Ecological Informatics 34 (2016) 102–107

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Acoustic analysis of big ocean data to monitor fish sounds

Farook Sattar [a,*], Sarika Cullis-Suzuki [b], Feng Jin [c]

[a] Dept of Elect. & Comp. Eng., University of Waterloo, Waterloo, Ontario, Canada
[b] Ocean Networks Canada, University of Victoria, British Columbia, Canada
[c] Dept of Elect. & Comp. Eng., Ryerson University, Toronto, Ontario, Canada

ARTICLE INFO

Article history:
Received 9 February 2016
Received in revised form 7 May 2016
Accepted 9 May 2016
Available online 13 May 2016

Keywords:
Noisy big data
Ocean acoustics
Fish sound monitoring

ABSTRACT

This paper presents a novel framework for monitoring fish sounds based on acoustic analysis of noisy big ocean data. The proposed method involves multiresolution acoustic features (MRAF) extraction and RPCA (robust principal component analysis) based feature selection for monitoring of natural fish sounds produced *in situ* by the plainfin midshipman (*Porichthys notatus*); here, we investigate this fish's grunts, growls and groans. Both local and contextual information are exploited by MRAF, while sparse components of the MRAF matrix obtained through RPCA is found to be more robust to overlapping low-frequency spectral contents among different classes. The simulation results obtained from real-recorded ocean data reveal the advantages of the proposed scheme for monitoring underwater soundscapes and determining a variety of fish sounds in natural marine habitats.

1. Introduction

Soundscapes offer information about our surroundings and environments, and include natural and man-made sounds (Pijanowski et al., 2011). They can describe the biological communities that reside within, including information on species interactions (Fine and Thorson, 2008), communication (Tricas et al., 2006), and even abundance (Rountree et al., 2006). Underwater soundscapes and the interactions of organisms within them, including how marine life benefits from and exploits them are complex, and to date, poorly understood (Fay, 2009; Nedelec et al., 2015).

Contrary to previous assumptions, we now know that marine soundscapes are full of natural noise (Slabbekoorn et al., 2010), including 'biophony', noise produced by animals (Krause, 2012). Fish in particular produce a lot of noise, usually through contracting muscles or stridulation (rubbing of bones together; Kasumyan, 2008), and can make up the "natural acoustic background" heard underwater (Kasumyan, 2009; Slabbekoorn et al., 2010). Determining which sounds are created by which fish can be a daunting task as the number of fish that produce sounds is at least 800, and likely much more (Rountree et al., 2006; Kasumyan, 2008; Krause, 2012). When it comes to biological sounds, establishing those produced by marine life is a difficult process, as each sound must first be detected, identified, and then categorized (e.g. type), thus assuming a basic knowledge of each organism and its behavior in its natural environment. Such a process is generally done manually, which proves difficult, costly, and very time-intensive (Rountree et al., 2006).

Having an automatic classification system whereby fish sounds are categorized by species and vocalization type would allow large acoustic datasets to be analyzed over short timescales, and would yield information on marine soundscape composition and condition. For example, by identifying and classifying fish sounds, fish location (including spawning sites), migration patterns, abundance and other parameters could all be determined (Rountree et al., 2006).

The plainfin midshipman (*Porichthys notatus*) is a highly vocal species of fish found along the Pacific northeast coast. Also known as the 'singing fish', it is recognized for producing unique and varied sounds (Bass et al., 1999; Cullis-Suzuki, 2015). The 'hum' is by far its best understood call: associated with reproduction, the hum is emitted by alpha males in search of females who will come and mate (Brantley and Bass, 1994; Sisneros, 2009). The midshipman's other calls — the grunt, grunt train, growl and groan — are in comparison not well established. Unraveling the cause of vocalization emission and determining how the calls relate to temporal, spatial, and frequency features, would yield important insights into fish behavior and acoustic communication. For example, if growls were emitted as agonistic responses to predators, an increase in growls might signify a high abundance of predators. Further, these findings would influence our understanding of noise and how it might impact marine life and ocean soundscapes. We investigate the plainfin midshipman fish in this study, as it has proven to be an exceptional species for studies on wild fish communication (McIver et al., 2014; Cullis-Suzuki, 2016).

