An approach for automatic classification of grouper vocalizations with passive acoustic monitoring

Ali K. Ibrahim, Laurent M. Chérubin, Hanqi Zhuang, Michelle T. Schärer Umpierre, Fraser Dalgleish, Nurgun Erdol, B. Ouyang, and A. Dalgleish

---

Articles you may be interested in

Passive, broadband suppression of radiation of low-frequency sound
The Journal of the Acoustical Society of America **143**, EL67 (2018); 10.1121/1.5022192

Time-domain sound field reproduction using the group Lasso
The Journal of the Acoustical Society of America **143**, EL55 (2018); 10.1121/1.5022280

Acoustic noise interferometry in a time-dependent coastal ocean
The Journal of the Acoustical Society of America **143**, 595 (2018); 10.1121/1.5022287

Beamforming using chip-scale atomic clocks in a controlled environment
The Journal of the Acoustical Society of America **143**, EL74 (2018); 10.1121/1.5022191

Elastic feature of cylindrical shells extraction in time-frequency domain using biomimetic dolphin click
The Journal of the Acoustical Society of America **142**, 3787 (2017); 10.1121/1.5017835

A systematic method for isolating, tracking and discriminating time-frequency components of bat echolocation calls
The Journal of the Acoustical Society of America **143**, 716 (2018); 10.1121/1.5023205

An approach for automatic classification of grouper vocalizations with passive acoustic monitoring

Ali K. Ibrahim,[1,a] Laurent M. Chérubin,[2] Hanqi Zhuang,[1] Michelle T. Schärer Umpierre,[3] Fraser Dalgleish,[2] Nurgun Erdol,[1] B. Ouyang,[2] and A. Dalgleish[2]

[1]*Department Computer and Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, Florida 33431, USA*

[2]*Harbor Branch Oceanographic Institute, Florida Atlantic University, 5600 US1 North, Fort Pierce, Florida 34946, USA*

[3]*Department of Marine Sciences, University of Puerto Rico, Mayagüez 00681, Puerto Rico, USA*

Grouper, a family of marine fishes, produce distinct vocalizations associated with their reproductive behavior during spawning aggregation. These low frequencies sounds (50–350 Hz) consist of a series of pulses repeated at a variable rate. In this paper, an approach is presented for automatic classification of grouper vocalizations from ambient sounds recorded *in situ* with fixed hydrophones based on weighted features and sparse classifier. Group sounds were labeled initially by humans for training and testing various feature extraction and classification methods. In the feature extraction phase, four types of features were used to extract features of sounds produced by groupers. Once the sound features were extracted, three types of representative classifiers were applied to categorize the species that produced these sounds. Experimental results showed that the overall percentage of identification using the best combination of the selected feature extractor weighted mel frequency cepstral coefficients and sparse classifier achieved 82.7% accuracy. The proposed algorithm has been implemented in an autonomous platform (wave glider) for real-time detection and classification of group vocalizations. © 2018 Acoustical Society of America. https://doi.org/10.1121/1.5022281

[JFL]   Pages: 666–676

I. INTRODUCTION

Mature adults of many fish species swim long distances and gather in high densities for mass spawning at precise locations and times.[1] This widespread reproductive strategy is typically shared among the groupers, which are both keys to the trophic balance of marine ecosystems and targeted by humans. Worldwide depletion of large predatory fishes has already caused top-down changes in coral reef ecosystems and biodiversity loss.[2,3] Moreover, most known fish spawning aggregations (FSAs) sites are shared by many species at different times[4] and as such, represent breeding hotspots. It is critical that their role in the persistence of marine populations be elucidated. FSAs share common features such as large body-sized individuals, strong site fidelity, and geomorphological attributes, (i.e., shelf-break, capes).[5,6] Once located, they are easily over-exploited and depleted[7,8] ICRS 2004. Despite numerous historical records of Caribbean-wide FSAs[9–23] only a few are viable to date and many remain unprotected.

These FSAs in the Caribbean Sea, Gulf of Mexico, and the Bahamas Region (i.e., the intra-America seas) are where a number of vocalizing grouper species such as the Nassau (*Epinephelus striatus*), yellowfin (*Mycteroperca venenosa*), red hind (*Epinephelus guttatus*), and black grouper (*Mycteroperca bonaci*), among others, aggregate to spawn. Most of these species spawn during the winter and spring months in the northern hemisphere. The timing of spawning is usually cued to the moon and daylight, but also to water temperatures and the local current conditions. Because FSAs often occur at remote locations, at dusk, and are in water depths between 30 and 80 m, near the shelf break, spawning activities and fish population are challenging to observe, and thus to monitor.

Studies have shown that more than 800 fish species can produce sounds for diverse purposes.[24,25] Most of the sounds are emitted at low frequencies,[26] usually below 1000 Hz. However, some pulses can reach 8 kHz[27,28] or present more complex characteristics.[29] In addition, these emissions are typically broadband short-duration signals (see Fig. 1). Fish generate sounds through several mechanisms, which depend on the species and a variety of circumstances, such as courtship, threats or defending territory.[30] Passive acoustics sensors can record species-specific acoustic signals associated with fish behaviors. The analysis of recordings of the sonorous species at FSAs has recently become a new approach in addition to underwater visual observations to monitor fish activity, such as courtship behavior, presence, and residence time. This approach is also used to scout the shelf edge where FSAs are likely to exist, which could reveal unknown aggregation sites or the recovery of overfished FSAs along with the species that visit the FSA.[31] Passive acoustic recordings are usually conducted at fixed stations with long-term acoustic recorder that can last several months underwater. Large volumes of acoustic data are usually generated and are manually classified using sounds spectrograms for visual identification and auditory classification, which can be tedious, time consuming, and prone to errors.

a)Electronic mail: aibrahim2014@fau.edu
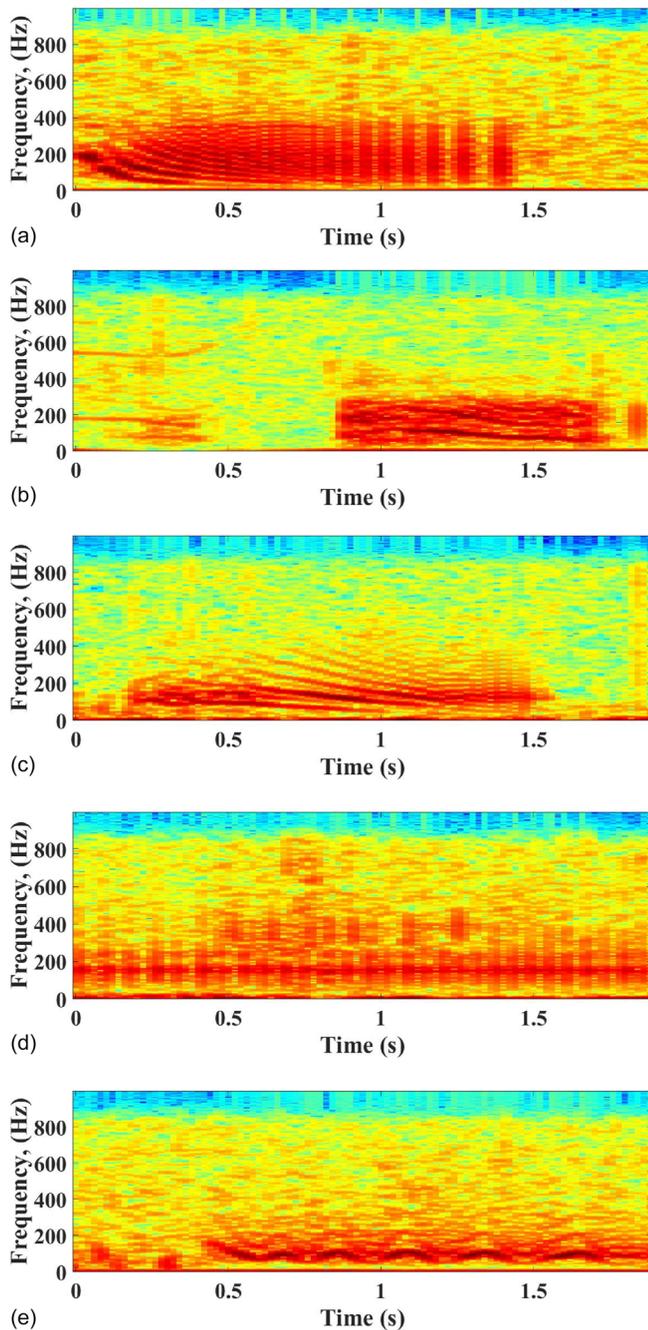
FIG. 1. (Color online) (a) Red hind (*E. guttatus*) sound spectrogram; (b) Nassau grouper (*E. striatus*) sound spectrogram; (c) Yellowfin (*M. Venenosa*) tonal sound spectrogram; (d) Yellowfin (*M. venenosa*) pulse train spectrogram; (e) Black grouper sound spectrogram.

