The Journal of the Acoustical Society of America

SEPTEMBER 19 2018

Automatic classification of grouper species by their sounds using deep neural networks

Ali K. Ibrahim; Hanqi Zhuang; Laurent M. Chérubin; Michelle T. Schärer-Umpierre; Nurgun Erdol

Check for updates

J. Acoust. Soc. Am. 144, EL196–EL202 (2018)
https://doi.org/10.1121/1.5054911

View Online
Export Citation
CrossMark

ASA
ACOUSTICAL SOCIETY
OF AMERICA

Automatic classification of grouper species by their sounds using deep neural networks

Ali K. Ibrahim[a] and Hanqi Zhuang
Department of Computer & Electrical Engineering and Computer Science, Florida Atlantic University, 777 Glades Road, Boca Raton, Florida 33431, USA
Aibrahim2014@fau.edu, zhuang@fau.edu

Laurent M. Chérubin
Harbor Branch Oceanographic Institute, Florida Atlantic University, 5600 US1 North, Fort Pierce, Florida 34946, USA
lcherubin@fau.edu

Michelle T. Schärer-Umpierre
HJR Reefscaping, P.O. Box 1442, Boquerón 00622, Puerto Rico
michelle.scharer@upr.edu

Nurgun Erdol
Department of Computer & Electrical Engineering and Computer Science, Florida Atlantic University, 777 Glades Road, Boca Raton, Florida 33431, USA
erdol@fau.edu

Abstract: In this paper, the effectiveness of deep learning for automatic classification of grouper species by their vocalizations has been investigated. In the proposed approach, wavelet denoising is used to reduce ambient ocean noise, and a deep neural network is then used to classify sounds generated by different species of groupers. Experimental results for four species of groupers show that the proposed approach achieves a classification accuracy of around 90% or above in all of the tested cases, a result that is significantly better than the one obtained by a previously reported method for automatic classification of grouper calls.

[DRB]
Date Received: June 7, 2018 **Date Accepted:** September 1, 2018

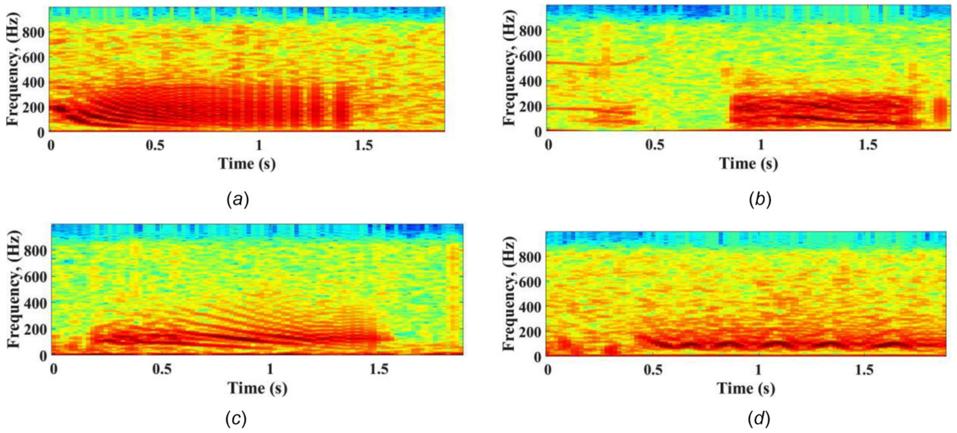
1. Introduction

Fish species produce sounds for multiple purposes, including courtship, navigation, and defending their territories from intruders.[1–5] Some groupers (fish family) produce courtship associated sounds (CAS) during spawning aggregation (Fig. 1) that are species specific. These sounds are in the 10–500 Hz frequency range and have distinctive characteristics as can be seen in sample spectrograms in Fig. 1. For instance, red hind (*E. guttatus*) calls are within the 100 to 200 Hz band.[6] The calls contain tonal segments that are produced at a variable pulse rate. Nassau grouper (*E. striatus*) calls consist of a pulse train with a varying number of short individual pulses and tonal sound in the 30 to 300 Hz band.[7] Yellowfin groupers (*M. venenosa*) produce calls similar to those of Nassau groupers, although they are longer in duration with frequencies ranging between 90 to 150 Hz.[8] Black groupers (*M. bonaci*) make at least two variations of frequency, modulated tonal calls between 60 and 120 Hz, but the calls have a longer duration than those of Nassau groupers.[9]

Passive acoustic monitoring (PAM) techniques have been used for many years to study the behavior of fishes.[10–14] A particular application of the PAM technique is to observe the reproductive cycles of fishes, including groupers. Many fish species swim long distances and gather in high densities for mass spawning at precise locations and times. This widespread reproductive strategy is typically shared among the groupers. Studying these spawning aggregations is vital to conservation efforts aimed at reversing worldwide depletion of endangered fishes and sustain marine biodiversity.