Other studies on other organisms have implemented a variety of automatic detection schemes for use on acoustic datasets. For example, an identification scheme is proposed in Chesmore and Ohya (2004) for Orthoptera species using temporal features based on shape of waveform and duration between consecutive zero-crossings followed by a multilayer perceptron (MLP) classifier. In Mellinger et al. (2011),

* Corresponding author.
E-mail address: farook_sattar@yahoo.com.sg (F. Sattar).

http://dx.doi.org/10.1016/j.ecoinf.2016.05.002
1574-9541/

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 in Hansson-Sandsten (2015), based on singular vectors of multitaper spectrogram and the similarity measures of two syllables using pairs of singular vectors. And finally, the study Starkhammar and Hansson-Sandsten (2015) presents an evaluation of different time-frequency representations for target detection applied to broadband echolocation signals of dolphins.

Here we focus on designing a robust acoustic analysis framework for big ocean data using robust principal component analysis (RPCA) and multiresolution acoustic features (MRAF) to monitor fish sounds. PCA uses the singular value decomposition (SVD) to find low rank representation of the data, while the robust version of PCA (RPCA) identifies a low rank representation, random noise, and a set of outliers by repeatedly calculating the SVD and applying "thresholds" to the singular values and error for each iteration (Guyon et al., 2012a; Bouwmans and Zahzah, 2016). The RPCA plays a significant role in tackling the key challenges involved with big data (Perez et al., 2015) by minimizing false alarms, reducing seasonal variability and processing the data that are not normally distributed. We further extract acoustic features using multiple window sizes (both 1D and 2D) from the same input data instead of the fixed window size (e.g. 20 ms). When multiple window sizes are used, multiple sets of feature vectors are derived for the same signal thereby increasing the number of examples. When features are extracted with multiple window sizes, the variations among the feature vectors are considerably increased, which will lead to better acoustic models. This multiresolution acoustic feature extraction technique can then be successfully used for building an efficient underwater monitoring system that can detect a variety of fish vocalizations automatically, thus providing information on the type and extent of communication taking place in underwater soundscapes.

The main contributions of this paper can be stated as: 1) We address the challenging task of fish sound monitoring from single channel audio when different fish vocalizations are overlapping. 2) The proposed MRAF are constructed based on both local and contextual information. 3) The proposed RPCA based feature selection is taken into account to reduce the non-distinctive features. 4) The advantage of the proposed two-stage scheme is that it provides a high performance for the input noisy raw data.

2. Method

The monitoring scheme we present is outlined in two steps: Firstly, we partition the raw hydrophone recordings (which contain fish sounds) into a number of segments. Secondly, we construct, select and use a set of multiresolution features as input to the MSVM classifier to track the types of fish vocalizations.

2.1. Data preprocessing

Manual spectrogram analysis involved examining the first 5 minutes of each hour in a day (i.e., 5 min × 24 h) for each of the three dates included in the analysis. The type and length of fish calls for each introductory 5-m segments were determined in Audacity (see Cullis-Suzuki, 2015; Sattar et al., 2016 for further details).

2.2. Multiresolution acoustic features

We have introduced a new multiresolution acoustic features (MRAF), which encodes the multi-resolution energy distributions in the time-frequency plan 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 different scales. 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

$$\begin{aligned} h(t) &= t^{(\eta-1)} e^{-2\pi B_{f_c} t} \cos(2\pi f_c t) \quad (t \geq 0) \\ &= 0 \quad (t \leq 0) \end{aligned} \tag{1}$$

where parameter η is the order of the filter, f_c denotes the center frequency while B_{f_c} refers to the bandwidth given f_c. 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 B_{f_c} and f_c is given by

$$B_{f_c} = 1.019 \times ERB(f_c) = 1.019 \times 24.7(4.37 \times f_c/1000 + 1). \tag{2}$$

Then each response signal from the gammatone filter bank is divided into 20 ms frames with a 10 ms frame shift; the cochleagram is obtained by calculating the energy of each time frame at each frequency channel. Each T-F unit in the cochleagram contains only local information, which may not be sufficient to accommodate the diversity in the ocean data. To compensate for this, the 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.

(1) Given input ocean data, compute the first 64-channel cochleagram (CB1) followed by a log operation applied to each T-F unit.
(2) Similarly, the second cochleagram (CB2) is computed with the frame length of 200 ms and frame shift of 10 ms.
(3) The third cochleagram (CB3) is derived by averaging CB1 using a rectangular window of size (5 × 5) including 5 frequency channels and 5 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).
(4) The fourth cochleagram CB4 is computed in a similar way to CB3, except that a rectangular window of size (11 × 11) is used.
(5) Concatenate CB1–CB4 to generate a feature matrix F and integrate it along the time frame to obtain a set of MRAF features of dimension (256 × 1).