Few automatic identification methods exist, which mostly focused on sounds of higher frequency than those of fish vocalizations. Chesmore and Ohya[32] proposed an identification scheme for *Orthoptera* species using temporal features based on the shape of waveform and duration between consecutive zero-crossings followed by a multilayer perceptron (MLP) classifier. In Ibrahim *et al.*,[33] a hybrid features extraction using discrete wavelet transform and mel frequency cepstral coefficients (MFCC) is proposed for detection North Atlantic Right Whales. In Mellinger *et al.*,[34] a complex detection method is presented for humpback whales by frequency contour tracing and by multiple parameter optimization. An unsupervised classification method for bird song syllables has been proposed by Hansson-Sandsten,[35] based on singular vectors of multitaper spectrogram and the similarity measures of two syllables using pairs of singular vectors. And finally, the study by Starkhammar and Hansson-Sandsten[35] presents an evaluation of different time-frequency representations for target detection applied to broadband echolocation signals of dolphins.

Fish sound production, including that of groupers, has long been known.[36] Some fish sounds are species-specific in frequency and pulse rate, which allows their presence to be detected from acoustic recordings.[36–39] Sattar *et al.*[40,41] proposed an automated approach to quantify and identify the sound of the plainfin midshipman (*Porichthys notatus*), a high vocal species of toadfish found in the northeast Pacific Ocean. The frequency range for this species is within 100 Hz, which is also close to the lower end of the frequency range for the grouper species targeted in this work. Sattar *et al.*[40] proposed an automated identification scheme for fish vocalizations based on the auditory analysis for feature extraction followed by a machine-learning algorithm for classification. The auditory analysis uses amplitude modulation spectrogram (AMS). The information containing amplitude modulations of the input signal is analyzed and represented in two-dimensional AMS from which the high-resolution features are extracted. A support vector machine (SVM) classifier is then trained on a large number of pre-selected AMS patterns, which classifies the input signal. They also evaluated the AMS feature extraction against the MFCC method, which had a lower accuracy that than the AMS method at detecting the growls and grunts made by *P. notatus*. In Sattar *et al.*,[41] the authors evaluated the performance of the multiresolution acoustic features (MRAF) extraction method and robust principal component analysis (RPCA) based feature selection to identify grunts, growls and groans from *P. notatus*. RPCA identifies a low rank representation, random noise, and a set of outliers by repeatedly calculating the singular value decomposition (SVD) and applying "thresholds" to the singular values and error for each iteration.[42,43] The RPCA plays a significant role in tackling the key challenges involved with big data[44] by minimizing false alarms, reducing seasonal variability and processing the data that are not normally distributed. Their classification accuracy showed improvement over the MFCC feature extraction method and an increased capacity at distinguishing between features because of RPCA robustness to overlapping low-frequency spectral contents among different classes. Noda *et al.*[45] were able to successfully classify 102 different fishes species using linear frequency cepstral coefficients (FCC) and MFCC, Shannon entropy and syllable length features extraction methods. For the classification, they used three widely used machine-learning algorithms: K-nearest neighbors (KNN), random forest (RF), and support vector machine (SVM). Their experimental results show an average classification accuracy of 95.24%, 93.56%, and 95.58%, respectively. All these methods have the advantage of being automated, and computationally cost efficient. They can be implemented on a small electronic chip, which can be installed on portable and autonomous devices for real-time detection.

J. Acoust. Soc. Am. **143** (2), February 2018

Ibrahim *et al.* 667

In this study, we propose to use a similar features extraction methods and machine learning classifiers association to identify four species of grouper that co-occur at spawning aggregation sites in the US Caribbean and whose courtship associated sound (CAS) have been described (Table I). The species-specific vocalizations are distinctive in duration, peak frequency, and tonal characteristics and are easily distinguished from each other audibly and visually in spectrograms. Figure 1 shows the spectrogram of the four species targeted in this study. Red hind (*E. guttatus*), whose vocalizations are within the 100–200 Hz band[36] and consist of a variable number of pulses, with one or more portion of the call being tonal, at a higher pulse rate than the rest of the pulses [Fig. 1(a)]: Nassau grouper (*E. striatus*), whose vocalizations consist of a pulse train made up of a varying number of short individual pulses and tonal sound in the 30–300 Hz band[37] [Fig. 1(b)]; Yellowfin grouper (*M. venenosa*), whose vocalizations consist of the same types of calls as the Nassau grouper, although longer in duration, with frequency ranging between 90 and 150 Hz[38] [Figs. 1(c) and 1(d)]; Black grouper (*M. bonaci*), which produce at least two variations of a low frequency, modulated tonal call, which ranges between 60 and 120 Hz, but generally has a longer duration than *E. striatus*.[39]

The features extraction methods used in this study are MFCC and MRAF. The MFCCs are short-term spectral based features, which despite being a powerful representation do not work well under noisy condition due to its mismatch problem. In this study, an optimized version of MFCC that consists of the application of weighting dynamic features was used along with a more robust feature, weighted MRAF (WMRAF) to classify the CAS of the four grouper species.

The paper is organized as follows. In Sec. II, we describe the measurements, their location, and recording characteristics. We also provide the human ear detection method used to validate the electronic classification method. In Sec. III we describe the two proposed types of features extraction using WMFCC and WMRAF. In Sec. IV, we describe the classification method. Experimental results with combinations of feature extraction and classification are discussed in Sec. V, and concluding remarks are given in Sec. VI.

II. DATASETS

The proposed system for classification of grouper CAS was tested on three datasets, collected at different spawning aggregation sites. The first dataset was recorded on the west coast of Puerto Rico at Abrir La Sierra (ALS),[37] a site known to have a spawning aggregation of red hind at a depth of 25 m.[36] The second dataset was collected at Bajo de Sico (BDS) Bank, which is a submerged seamount approximately 27 km west of Puerto Rico, surrounded by depths of over 250 m to the southeast near the Puerto Rico insular shelf and over 1000 m to the north. Currently, there is a six-month seasonal closure to reef fish fishing from October to March of each year.[20] Nassau grouper have been reported at BDS during the non-spawning time, and more recently a spawning aggregation of approximately 100 individuals has been documented.[37] The third dataset was collected at Mona Island, located offshore, 72 km west of Puerto Rico,[38] within the no-take marine reserve. At this site Yellowfin grouper spawning aggregation have been repeatedly reported.[38]

At each site, a DSG-Ocean (Loggerhead Instruments) recording unit was deployed in December, prior to the grouper spawning season and recovered in June. Each unit was programmed to record ambient sounds for 20 s at 5 min intervals at a sample rate of 10 kHz to optimize battery life. These recordings were made onto an SD memory card and downloaded as one.wav file for each 20 s recording. This cycle generates 288 files per day that are stored in independent folders with 9999 files each. Each file can be heard with noise canceling headphones or visualized with acoustic analysis software. Grouper sounds were quantified per file by visual inspection of spectrograms using Ishmael Bioacoustics software. Each file was displayed and classified by an observer depending on the pattern, duration, and frequency of each signal. Sounds were summed by species for each file and pooled as necessary for comparison with the algorithm detections. The presence of at least one CAS per species was compared with the results of the algorithm detections. The files used for the human analysis were not modified for the automated classification.

III. FEATURES EXTRACTION

We now present the sequence of the acoustic files processing for each features extraction method, MFCC, and MRAF, respectively.

