basis-function networks. *Neural Computation* 5, 305–316.
- Pavan G. (2007) Acoustic Risk Mitigation in the Mediterranean Sea. Current situation and recommendations. In *Proceedings of the Underwater Defence Technology (UDT)*, Naples (Italy).
- Popper A.N. (1980) Sound emission and detection by delphinids. In Herman L.M. (ed.) *Cetacean behavior: mechanisms and function*. New York: John Wiley & Sons, pp. 1–52.
- Richardson W.J., Green C.R., Malme C.I. and Thomson D.H. (1995) *Marine mammals and noise*. San Diego, CA: Academic Press Inc.
- Rickwood P. and Taylor A. (2008) Methods for automatically analyzing humpback songs units. *Journal of the Acoustical Society of America* 123, 1763–1772.
- Roch M.A., Soldevilla M.S., Burtenshaw J.C., Henderson E.E. and Hildebrand J.A. (2007) Gaussian mixture model classification of odontocetes in the Southern California Bight and the Gulf of California. *Journal of the Acoustical Society of America* 121, 1737–1748.
- Sanchez-García A., Rodrigo F.J. and Sancho-Gómez J.L. (2008) Sonar transient false alarm reduction based on automated detection and characterization of marine mammal sounds. In *Proceedings of the Underwater Defence Technology (UDT)*, Glasgow (UK).
- Sanner R.M. and Slotine J.E. (1994) Gaussian network for direct adaptive control. *IEEE Transactions on Neural Network* 3, 837–863.
- Searby A. and Jouventin P. (1995) The double vocal signature of crested penguins: is the identity coding system of rockhoppers (*Eudyptes chrysocome*) due to phylogeny or ecology? *Journal of Avian Biology* 36, 449–460.
- Simmonds M.P. and Lopez-Jurado L.F. (1991) Whales and the military. *Nature* 351, 448.
- Theriault J.A. (2005) Marine mammals and active sonar. *Sea Technology* November, 23–29.
- Trifa V.M., Kirschel A.N.G., Taylor C.E. and Vallejo E.E. (2008) Automated species recognition of antbirds in a Mexican rainforest using hidden Markov models. *Journal of the Acoustical Society of America* 123, 2424–2431.
- Tyack P.L. and Clark C.W. (2000) Communication and acoustic behaviour of dolphin and whales. In Au W.W.L., Popper A.N. and Fay R.R. (eds) *Hearing by whales and dolphins*. New York: Springer Verlag, pp. 156–224.
- and
- Van Ijsselmuide S.P. and Beerens S.P. (2004) Detection and classification of marine mammals using an LFAS system. *Canadian Acoustics* 32, 93–106.

Correspondence should be addressed to:

A. Sánchez-García
 Sociedad Anónima de Electrónica Submarina (SAES)
 Carretera de la Algameca s/n
 30205-Cartagena (Spain)
 email: a.sanchez@electronica-submarina.com