2.3. Feature selection

The feature selection is motivated by the idea of decomposition of feature matrix into low-rank and sparse matrices based on the alternating direction method (ADM) (Yuan and Yang, 2009; Guyon et al., 2012b). It leads to feature selection via RPCA (robust principal component analysis) based on convex optimization. The RPCA is basically a matrix decomposition problem where it is assumed that the input feature matrix F is composed by a low-rank matrix, L, and a sparse matrix, S. Then the recovery of L and S matrix can be accomplished by solving the following convex programming problem:

$$\min_{S,L} \quad \gamma ||S||_{l_1} + ||L||_* \quad \text{subject to} \quad S + L = F \tag{3}$$

where $||\cdot||_{l_1}$ is the l_1 norm, $||\cdot||_*$ is the nuclear norm defined by the sum of all singular values, and γ is a positive regularization parameter.

The convex optimization problem in Eq. (3) can be solved by the ADM approach (Kontogiorgis and Meyer, 1998), which is easily implementable, computationally efficient using SVD (singular value decomposition) (Larsen, n.d.), and based on augmented Lagrangian

Table 1
The number of fish vocalization segments used for performance evaluation.

Day	Growl	Groan	Grunt
June 7, 2012	65	89	854
June 15, 2012	215	6	588
June 22, 2012	158	261	1344
Total	438	356	2786

optimization (Glowinski and Le Tallec, 1989). The corresponding augmented Lagrangian function is

$$\mathcal{L}(S,L,Z) = \gamma||S||_{l_1} + ||L||_* - \langle Z, S+L-F\rangle + \frac{\beta}{2}||S+L-F||^2. \tag{4}$$

In Eq. (4), $Z\in\mathcal{R}^{m\times n}$ is the multiplier of the linear constraint, $\langle\rangle$ is the trace inner product for matrix $\langle X,Y\rangle = trace(X^TY)$.

Then, the iterative scheme of ADM is

$$\begin{aligned} S^{k+1} &\in \operatorname{argmin}_{S\in\mathcal{R}^{m\times n}}\ \mathcal{L}\left(S, L^k, Z^k\right) \\ L^{k+1} &\in \operatorname{argmin}_{L\in\mathcal{R}^{m\times n}}\ \mathcal{L}\left(S^{k+1}, L, Z^k\right). \\ Z^{k+1} &= Z^k - \beta\left(S^{k+1} + L^{k+1} - F\right) \end{aligned} \tag{5}$$

We set the regularization parameter $\gamma = (tol/1\text{-}tol)$ and the parameter $\beta = 0.25nm/||F||_{l_1}$ $(F\in\mathcal{R}^{n\times m})$. Also, the value of $tol = 0.1$ is set within a suitable interval [0.03, 0.12] providing high-quality recovery (Yuan and Yang, 2009), while the iteration is stopped whenever the relative change $\mathrm{RC} = ||\frac{(S^{k+1},L^{k+1})-(S^k,L^k)||_{l_2}}{||(S^k,L^k)||_{l_2}+1} < tol$, where $||\cdot||_{l_2}$ denotes the Euclidean norm.

2.4. MSVM classification

Separating *grunts/growls/groans* in non-stationary noise, related to *P. notatus*, is a multiple classification based monitoring problem, which is solved here by considering all data in one optimization formulation based on the Crammer and Singer (CS) model (Crammer and Singer, 2001) for a multiclass support vector machine (MSVM) which has fast convergence and high accuracy (Dogan et al., 2011). In general, a MSVM classifier solves a d-class classification problem by constructing decision functions of the form:

$$x \mapsto \arg \min_{c\in\{1,\ldots,d\}} \{\langle w_c, \phi(x)\rangle + b_c\} \tag{6}$$

given *i.i.d.* training data $((x_1,y_1),\ldots,(x_l,y_l))\in(X\times\{1,\cdots,d\})^l$. Here, $\phi: X\rightarrow\mathcal{H}, \phi(x) = k(x,\cdot)$, is a feature map into a reproducing kernel Hilbert space $\mathcal{H}$ with corresponding kernel k, and $w_1\cdots, w_d\in\mathcal{H}$ are class-wise weight vectors. The CS machine is usually only defined for hypotheses without bias terms, that is, for $b_c = 0$. This CS based MSVM classifier is trained by solving the primal problem

$$\min_{w_c} \frac{1}{2}\sum_{c=1}^{d}\langle w_c, w_c\rangle + C\sum_{n=1}^{l}\eta_n \tag{7}$$

subject to

$$\forall n\in\{1,\cdots,l\}, \forall c\in\{1,\cdots,d\}\backslash\{y_n\}: \langle w_{y_n} - w_c, \phi(x_n)\rangle \geq 1-\eta_n$$

and

$$\forall n\in\{1,\cdots,l\}: \eta_n \geq 0$$

where η refers to the 'slack' variable. For learning structured data, CS's method is usually the MSVM algorithm of choice taking all class