A. Mel frequency cepstral coefficients

An MFCC process converts linear spectrum into nonlinear mel-spectrum. The corresponding relationship between the linear-scale frequency f and the mel-scale frequency f_{mel} is shown below

$$f_{mel} = 2595 \log_{10}\left(1 + \frac{f}{700}\right). \tag{1}$$

TABLE I. Groupers sounds characteristics.

Type of Species	Frequency range (Hz)	Peak frequency (Hz)	Bandwidth (Hz)	Duration (s)
Red hind	50–350	213 ± 23	38.2 ± 18.5	1.78 ± 1.02
Nassau Grouper	90–150	99 ± 33.6	22.4 ± 12.2	1.6 ± 0.3
Yellowfin Pulse train	101.4–132.4	120.46 ± 7.45	33.03 ± 6.13	2.96 ± 0.97
Yellowfin Tonal call	88.9–141.7	$121.04 + 12.57$	43.18 ± 4	$3.14 + 0.95$
Black Grouper	60–150	108 ± 9	31 ± 6.3	1.7 ± 0.85

MFCC is calculated using a 1024 point fast Fourier transform with a 0.1 s frame length using a Hamming window and overlap 50%. The Hamming window function is multiplied with the signal results in a 1024 length non-zero frame. The Hamming window was chosen because of its high frequency domain resolution and low spectral leakage. This results in a FFT with 1024 unique points for each frame. Once the FFT is calculated for each frame the spectrogram is estimated by taking the squared absolute value of each frame's FFT, and triangular filer bank were applied to the squared absolute value. Discrete Cosine transform and Lifer were applied to the output of the filter bank to get MFCC's coefficients. The flow chart of calculating the traditional MFCC is shown in Fig. 2.

The traditional MFCC only represent the sound feature but does not consider the dynamic characteristics of the sound. In order to improve the performance, the popular method is to combine the traditional MFCC and its first-order differential coefficients,[33] which can effectively reflect the dynamic characteristics of the sound. However, this approach increases both the dimension of the parameters and the computational complexity of the sound recognition system.

In this paper, an optimization algorithm is proposed using the weighted dynamic MFCC.[46] Assuming the new weighted dynamic MFCC as WMFCC, the equation is shown as follows:

$$\text{WMFCC} = \text{MFCC} + \alpha_1\, \Delta\text{MFCC} + \alpha_2\, \Delta^2\text{MFCC}, \quad (2)$$

where ΔMFCC is the first-order differential coefficient, Δ^2MFCC is the second-order differential coefficient, with α_1 and α_2 as their weights, respectively. The weight parameters are calculated by using the simple cyclic rule. In the cyclic rule, one weight kept constant and the other weight is selected to minimize the cost function:

$$\arg\min_{\alpha_1,\alpha_2} \frac{1}{n}\sum_{i=1}^{N}\left[l_i - \hat{l}_i(\alpha_1,\alpha_2)\right]^2, \quad (3)$$

where l_i, $\hat{l}_i$ are, respectively, training label and classified label for the ith data point. The process is repeated with the role reversal for the two weights until the weight variations are below a preset threshold. In each iteration, the golden rule is applied to select an optimal value of the weight.[47] The golden rule is an extension of the bisection rule. In the
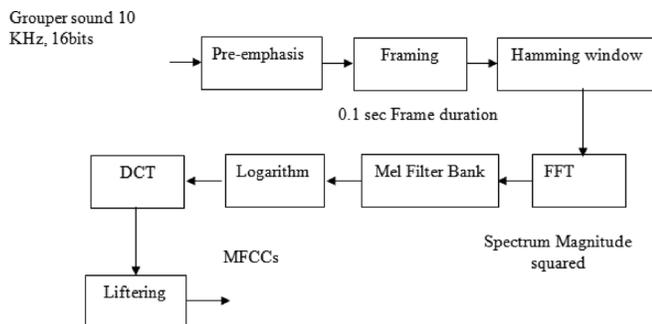
bisection rule, an interval is repeatedly bisected with a ratio of 1:1 and a sub-interval is selected in which the optimal solution must lie. The golden rule cuts the internal into two sub-intervals with a golden ratio of 1:1.618 instead of 1:1.

The weighted dynamic MFCC could characterize the grouper sound's voiceprint features and the dynamic characteristics of the sound. As a result, by using these feature parameters, the computational complexity of the system is significantly reduced while maintaining a high recognition rate. In Eq. (2) of this new coefficient WMFCC represents the new features, the element MFCC depicts the sound channel characteristics, ΔMFCC reflects the sound rate, and Δ^2MFCC provide information similar to acceleration of sound. Let M = MFCC, ΔM = ΔMFCC, and $\Delta^2 M$ = Δ^2MFCC, we have

$$\Delta M_{i,n} = \sum_{k=-2}^{2} k M_{i-k,n}, \quad (4)$$

$$\Delta^2 M_{i,n} = \sum_{k=-2}^{2} k \Delta M_{i-k,n}, \quad (5)$$

where $i = 3, 4, 5, \ldots, T{-}2$ is the frame of the feature parameters and $n = 1, 2, 3, \ldots, N$ is the dimension of the feature parameters.

The weighted dynamic MFCC has the same dimension as MFCC and takes advantages of dynamic features. The feature parameters with the concept of the weighted dynamic MFCC would have better performance at capturing underwater sounds. However, the dimension of the coefficient matrix in Eq. (2) is less by 67% in comparison to the matrix in Eq. (6):

$$\text{WMFCC} = \begin{bmatrix} \text{MFCC} \\ \alpha_1\, \Delta\text{MFCC} \\ \alpha_2\, \Delta^2\text{MFCC} \end{bmatrix}. \quad (6)$$

The new chart to derive the WMFCC method if now shown in Fig. 3.

B. Multiresolution features

We now describe the MRAF, which encodes the multiresolution energy distributions in the time-frequency plane based on the cochleagram representation of an input signal. We incorporate a number of cochleagrams at different resolutions to design the MRAF features set. The cochleagram with high resolution captures the local information, while the other low-resolution cochleagrams capture the contextual information at

FIG. 2. Block diagram of MFCC extraction.
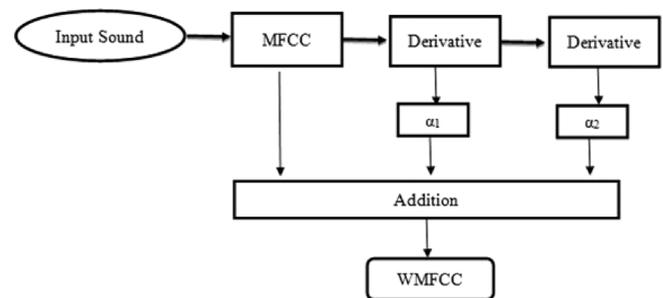
FIG. 3. Block diagram of WMFCC extraction.

J. Acoust. Soc. Am. **143** (2), February 2018

Ibrahim *et al.* 669

different scales. These scales are used to catch the time frequency behavior of the sound signal at different resolutions.[48] To compute the cochleagram, we first pass an input signal to a gammatone filter bank, where the impulse response of a particular gammatone filter has an impulse response given by

$$h(t) = t^{a-1} e^{-2\pi\beta_{f_c}t} \cos(2\pi f_c t), \qquad (7)$$

where parameter a is the order of the filter, f_c denotes the center frequency while β_{f_c} refers to the bandwidth given f_c. Figure 4 shows the frequency response of gammatone filter banks.

The gammatone filter function is used in models of the auditory periphery representing critical-band filters where the center frequencies f_c are uniformly spaced on the equivalent rectangular bandwidth (ERB) scale. The relation between β_{f_c} and f_c is given by

$$\beta_{f_c} = 1.019 * \mathrm{ERB}(f_c) = 1.019 * 24.7(4.37 * f_c/1000 + 1). \qquad (8)$$

Each response signal from the gammatone filter bank is divided into 20 ms frames with a 10 ms frame shift to extract high resolution features. Furthermore, each response signal from the gammatone filter bank is divided into 200 ms frames with a 100 ms frame shift to obtain low resolutions. The low resolution and high resolutions cochleagram features represent the local information and global information of the sound signal, respectively. It has been shown that cochleagram features at a low resolution, i.e., frame length = 200 ms, can detect patterns of noisy speech better than those at only a high resolution. Meanwhile, high resolution features complement low resolution ones.

MRAF feature set provides contextual information by including the energy distribution in the neighborhood of each T-F unit. The steps for computing MRAF are as follows.

Step 1: Given an input of sound data, compute the first 64-channel cochleagram, CG1. A log operation is applied to each T-F unit.

Frequency Response of Gammatone Filter bank
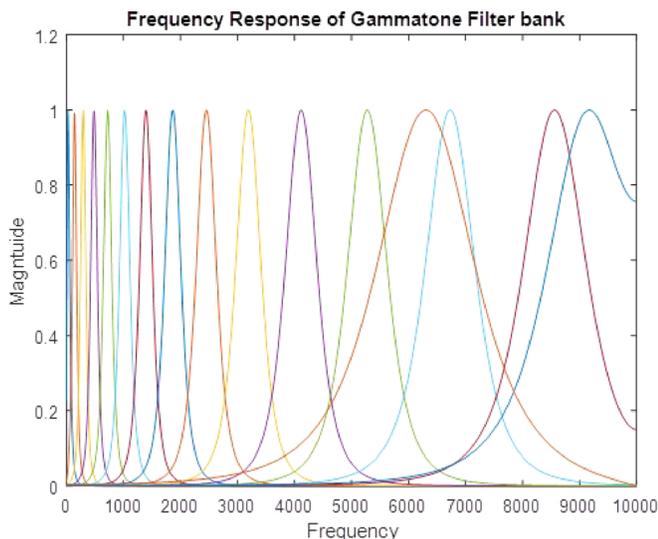
FIG. 4. (Color online) An example of gammatone filter bank.

Step 2: Similarly, compute CG2 with the frame length of 200 ms and frame shift of 100 ms.

Step 3: CG3 is derived by averaging CG1 across a square window of 11 frequency channels and 11 time frames centered at a given T-F unit. If the window goes beyond the given cochleagram, the outside units take the value of zero (i.e., zero padding).

Step 4: CG4 is computed in a similar way to CG3, except that the window size is (5*5).

Step 5: Concatenate CG1-4 to obtain the MRAF feature vector.

Step 6: Calculate the dynamic features delta and delta square.

Step 7: Weight the dynamic features ΔMRAF and Δ^2MRAF, and add them to the static features.

$$\mathrm{WMRAF} = \mathrm{MRAF} + w_1 \Delta\mathrm{MRAF} + w_2\Delta^2\mathrm{MRAF}, \qquad (9)$$

where w_1 and w_2 are weighted parameters. The weights parameters are calculated by using golden rule.[47] Considering their different contribution to the speech feature parameters, the constraint condition should be set as: $w_1 < w_2 < 1$ ($w_1 = 0.27$, $w_2 = 0.571$).

Figure 5 shows the WMRAF algorithm.

IV. CLASSIFIERS

In this section, a summary of classifiers is given.

A. K-nearest neighbors

KNN is a simple classifier, which is based on the clustering of the elements that have the same characteristics. It decides the class category of a test example based on the classes of its k neighbors that are near to it. The value of k in the KNN depends on the size of dataset and the type of the classification problem.[49]

B. Support vector machine

SVM, derived from the theory of Structural Risk Minimization, was first introduced by Vapnik[50] Detailed information and further references can be found in Refs. 51 and 52 SVMs are classifiers that can be separate objects into their respective groups using lines or hyperplanes that are derived from the objects. When it is not possible for a straight line to separate the objects, a kernel transformation can rearrange the objects so that their separation by a hyperplane is possible. Next we go over the SVM algorithm.

Given a training set $\{x_i, y_i\}$ consisting of training vectors $x_i \in R_n$ and their corresponding labels $y_i \in \{-1, +1\}$, a kernel function $K(x, y)$, and a parameter C, the SVM classifier finds an optimal separating hyperplane in F. This is done by solving the following quadratic programming problem: Choosing the vector λ, a collection of α_i, which maximizes

$$W(\lambda) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{J=1}^{N} \lambda_i\lambda_j y_i y_j k(x_i, x_j) \qquad (10)$$

under the constraints
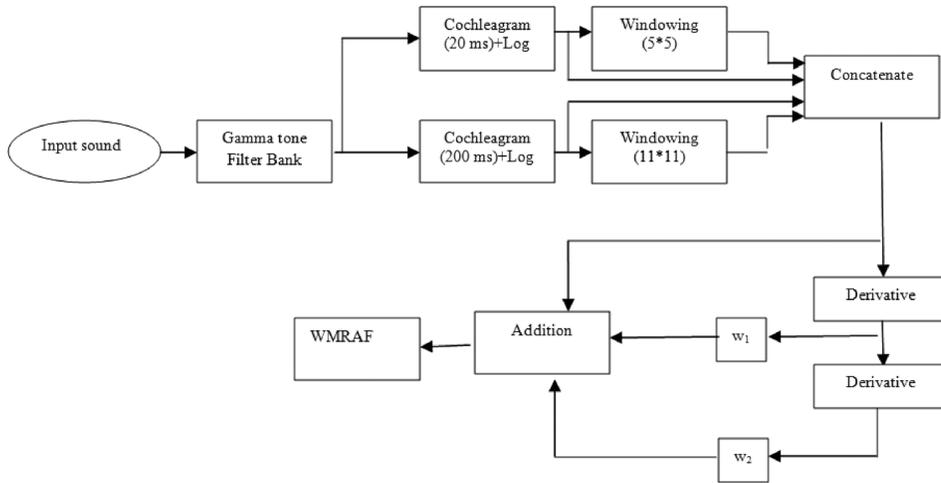
FIG. 5. A proposed WMRAF block diagram.

$$\sum_{i=1}^{N} \lambda_i y_i = 0 \quad \text{and} \quad 0 \leq \lambda_i \leq C \quad \forall i. \tag{11}$$

The parameter $C > 0$ allows us to specify how strictly we want the classifier to fit to the training data (a larger C meaning more strictly). The output of SVM is

$$f(x) = \sum_{i=1}^{N} \lambda_i y_i K(x_i, x), \tag{12}$$

where $f(x) > 0$ means that x is classified to class $+1$. The training vectors x_i for which the λ_i are greater than zero are called support vectors.

C. Sparse classification

In a classification problem, labeled training data feature vectors from different classes are used and then the test data feature vector v is assigned to a particular class, using an algorithm. For example, in a grouper sound task, feature vectors are obtained by using WMFCC and converted to a vector of length M ($M = 5004$). Assuming that all training feature vectors, from the ith grouper sound are placed in a matrix A_i as column vectors to serve as exemplars; $A_i = [v_{i1} \, v_{i2} \, \cdots \, v_{iN}]$, where v_{iN} represents the Nth training feature vectors of the ith grouper sound.

If y is a test feature vector ($M \times 1$) from the ith grouper sound, then y can be represented as a weighted linear combination of all entries in A_i,

$$y = \gamma_{i1} v_{i1} + \gamma_{i2} v_{i2} + \cdots + \gamma_{iN} v_{iN}, \tag{13}$$

where γ_{ij} are scalar quantities (weights) to be determined. In order to determine the class of y (i.e., finding the weights), a global exemplar dictionary matrix A ($M \times L$) needs to be developed to include training feature vectors from all k classes (groupers) by concatenating A_i ($i = 1, \ldots, K$).

$$A = [A_1, A_2, \ldots, A_K]$$
$$= \left[v_{11} \, v_{12} \, \cdots \, v_{1N} \vdots v_{21} \, v_{22} \, \cdots \, v_{2N} \vdots v_{K1} \, v_{K2} \, \cdots \, v_{KN} \right]_{M \times L},$$

where $L = K*N$. The number of column vectors in $A_1, A_2, A_3, \ldots$ depends on the number of training available for each grouper species. For simplicity, it is assumed that all classes have the same number of training features (N). Now the test vector y can be represented as a linear combination of all k classes of training feature vectors in the matrix A above by using vector x, i.e.,

$$[y]_{M \times 1} = [A]_{M \times L} [x]_{L \times 1}. \tag{14}$$

The linear system of Eq. (14) can be solved and the class of y can be found by using the information in x. For example, if y belongs to red hind grouper, then the weights of x (i.e., $'s$) that are not associated with the red hind grouper should ideally be zero.

$$\mathbf{x} = \left[\gamma_{11} \, \gamma_{12} \, \cdots \, \gamma_{1N} \vdots 0, 0, \ldots, 0 \vdots 0, 0, 0, \ldots \right]^T. \tag{15}$$

Ideally the vector x will exhibit a high level of sparsity, and the non-zero weights will correspond to exemplars from red hind grouper. In practice this sparsity condition is enforced by choosing an appropriate solution to the system of linear equations in Eq. (14). The feature vector dimension M is much smaller than L; ($M \ll L$). Therefore, the system, $y = Ax$, has more unknown a's than equations (i.e., an underdetermined system) and has infinite solutions.[53,54] In order to obtain a single well-defined solution, and preferably in this case the sparsest solution, additional criteria are needed and this is achieved using the following optimization techniques.