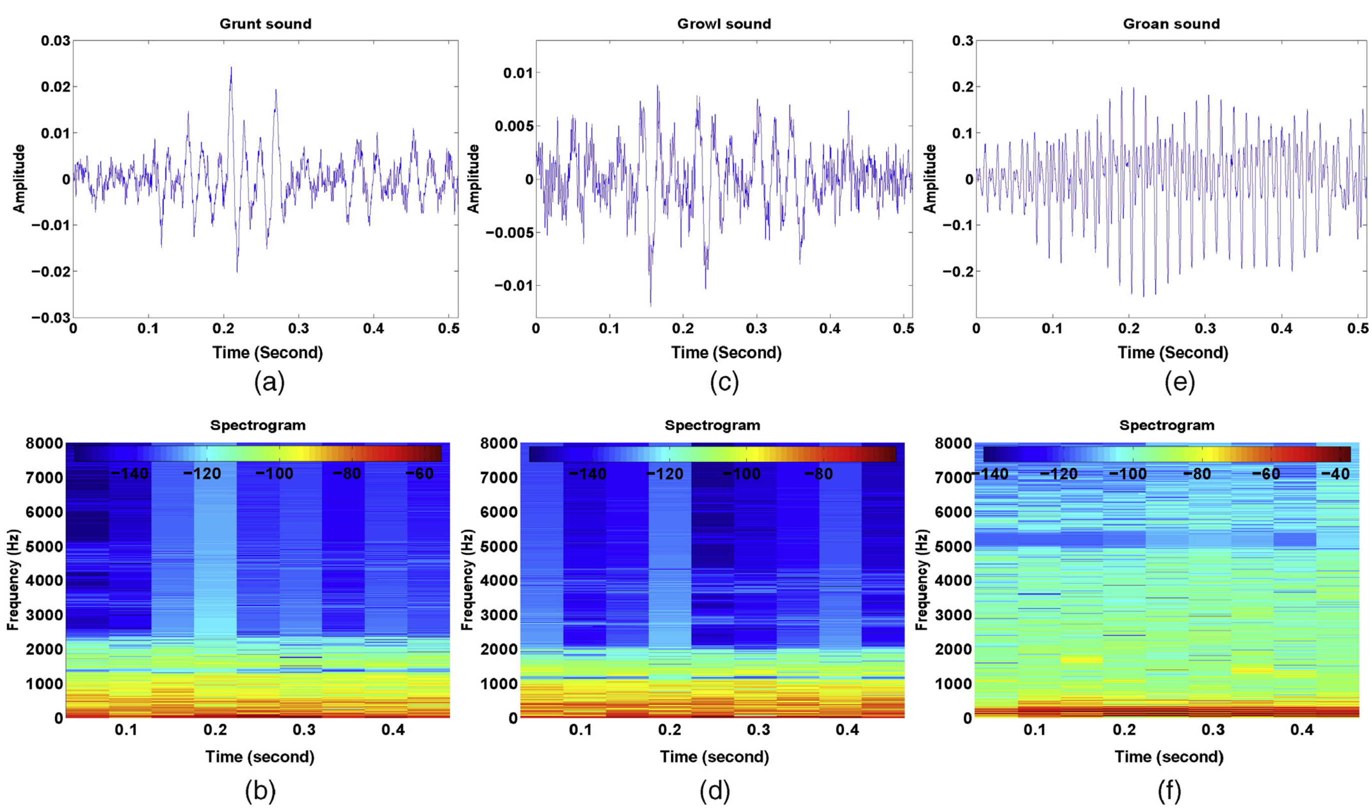

Fig. 1. Example of various typical segments and their spectrograms: (a)–(b) grunt; (c)–(d) growl; (e)–(f) groan.