Find the weight vector, x, such that $y = Ax$ and $\|x\|_1$ is minimized. That is,

$$\min_x \|x\|_1 \text{ subject to } y = Ax, \tag{16}$$

where $\|x\|_1$ is the ℓ_1 norm. This will lead to a sub-optimal solution where iterative methods like the matching pursuit and orthogonal matching pursuit are used and the measurements are noisy. A generalized version of Eq. (16) is given below:

$$\min_x \|y - Ax\|_1 + \beta \|x\|_1, \tag{17}$$

where the vector y is noisy and assumed to be generated by $y = Ax + \varepsilon$ and ε is a Gaussian white noise vector. The
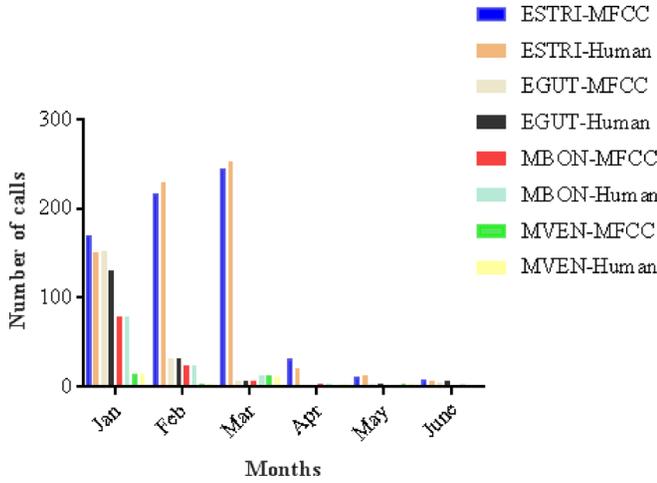
FIG. 6. (Color online) Monthly sum of files with fish sounds from the human classification and the automated classification based on MFCC at Bajo de Sico. ESRTI = *E. striatus*; EGUT = *E. guttatus*; MBON = *M. bonaci*; MVEN = *M. venenosa*.

FIG. 7. (Color online) Monthly sum of files with fish sounds for *E. guttatus* from the human classification and the automated classification based on MFCC at ALS.

regularization parameter $\beta\, (0 < \beta < 1)$ controls the weight of the ℓ_1-norm. Equation (17) is known as the LASSO problem and can be modified to impose a mixture of an ℓ_1-norm and ℓ_2-norm constraints on x and is given below:

$$\min_x \|\mathbf{y} - A\mathbf{x}\|_2^2 + (1 - \beta)\|\mathbf{x}\|_1. \tag{18}$$

Equation (18) is known as the elastic net problem. The ℓ_1 term enforces the sparsity of the solution and the ℓ_2 penalty has a smoothing effect that stabilizes the obtained solution. Once the underdetermined linear system $\mathbf{y} = A\mathbf{x}$ is solved using Eqs. (16) or (17) or (18), the weight vector $\mathbf{x}$ can be used as a new feature extracted from the test vector $\mathbf{y}$ for classification purposes. The new feature vector x, should be as discriminative as possible between many classes (or groupers). Ideally, the new feature vector $\mathbf{x}$ should have non-zero entries associated with the class (grouper species) of test vector $\mathbf{y}$ as in Eq. (13). The modeling error can cause non-zero values at the entries of x other than those of the class of $\mathbf{y}$. Therefore, the contribution of individual class (or grouper species) in the dictionary A is represented by the test vector y, should be calculated in terms of residual error for classification purposes. The residual error R_i of the ith class is calculated by retaining the weights associated with that class in the vector $\mathbf{x}$ and setting all the other entries in x that are not associated with the ith class to zero. The residual error for the ith class is given in Eq. (19):

$$R_i = \|\mathbf{y} - A\mathbf{x_i}\|_2 \tag{19}$$

where

$$\mathbf{x_i} = \begin{bmatrix} 000\cdots \vdots \cdots \ldots \vdots \gamma_{i1}\ \gamma_{i2}\ \cdots\ \gamma_{iN} \vdots 0,0,\ldots,0 \vdots 0,0,0,\ldots \end{bmatrix}^T.$$

The residual error can be normalized as follows:

$$R_{\text{in}} = \frac{\|\mathbf{y} - A\mathbf{x_i}\|_2}{\|\mathbf{y}\|_2}. \tag{20}$$

V. EXPERMENTAL RESULTS

The results of human detection analysis were totaled by day by adding the files with positive grouper detections per species within a calendar day. These were then compared to the algorithm detections per species for the same dataset.
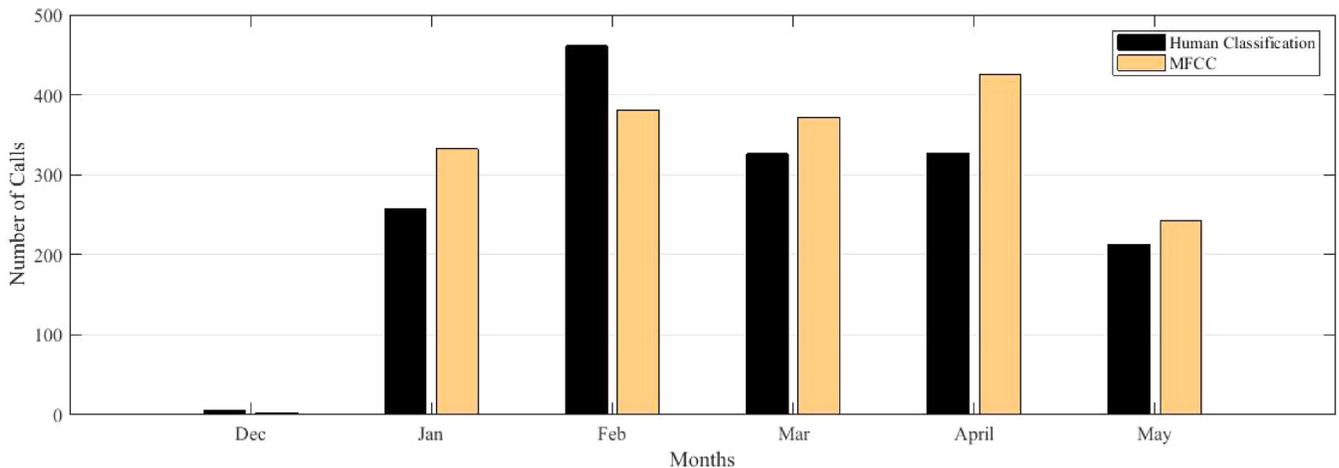
FIG. 8. (Color online) Monthly sum of files with fish sounds from the human classification and the automated classification based on MFCC at Mona Island for *M. venenosa*.

TABLE II. Classifications results based on MFCC at BDS.

	E. striatus	E. guttatus	M. venenosa	M. bonaci
January	170	152	78	14
February	217	31	24	4
March	245	6	7	13
April	31	1	4	2
May	11	3	2	3
June	8	5	2	0

The classification methods were evaluated in terms of classification results for each of the four grouper species. The identification results based on MFCC and Sparse classifier are shown in Figs. 6, 7, and 8 for BDS, ALS, and Mona Island, respectively.