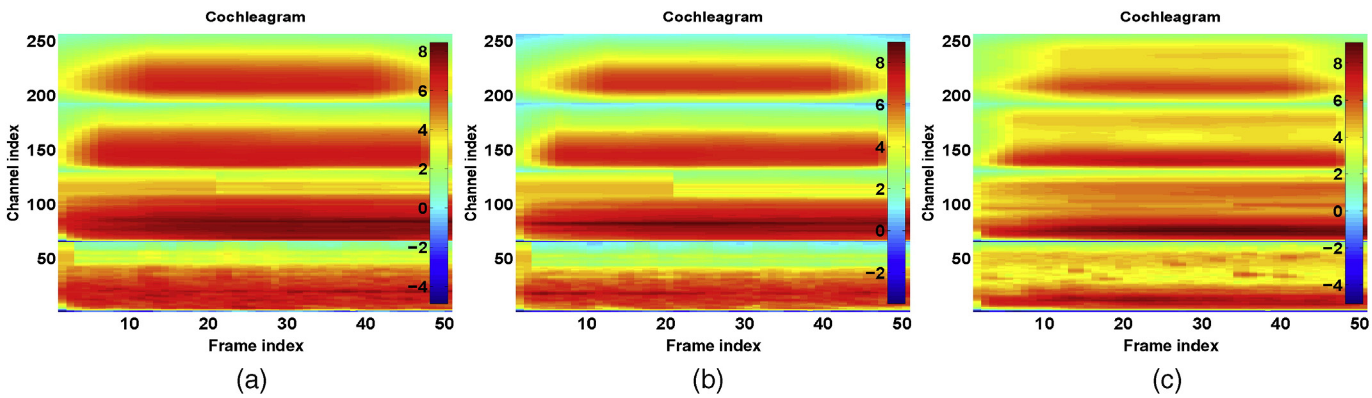

Fig. 2. Examples of cochleagrams (concatenated version: Channel indices [1:64] ∈CB1, [65:128] ∈ CB2, [129:192] ∈ CB3, [193:256] ∈ CB4) for the typical segments shown in Fig. 1; (a) grunt, (b) growl, (c) groan.

relations into account at once to solve a single optimization problem with fewer slack variables.

3. Experiment

3.1. Experimental dataset

Audio data obtained for this analysis were collected passively off a private dock on the east coast of Quadra Island, British Columbia, using an HTI-96-MIN hydrophone (Wildlife Acoustics, MA, USA) secured to the seafloor (as in Cullis-Suzuki, 2015; Sattar et al., 2016). All data were collected in June 2012 and three dates were selected for this paper: June 7th, 15th, and 22nd; all three fish calls—grunts, growls and groans—were present during these three days. The three continuous 24-hour signals have been adopted for performance evaluation. Manual spectrogram analysis of fish sounds was carried out in Audacity 2.0.6 (see Cullis-Suzuki, 2015 and Sattar et al., 2016 for more details). The resulting sound segments were also used for performance evaluation as depicted in Table 1. Some example spectrograms showing typical segments of different sound types are displayed in Fig. 1. The corresponding concatenated versions of the cochleagrams are also illustrated in Fig. 2.

3.2. Results and performance

The scatter plots of grunts, growls and groans are shown in Fig. 3. The corresponding correlation coefficients between PC 1 (principal component 1) and PC 2 (principal component 2) variables are 0.000024, 0.0098, and 0.00033 for the grunts, growls and groans, respectively, emphasizing the challenges in classification.

The proposed scheme is evaluated in terms of classification results for real recorded fish data. Results are obtained over 100 different runs in which the feature sets are split randomly by segment where 2/3 of the data are used for training and 1/3 of the data are retained for testing. In each case, the feature set is normalized to have zero mean and unit standard deviation. Here we have selected the default MSVM parameter C (regularization parameter) and γ (bandwidth parameter) of the radial Gaussian kernel $k(x;x') = \exp(-\gamma||x-x'||^2)$ as $C=10$ and $\gamma=2$ (Lauer, 2014).

Table 2 shows the performance of the proposed scheme in terms of classification accuracy (see Eq. (8)) given a feature set of size M, as derived from the fish data. The accuracy is defined as:

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{(\text{TP}+\text{FP})+(\text{TN}+\text{FN})} \tag{8}$$

where TP: True Positive, FP: False Positive, TN: True Negative, FN: False Negative. Here, as we see, accuracy (%) increases with M where M value is incremented by 64 to take into account the effects by integrating a different 64-channel cochleagram at each time.

Note that the June 15 data sets produce considerably better results than the other days. The possible reason can be attributed to its higher average SNR (signal-to-noise ratio), which is ≈7 dB, compared to the average SNRs for the June 7 and June 22 data sets, which are ≈5 dB and ≈3.5 dB, respectively.

For illustration, the confusion matrices of the MSVM classification for the three data sets (recorded on June 7, June 15, June 22) are presented in Table 3. As we can see, the mean accuracies (%) are high between 85 and 95% and vary with different days or data sets. Note that the sensitivity and specificity are very high for grunts, while they are relatively low for growls and groans due to the small number of data samples

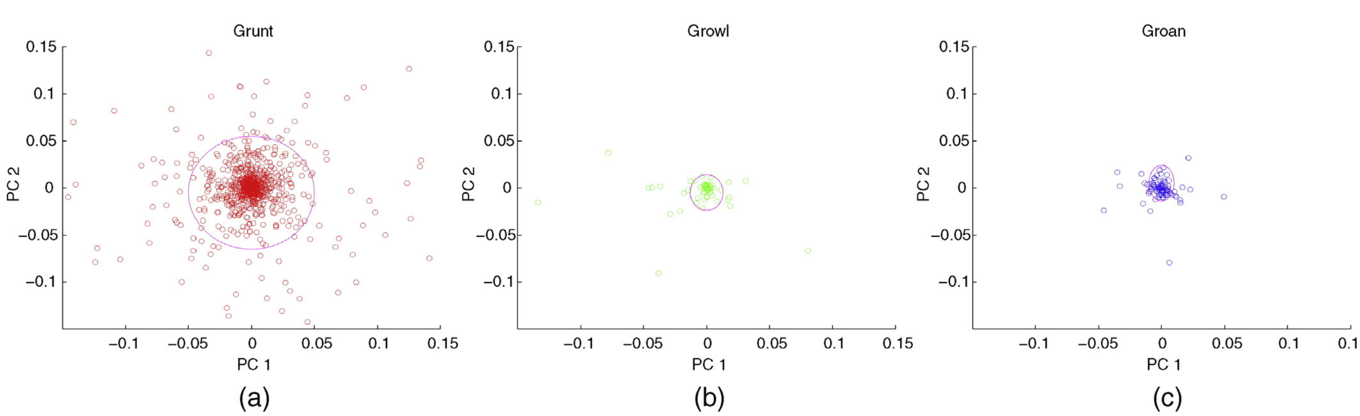

Fig. 3. Scatter plots of PC 1-vs-PC 2: (a) grunt; (b) growl; (c) groan.

Table 2
Percentage of average accuracy (classification rate) given M features derived from 24-hour fish data for various days/data sets (μ: mean, σ: standard deviation).

Day	M	Accuracy(%) $\mu \pm \sigma$
June 7, 2012	64	87.19 ± 1.48
	128	87.74 ± 1.41
	192	88.80 ± 1.54
	256	89.57 ± 1.51
June 15, 2012	64	94.03 ± 1.38
	128	94.31 ± 1.01
	192	95.02 ± 1.20
	256	95.06 ± 1.03
June 22, 2012	64	84.50 ± 2.19
	128	84.59 ± 1.12
	192	85.34 ± 1.75
	256	85.83 ± 1.33

(i.e. vocalizations) as well as low SNRs (average SNR of June 22 data; grunt: 5.10 dB, growl: 1.65 dB, groan: 3.40 dB). The results of the proposed method with the RPCA based feature selection are presented in Table 4, and show improvement of the average classification accuracy with smaller sized feature sets (c.f. Table 2) due to the selection of distinctive features. Note also that here we make use of feature-level fusion strategy (as performed by merging the calculated features from each source, i.e. cochleagram, into a cumulative structure), which gives better classification accuracy for 256 features with RPCA over the features without RPCA.

Table 3
Confusion matrices for the three data sets recorded on June 7, June 15 and June 22. (The classification accuracy as indicated in the right bottom corner (bold face) for the respective confusion matrix is calculated as $\frac{Sum\ of\ diagonal\ elements}{Sum\ of\ all\ elements}$).

Day		Grunt	Growl	Groan	Specificity
June 7, 2012	Grunt	258	0	0	1
	Growl	3	10	2	0.66
	Groan	11	18	1	0.03
	Sensitivity	0.95	0.35	0.5	**88.78**
June 15, 2012	Grunt	169	6	0	0.96
	Growl	5	59	0	0.92
	Groan	0	3	1	0.25
	Sensitivity	0.97	0.86	1	**94.24**
June 22, 2012	Grunt	398	1	2	0.99
	Growl	23	23	7	0.43
	Groan	34	10	31	0.41
	Sensitivity	0.87	0.67	0.77	**85.44**

Table 4
Percentage of average accuracy (classification rate) given M features derived and selected from 24-hour fish data for various days/data sets (μ: mean, σ: standard deviation).

Day	M	Accuracy (%) $\mu \pm \sigma$
June 7, 2012	64	93.66 ± 0.47
	128	93.36 ± 1.44
	192	93.36 ± 1.11
	256	94.08 ± 0.89
June 15, 2012	64	96.71 ± 0.75
	128	96.76 ± 0.84
	192	96.76 ± 0.94
	256	96.87 ± 0.77
June 22, 2012	64	90.46 ± 1.38
	128	90.41 ± 0.87
	192	89.99 ± 0.98
	256	90.57 ± 1.25

Table 5
Comparison results with MFCC features and three data sets recorded on June 7, June 15, June 22. (The classification accuracy as indicated in the right bottom corner (bold face) for the respective confusion matrix is calculated as $\frac{Sum\ of\ diagonal\ elements}{Sum\ of\ all\ elements}$).