The dominant species at BDS (Fig. 6), regardless of the month was E. striatus followed by E. guttatus, M. venenosa, and M. bonaci as shown in Table II. Figure 6 shows a small but consistent difference between MFCC and the human classification of O (10) mostly for E. striatus. At ALS (Fig. 7), the dominant species is E. guttatus, but the difference between human and MFCC based detection is quite significant (O (800)) at the peak of the aggregation in January.

At Mona Island (third site) the dominant species is M. venenosa (Fig. 8). The difference between the human classification and the automated classification based on MFCC is lower than that observed at ALS. It was about the same order of magnitude as BDS for E. striatus (not shown).

The following parameters were used to extract the acoustic features with MFCC at all sites: window length $= 0.1\,$s, number of MFCC features $= 13$, MFCC window overlapping $= 50\%$. WMFCC parameter $a_1 = 0.6$ and $a_2 = 0.34$ were used to combine dynamic features and reduce the complexity of the feature space. The percentage of average classification accuracy was calculated as $Tp/(Tp+Fn)$, where Tp is the number of true positive detections and Fn the number of false negative detection estimated from the human detection analysis. The accuracy with 95% confidence interval obtained by using MFCC features for the three datasets and all species are 77.56% ± 1.56%, 74.93% ± 2.46%, 76.46% ± 1.61%, for ALS, BSD, and Mona Island, respectively. The addition of dynamic characteristics of the sounds to MFCC improved the accuracy of MFCC as shown in Fig. 9. The percentages of average classification accuracy with the WMFCC features for the same three datasets became 82.81% ± 1.85%, 80.4% ± 2.29%, 83.64% ± 1.78% with 95% confidence interval, respectively. One-tail t-tests also confirmed with 95% confidence that the accuracy performance of WMFCC features is superior to that of the MFCC features. For both features extraction methods, the performance of the classification was not uniform between species. E. guttatus and E. striatus had the highest detection rate, while M. bonaci detection rate was 30% lower than the previous two (Fig. 9). This difference was however reduced by 15% with WMFCC based features extraction.

A similar comparison was conducted between MRAF and WMRAF (Fig. 10) and a similar improvement per species was obtained. As for MFCC and WMFCC, the two highest identification percentage in both methods was for E. guttatus and E. striatus. M bonaci had the lowest classification performance and the improvement due to the dynamics features was not as significant as for the WMFCC method (Fig. 10).

We now compare the results for all four features extraction methods, for each species in Table III. This shows that WMFCC is the best features extraction method for classification of E. guttatus and M. bonaci, with an accuracy of 86.1% and 67.3%, respectively. However, WMRAF has highest identification rate for classification of E. striatus, which is 85.3% and M. venenosa, which is 80.1%. These results suggest that the performance of the features extraction method is species dependent.

Finally using WMFCC as features extraction method, we evaluated the relative performance of the classifiers KNN, sparse and multiclass support vector machine (MSVM). Results in Table IV show that the sparse classifier outperformed in average the other classifiers for the classification of sounds of all four grouper species.

Moreover, a twofold cross validation procedure was adopted for 100 data files for each class to evaluate their
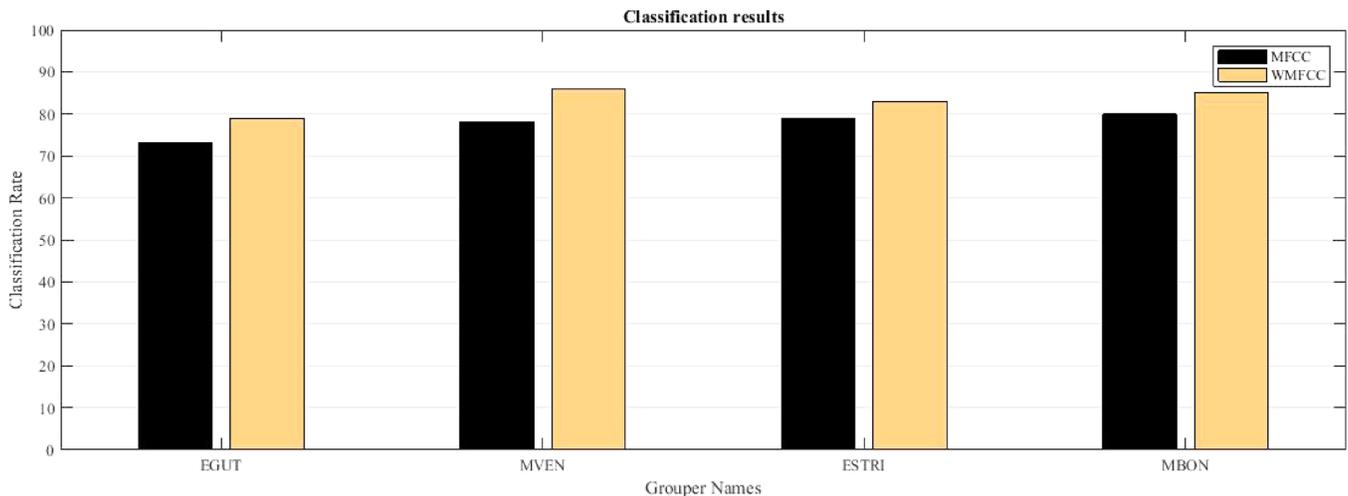
FIG. 9. (Color online) Accuracy of MFCC and WMFCC in sound classification per species. See Fig. 6 for the definition of species' names.
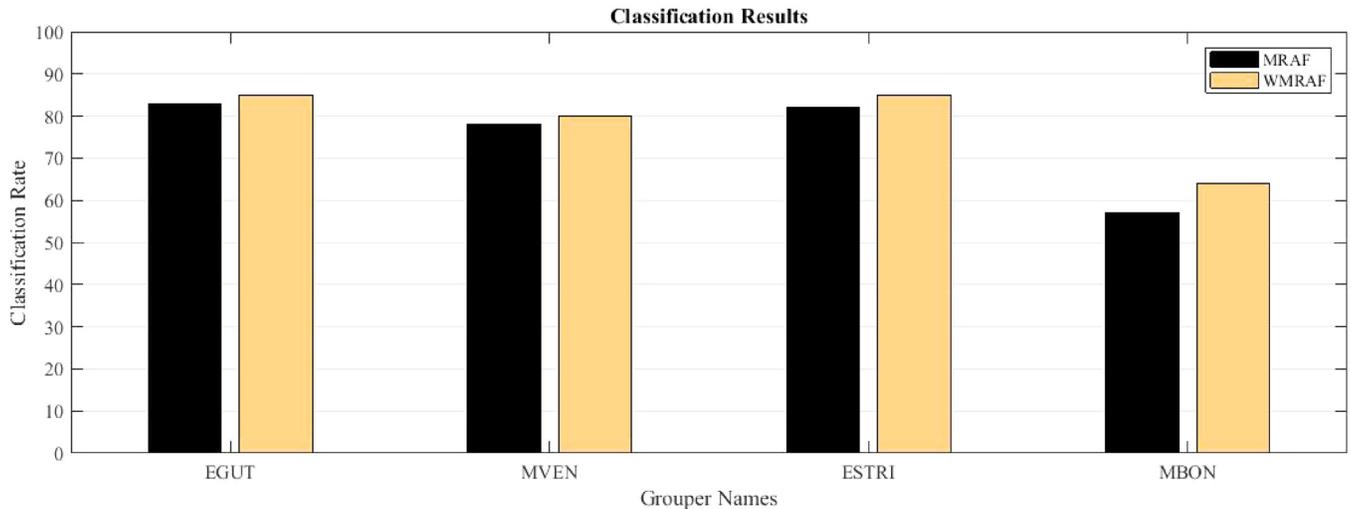
FIG. 10. (Color online) Accuracy of MRAF and WMRAF in sounds classification by species.

TABLE III. Classification results using the four features extraction methods, MFCC, WMFCC, MRAF, and WMRAF for sounds produced by all four species.

	EGUT	MVEN	ESTRI	MBON
MFCC	79.34%	73.64%	82.38%	54.81%
WMFCC	86.10%	78.30%	83.6%	67.25%
MRAF	83.46%	78.79%	82.17%	56.30%
WMRAF	85.31%	80.10%	85.3%	63.70%

TABLE IV. Comparison of different classifiers, KNN, Sparse and MSVN.

	EGUT	MVEN	ESTRI	MBON
KNN	87.30	63.60	71.60	48.30
Sparse	86.10	78.30	83.41	67.25
MSVM	85.74	72.14	80.19	63.49

TABLE V. Twofold cross validation for 100 data files per species, for four feature extraction methods.

	EGUT	MVEN	ESTRI	MBON
MFCC	86%	84%	78%	69%
WMFCC	91%	86%	82%	72%
MRAF	88%	81%	80%	70%
WMRAF	89%	81%	81%	70%

TABLE VI. WMFCC vs CMFCC + PCA with 100 dataset per species.

	EGUT	MVEN	ESTRI	MBON
WMFCC	91%	86%	82%	72%
CMFCC	96%	88%	78%	80%
CMFCC+PCA (33% dimension reduction)	94%	84%	72%	70%
CMFCC+PCA (67% dimension reduction)	88%	72%	84%	80%

respective influence on the performance of the classification algorithm and remove any bias due to a different number per class. Table V shows the results for all four feature extraction methods with the Sparse classifier. The absolute and relative accuracy increased for all classes, with the most significant benefit for *M. bonaci*. This confirms that *M. bonaci* remains the species the most difficult to digitally classify among the four (although the easiest to classify by humans because it is very different from the others), while *E. guttatus* classification seems to be the most accurate.

In addition to the overall better performance of WMFCC over WMRAF, there is one other advantage in using the WMFCC features extraction method over WMRAF. In contrast to the 2D cochleagram based approach in WMRAF, the features used in WMFCC are extracted from 1D time signals. Therefore the computational cost of the WMFCC algorithms is orders of magnitude lower, which makes it more suitable for real-time applications on energy limited autonomous platforms.

We also compared WMFCC with the concatenated MFCC (CMFCC) feature extraction algorithm. The CMFCC extractor lists MFCC, its derivative and double derivative in a single vector. Table VI provides the result using the sparse classifier. It can be seen that in terms of accuracy, CMFCC performs better in some cases and worse in other cases in comparison to WMFCC. We also used the principal component analysis (PCA) algorithm to reduce the dimension of the feature vector. Again after significant dimension reduction on CMFCC, the accuracy performance of the algorithm is still close to that of WMFCC.

VI. CONCLUSIONS

This paper introduces an automated detection and classification method for sounds produced by four grouper species, red hind (*E. guttatus*), Nassau (*E. striatus*), yellowfin (*M. venenosa*), and black (*M. bonaci*) grouper. The data were recorded at known spawning aggregation sites during the reproductive season, and most of the grouper vocalizations are associated with courtship behavior and consist of

tonal and pulse train calls. The acoustic dataset was first analyzed by human visual and auditory identification and then used as verification for the automated detection. We evaluated the performance of four acoustic features extraction methods. First MFCC and MRAF were evaluated on all dataset and compared. Then, the weighted dynamic acoustic features were applied to each method and their new accuracy compared. Experimental results have shown improved performance by WMFCC over MFCC features, and that this method outperformed MRAF and WMRAF methods overall. Further empirical studies revealed that the CMFCC algorithm also performs very well, even after a significant dimension reduction. The proposed methods transform ocean sounds and extract features in the time-frequency space. The Sparse representations classification of the WMFCC or CMFCC showed the best results over KNN and MSVN for all species. However, the features extraction methods used in this study consistently showed the same pattern of accuracy per species. *E. gutattus* and *E. striatus* were the most successfully classified species, while *M. venenosa* was slightly lower than the previous two and *M. bonaci* had the lowest accuracy rate of all. When the bias due to the different number of calls per class was removed, the difference between all four features extraction methods was reduced and *M. venenosa* became the second best classified (Table VI). But *M. bonaci* (*E. guttatus*) remained the least (most) accurate. All features extraction methods delivered almost identical results to the human detection, which provides confidence in the usefulness of such methods at classifying large datasets of ocean acoustic data. However, there were significant differences between the human and digital classification according to the aggregation site (Figs. 6, 7, and 8).

ACKNOWLEDGMENTS

[1]M. L. Domeier and P. L. Colin, "Tropical reef fish spawning and aggregations: Defined and reviewed," Bull. Mar. Sci. **60**(3), 698–726 (1997).

[2]P. J. Mumby, C. P. Dahlgren, A. R. Harborne, C. V. Kappe, F. Micheli, D. R. Brumbaugh, K. E. Holmes, J. M. Mendes, K. Broad, J. N. Sanchirico, K. Buch, S. Box, R. W. Stoffle, and A. B. Gill, "Fishing, Trophic cascades, and the process of grazing on coral reefs," Science **311**, 98–101 (2006).

[3]R. M. Starr, E. Sala, E. Ballesteros, and M. Zabala, "Spatial dynamics of the Nassau grouper *Epinephelus striatus* in a Caribbean atoll," Mar. Ecol.: Prog. Ser. **343**, 239–249 (2007).

[4]W. D. Heyman, B. Kjerfve, R. T. Graham, K. L. Rhodes, and L. Garbutt, "Spawning aggregations of *Lujanus cyanopterus* (cuvier) on Belize Barrier Reef over a 6 year period," J. Fish Biol. **67**, 83–101 (2005).

[5]W. D. Heyman, R. T. Graham, B. Kjerfve, and R. E. Johannes, "Whale sharks *Rhincodon typus* aggregate to feed on fish spawn in Belize," Mar. Ecol.: Prog. Ser. **215**, 275–282 (2001).

[6]R. Claro and K. C. Lindeman, "Spawning aggregation sites of snapper and grouper species (Lutjanidae and Serranidae) on the insular shelf of Cuba," Gulf Caribb. Res. **14**(2), 91–106 (2003).

[7]Y. Sadovy, "The case of the disappearing grouper: *Epinephelus striatus* (Pisces: Serranidae)," J. Fish Biol. **46**(6), 961–976 (1997).

[8]E. Sala, E. Ballesteros, and R. M. Starr, "Rapid decline of Nassau grouper spawning aggregations in Belize: Fishery management and conservation needs," Fisheries **26**(10), 23–30 (2001).

[9]C. L. Smith, "A spawning aggregation of Nassau grouper, *Epinephelus striatus* (Bloch)," Trans. Am. Fish Soc. **2**, 257–261 (2011).

[10]P. L. Colin and I. E. Clavijo, "Spawning activities of fishes producing pelagic eggs on a shelf edge coral reef, southwestern Puerto Rico," Bull. Mar. Sci. **43**(2), 249–279 (1988).

[11]D. A. Olsen and J. A. Laplace, "A study of a Virgin Islands grouper fishery based on a breeding aggregation," Proc. Gulf Caribb. Fish. Inst. **131**, 130–144 (1978).

[12]D. Y. Shapiro, "Reproduction in groupers," in *Tropical Snappers and Groupers: Biology and Fisheries Management*, edited by J. J. Polovina and S. Ralston (Westview Press, Boulder, CO, 1987), pp. 295–328.

[13]J. W. Tucker, P. G. Bush, and S. T. Slaybaugh, "Reproductive patterns of Cayman Islands Nassau grouper (*Epinephelus striatus*) populations," Bull. Mar. Sci. **52**(3), 961–969 (1993).

[14]A. Aguilar-Perera, "Preliminary observations of the spawning aggregation of Nassau grouper, *Epinephelus striatus*, at Mahahual, Quintana Roo, Mexico," Proc. Gulf Caribb. Fish Inst. **43**, 112–122 (1994).

[15]J. Carter and D. Perrine, "A spawning aggregation of dog snapper, *Lutjanus jocu* (*Pisces*: *Lutjanidae*) in Belize, Central America," Bull. Mar. Sci. **55**, 228–234 (1994).

[16]P. L. Colin, "Longevity of some coral reef fish spawning aggregations," Copeia **1996**, 189–192.