Day		Grunt	Growl	Groan	Specificity
June 7, 2012	Grunt	252	2	1	0.98
	Growl	2	5	16	0.21
	Groan	7	8	10	0.40
	Sensitivity	0.96	0.33	0.37	**88.12**
June 15, 2012	Grunt	157	15	1	0.90
	Growl	10	56	3	0.81
	Groan	1	0	0	0
	Sensitivity	0.93	0.78	0	**87.65**
June 22, 2012	Grunt	367	16	14	0.92
	Growl	14	18	17	0.36
	Groan	12	22	49	0.59
	Sensitivity	0.93	0.39	0.61	**82.04**

3.3. Comparison results

Our results have been compared with the relevant method in Chesmore and Ohya (2004), since to the best of our knowledge, the later yields the best results among related studies in the literature. We determined that the average classification accuracies for the three data sets analyzed are much lower (≤50%) in Chesmore and Ohya (2004) than the method we present. This could be attributable to the susceptibility of the method in Chesmore and Ohya (2004) to noise and distortions, which produces a lower classification accuracy; such a result occurs due to its dependence on the zero-crossings and the local maximas of the input time-domain signal. It can be noted that the size of these temporal features (i.e. zero-crossings and local maximas) are varied and roughly related to the fundamental frequency and length of the input signal.

The comparison results with MFCC (mel frequency ceptral coefficients; Devi and Ravichandran, 2013) features are presented in Table 5. The following parameters are used: MFCC window length = 20 ms (320 samples), number of MFCC features = 12, MFCC window overlapping = 50%. The percentage of average classification accuracy ($\mu \pm \sigma$, μ: mean, σ: standard deviation) with the MFCC features for the three data sets are 86.78 ± 1.32, 87.93 ± 1.69, 84.46 ± 1.11, for June 7th, June 15th and June 22nd, respectively.

4. Conclusion

This paper introduces a new method for long-term monitoring of fish sounds using noisy ocean data. This framework is found to be effective in classifying the three types of fish sounds analyzed herein, a challenge due to the sounds' identical and overlapping spectral contents. Experimental results have shown improved performance by MRAF over MFCC features, and that this method outperforms the comparative methods. The feature selection by sparse representations of the MRAF further improved our results. Here we have proposed a supervised approach; in the next step of our work, an unsupervised approach will be developed for automatic grouping and labeling of different fish vocalizations and species.

Acknowledgments

The authors gratefully acknowledge the anonymous reviewers who enhanced this paper through constructive feedback.

References

Bass, A.H., Bodnar, D., Marchaterre, M., 1999. Complementary explanations for existing phenotypes in an acoustic communication system. In: Hauser, M.D., Konishi, M. (Eds.), The Design of Animal Communication. MIT Press, pp. 493–514.
Bouwmans, T., Zahzah, E., 2016. Robust principal component analysis via decomposition into low-rank and sparse matrices: an overview. Handbook on Robust Low-Rank

and Sparse Matrix Decomposition: Applications in Image and Video Processing. CRC Press, Taylor and Francis Group.

Brantley, R.K., Bass, A.H., 1994. Alternative male spawning tactics and acoustic signals in the plainfin midshipman fish *Porichthys notatus* Girard (Teleostei, Batrachoididae). Ethology 96, 213–232.

Chesmore, E.D., Ohya, E., 2004. Automated identification of field-recorded songs of four British grasshoppers using bioacoustic signal recognition. Bull. Entomol. Res. 94, 319–330.

Crammer, K., Singer, Y., 2001. On the algorithmic implementation of multiclass kernel-based vector machines. J. Mach. Learn. Res. 2, 265–292.

Cullis-Suzuki, S., 2015. Fish and ships: impacts of boat noise on the singing fish, *Porichthys notatus* (PhD Thesis) University of York, UK.

Cullis-Suzuki, S., 2016. Singing fish in an ocean of noise: effects of boat noise on the plainfin midshipman *Porichthys notatus* in a natural ecosystem. In: Popper, A.N., Hawkins, A. (Eds.), The Effects of Noise on Aquatic Life II, Advances in Experimental Medicine and Biology, Springer, pp. 171–179.

Devi, M.R., Ravichandran, T., 2013. A novel approach for speech feature extraction by cubic-log compression in MFCC. IEEE Conference on Pattern Recognition, Informatics and Mobile Engineering.