[17]C. C. Koenig, F. C. Coleman, L. A. Collins, Y. Sadovy, and P. L. Colin, "Reproduction in gag, *Mycteroperca microlepis* (Pisces: Serranidae) in the eastern Gulf of Mexico and the consequences of fishes spawning aggregations," in *Biology, Fisheries and Culture of Tropical Groupers and Snappers*, edited by F. Arreguin-Sanchez, J. L. Munro, M. C. Balgos, and D. Pauly, ICLARM Conf Proc. (1996), Vol. 48, pp. 307–323.

[18]Y. Sadovy, P. L. Colin, and M. Domeier, "Aggregation and spawning of the tiger grouper, *Mycteroperca tigris* (Pisces: Serranidae)," Copeia **1994**, 511–516.

[19]Y. Sadovy, A. Rosario, and A. Román, "Reproduction in aggregating grouper, the red hind, *Epinephelus guttatus*," Environ. Biol. Fishes **41**, 269–286 (1994).

[20]Y. Sadovy, "Reproduction of reef fishery species," in *Reef Fisheries*, edited by N. V. C. Polunin and C. M. Roberts, Chapman & Hall Fish and Fisheries Series (Springer Science+Business Media, Dordrecht, 1996), Vol. 20, pp. 15–59.

[21]P. L. Colin, W. A. Laroche, and E. B. Brothers, "Ingress and settlement in the Nassau grouper, *Epinephelus striatus* (Pisces: Serranidae), with relationship to spawning occurrence," Bull. Mar. Sci. **60**, 656–667 (1997).

[22]A. M. Eklund, D. B. McClennal, and D. E. Harper, "Black grouper aggregations in relation to protected areas within the Florida Keys National Marine Sanctuary," Bull. Mar. Sci. **66**(3), 721–728 (2000).

[23]Y. Sadovy, "The Nassau grouper 16 years on: Endangered, not just unlucky," in *59th Annual Gulf Caribb. Fish Inst.*, Belize City, Belize, Abstract 155 (2006).

[24]I. M. Kaatz, "Multiple sound-producing mechanisms in teleost fish and hypotheses regarding their behavioural significance," Bioacoustics **12**, 230–233 (2002).

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

[26]F. Ladich, "Sound production and acoustic communication," in *The Senses of Fish* (Springer, Berlin, 2004), pp. 210–230.

[27]R. Zelick, D. A. Mann, and A. N. Popper, "Acoustic communication in fish and frogs," in *Comparative Hearing: Fish and Amphibians* (Springer, Berlin, 1999), pp. 363–411.

[28]W. N. Tavolga, A. N. Popper, and R. R. Fay, *Hearing and Sound Communication in Fish* (Springer Science & Business Media, Berlin, 2012).

[29]R. O. Vasconcelos, P. J. Fonseca, M. C. P. Amorim, and F. Ladich, "Representation of complex vocalizations in the Lusitanian toadfish

J. Acoust. Soc. Am. **143** (2), February 2018

Ibrahim *et al.*   675

auditory system: Evidence of fine temporal, frequency and amplitude discrimination," Proc. Biol. Sci. **278**, 826–834 (2011).

[30] A. O. Kasumyan, "Sounds and sound production in fish," J. Ichthyol. **48**, 981–1030 (2008).

[31] T. J. Rowell, R. S. Appeldoorn, J. A. Rivera, D. A. Mann, T. Kellison, M. I. Nemeth, and M. T. Schäre, "Use of passive acoustics to map grouper spawning aggregations, with emphasis on red hind, *Epinephelus guttatus*, off western Puerto Rico," Proc. Gulf Caribb. Fish Inst. **63**, 139–142 (2011).

[32] E. D. Chesmore and E. Ohya, "Automated identification of field-recorded songs of four British grasshoppers using bioacoustic signal recognition," Bull. Entomol. Res. **94**, 319–330 (2004).

[33] A. K. Ibrahim, H. Zhuang, N. Erdol, and A. Muhamed Ali, "A new approach for North Atlantic Right Whale up-call detection," in *IEEE International Symposium on Computer, Consumer and Control*, Xi'an China (2016).

[34] D. K. Mellinger, S. W. Martin, R. P. Morrissey, L. Thomas, and J. J. Yosco, "A method for detecting whistles, moans, and other frequency contour sounds," J. Acoust. Soc. Am. **129**(6), 4055–4061 (2011).

[35] M. Hasson-Sandsten, "Classification of bird song syllables using singular vectors of multitaper spectrogram," in *3rd European Signal Processing Conference (EUSIPCO)* (2015), pp. 554–589.

[36] D. A. Mann, J. V. Locascio, M. T. Schärer, M. I. Nemeth, and R. S. Appeldoorn, "Sound production by red hind *Epinephelus guttatus* in spatially segregated spawning aggregations," Aquat. Biol. **10**, 149–154 (2010).

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

[38] M. T. Schärer, M. I. Nemeth, D. A. Mann, J. V. Locascio, R. S. Appeldoorn, and T. J. Rowell, "Sound production and reproductive behavior of Yellowfin grouper, *Mycteroperca venenosa* (Serranidae) at a spawning aggregation," Copeia **1**, 135–144 (2012).

[39] M. T. Schärer, M. I. Nemeth, T. J. Rowell, and R. S. Appeldoorn, "Sounds associated with the reproductive behavior of the black grouper (*Mycteroperca bonaci*)," Mar. Biol. **161**, 141–147 (2014).

[40] F. Sattar, S. Cullis-Suzuki, and F. Jin, "Acoustic analysis of big ocean data to monitor fish sounds," Ecol. Informat. **34**, 102–107 (2016).

[41] F. Sattar, S. Cullis-Suzuki, and F. Jin, "Identification of fish vocalizations from ocean acoustic data," Appl. Acoust. **110**, 248–255 (2016).

[42] C. Guyon, T. Bouwmans, and E. Zahzah, "Robust principal component analysis for background subtraction: Systematic evaluation and comparative analysis," in *Principal Component Analysis* (INTECH Publisher, London, 2012), pp. 223–238.

[43] T. Bouwmans and E. Zahzah, "Robust principal component analysis via decomposition into low-rank and sparse matrices: An overview," in *Handbook on Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing* (CRC Press, Boca Raton, FL, 2016).

[44] J. A. Perez, C. C. Y. Poon, R. D. Merrifield, S. T. C. Wong, and G. Z. Yang, "Big data for health," IEEE J. Biomed. Health Inf. **19**(4), 1193–1208 (2015).

[45] J. J. Noda, C. M. Travieso, and D. Sánchez-Rodríguez, "Automatic taxonomic classification of fish based on their acoustic signals," Appl. Sci. **6**, 443 (2016).

[46] Z. Weng, L. Li, and D. Guo, "Speaker recognition using weighted dynamic MFCC based on GMM," in *2010 International Conference on Anti–Counterfeiting, Security and Identification*, Chengdu (2010), pp. 285–288.

[47] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "Golden section search in one dimension," in *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge University Press, Cambridge, UK, 1992), p. 2.

[48] T. Chih, P. Ru, and S. Shamma, "Multiresolution spectrotemporal analysis of complex sounds," J. Acoust. Soc. Am. **118**, 887–906 (2005).

[49] D. Bremner, E. Demaine, J. Erickson, J. Iacono, S. Langerman, P. Morin, and G. Toussain, "Output-sensitive algorithms for computing nearest-neighbour decision boundaries," Discrete Comput. Geom. **33**(4), 593–604 (2005).

[50] V. Vapnik, *The Nature of Statistical Learning Theory* (Springer Verlag, New York, 1995).

[51] B. Scholkopf and A. J. Smola, *Learning with Kernels* (MIT Press, Cambridge, MA, 2002).

[52] C. J. Burges, "A tutorial on support vector machines for pattern recognition," Data Min. Knowl. Discovery **2**, 121–167 (1998).

[53] A. K. Ibrahim, H. Zhuang, N. Erdol, and A. Muhamed Ali, "EEG seizure detection by integrating slantlet transformation with sparse coding," in *EUSIPCO*, Greece, September 2017.

[54] A. Muhamed Ali, H. Zhuang, N. Erdol, and A. K. Ibrahim, "An approach for facial expression classification," Int. J. Biom. **9**(2), 96–112 (2017).