Dogan, U., Glasmachers, T., Igel, C., 2011. Fast training of multi-class support vector machines. Technical Report. Faculty of Science, University of Copenhagen, pp. 1–19.

Fay, R., 2009. Soundscapes and the sense of hearing of fishes. Integr. Zool. 4 (1), 26–32.

Fine, M.L., Thorson, R.F., 2008. Use of passive acoustics for assessing behavioral interactions in individual toadfish. Trans. Am. Fish. Soc. 137, 627–637.

Glowinski, R., Le Tallec, P., 1989. Augmented Lagrangian and operator splitting methods in nonlinear mechanics. SIAM Stud. Appl. Math.

Guyon, C., Bouwmans, T., Zahzah, E., 2012a. Robust principal component analysis for background subtraction: systematic evaluation and comparative analysis. Principal Component Analysis. INTECH Publisher, pp. 223–238.

Guyon, C., Bouwmans, T., Zahzah, E., 2012b. Foreground detection by robust PCA solved via a linearized alternating direction method. International Conference on Image Analysis and Recognition (ICIAR).

Hansson-Sandsten, M., 2015. Classification of bird song syllables using singular vectors of the multitaper spectrogram. 23rd European Signal Processing Conference (EUSIPCO), pp. 554–5589.

Kasumyan, A.O., 2008. Sounds and sound production in fishes. J. Ichthyol. 48 (11), 981–1030.

Kasumyan, A.O., 2009. Acoustic signaling in fish. J. Ichthyol. 49 (11), 963–1020.

Kontogiorgis, S., Meyer, R.R., 1998. A variable-penalty alternating directions method for convex optimization. Math. Program. 83, 29–53.

Krause, B., 2012. The great animal orchestra: finding the origins of music in the world's wild places. Little, Brown and Company, New York.

Larsen, R.M., . PROPACK-Software for large and sparse SVD calculations Available from http://sun.stanford.edu/srmunk/PROPACK/.

Lauer, F., 2014. MSVMpack: a multi-class support vector machine package. http://www.loria.fr/lauer/MSVMpack.

McIver, E.L., Marchaterre, M.A., Rice, A.N., Bass, A.H., 2014. Novel underwater soundscape: acoustic repertoire of plainfin midshipman fish. J. Exp. Biol. 217, 2377–2389.

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

Nedelec, S.L., Simpson, S.D., Holderied, M., Radford, A.N., Lecellier, G., Radford, C., Lecchini, D., 2015. Soundscapes and living communities in coral reefs: temporal and spatial variation. Mar. Ecol. Prog. Ser. 524, 125–135.

Perez, J.A., Poon, C.C.Y., Merrifield, R.D., Wong, S.T.C., Yang, G.-Z., 2015. Big data for health. IEEE J. Biomed. Health Inf. 19 (4).

Pijanowski, B.C., Villanueva-Rivera, L.J., Dumyahn, S.L., Farina, A., Krause, B.L., Napoletano, B.M., Gauge, S.H., Pieretti, N., 2011. Soundscape ecology: the science of sound in the landscape. Bioscience 61 (3), 203–216.

Rountree, R.A., Gilmore, R.G., Goudey, C.A., Hawkins, A.D., Luczkovich, J.J., Mann, D.A., 2006. Listening to fish: applications of passive acoustics to fisheries science. Fisheries 31 (9), 433–446.

Sattar, F., Cullis-Suzuki, S., Jin, F., 2016. Identification of fish vocalizations from ocean acoustic data. Appl. Acoust. 110, 248–255.

Sisneros, J.A., 2009. Seasonal plasticity of audotory saccular sensitivity in the vocal plainfin midshipman fish, *Porichthys notatus*. J. Neurophysiol. 102, 1121–1131.

Slabbekoorn, H., Bouton, N., van Opzeeland, I., Coers, A., ten Cate, C., Popper, A.N., 2010. A noisy spring: the impact of globally rising underwater sound levels on fish. Trends Ecol. Evol. 25 (7), 419–427.

Starkhammar, J., Hansson-Sandsten, M., 2015. Evaluation of seven time-frequency representation algorithms applied to broadband echolocation signals. Adv. Acoust. Vib. 2015, 1–13.

Tricas, T.C., Kajiura, S.M., Kosaki, R.K., 2006. Acoustic communication in territorial butterflyfish: test of the sound production hypothesis. J. Exp. Biol. 209 (24), 4994–5004.

Yuan, X., Yang, J., 2009. Sparse and low-rank matrix decomposition via alternating direction method. Pac. J. Optim.