In an earlier work, we designed an automated classification algorithm, FADAR (Fish Acoustic Detection Algorithm Research), which is capable of identifying fours species of grouper in their natural environment with a classification accuracy around 82%.[15] FADAR consists mainly of three stages: signal denoising, feature

[a]Author to whom correspondence should be addressed.



(*a*)

(*b*)

(*c*)

(*d*)

Fig. 1. (Color online) Sound spectrograms for (a) red hind (*E. guttatus*); (b) Nassau grouper (*E. striatus*); (c) yellowfin (*M. venenosa*); and (d) black grouper (*M. bonaci*).

extraction, and classification. Although it is an automated approach, this machine learning approach still relies heavily on a carefully designed preprocessing and feature extraction method whose performance may degrade in low signal to noise ratio (SNR) environments.

Deep learning-based detectors and classifiers do not need sophisticated preprocessing and hand-crafted feature extraction procedures. It has been demonstrated in the literature that deep learning algorithms, such as autoencoders, convolutional neural networks (CNNs), and recurrent neural networks (RNNs), can act as feature extractors and classifiers.[16] CNNs are especially effective in identifying spatial patterns from images. On the other hand, RNNs are known to be capable of extracting discriminative patterns from time signals. However, the phenomenon of vanishing gradients prevents a standard RNN from memorizing long-term dependency of an input time sequence. Long short-term memory (LSTM) networks solve this problem by introducing parameters that selectively memorize or forget certain attributes of an input sequence.[17–20]

In this paper, we report the effectiveness of using CNNs and LSTM networks for classification of sounds produced by four species of grouper. We first describe the architecture of our solution to the problem undertaken. We then compare the new approach with the previously reported one using a grouper sound datasets collected off the west coast of Puerto Rico, in the Caribbean Sea. The experimental results confirm the hypothesis that a data-driven feature extractor, like the one proposed in this paper, can outperform with a large margin a hand-crafted one, like the one reported in Ref. 15.

2. Solution strategy

A diagram of the proposed grouper sound classification algorithm is given in Fig. 2. The LSTM network, as an example, is used to implement the deep learning layers. Initial denoising is performed in the discrete wavelet transform (DWT) domain. Denoised data points are, subsequently, processed by a sequence of LSTM layers. These layers produce discriminative features, which in turn are used to classify the incoming signals. The classifier eventually outputs the grouper species that produced the call in the first place.

2.1 Signal denoising with discrete wavelet transform

As has been mentioned, grouper calls are concentrated in the frequency range between 10 and 500 Hz. A band-pass filter is a simple way to remove unwanted noise outside of the signal frequency band. Hydrophone data are subject to a variety of in-band noise

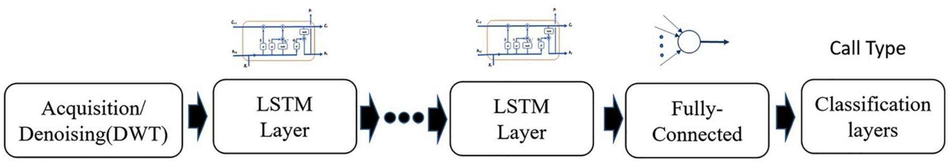


Fig. 2. (Color online) Grouper classification system by using deep learning.

sources. A prevailing low-frequency ocean ambient noise is generated by natural (wind, swells, currents) or anthropogenic sources (vessels, marine exploration, seismic air gun noise). Other marine mammals also contribute to the noise floor by the vestige of their vocalizations which are typically different, in frequency range and power, from sounds generated by groupers. Wavelets afford the flexibility of addressing the denoising problem in width-scaled bands. Wavelet denoising is an effective method for SNR improvement in environments with wide range of noise types competing for the same subspace.

DWT decomposes[21–25] signals over multiresolution subspaces making it a convenient tool for analyzing non-stationary signals. A cascade of filters and downsamplers separate the signal repeatedly into an approximation (low-pass) and a detail (high-pass) component. Denoising in the wavelet domain is accomplished by reducing the DWT detail components according to a threshold policy that is adjusted according to the frequency band or the multiresolution subspace. The general idea is that coefficients with amplitudes below a computed threshold are probably noise or can be removed without severely distorting the reconstructed signal. The noise threshold for the selected detail components is obtained by using decomposition level, which is dependent of $\lambda_j = \sigma_j^{\widehat{NOise}} \sqrt{2 \log_2 N_j}$, where N_j is the length of the jth detail component, and $\sigma_j^{\widehat{NOise}}$ is an estimate of the noise level. A hard or soft threshold function, defined below,[27–29] is then applied in an effort to remove the noise:

$$\widetilde{w_{j,i}} = \begin{cases} w_{j,i}, & |w_{j,i}| > \lambda_j, \\ 0, & |w_{j,i}| \le \lambda_j, \end{cases} \tag{1}$$

$$\widetilde{w_{j,i}} = \begin{cases} \mathrm{sgn}(w_{j,i}), & |w_{j,i} - \lambda_j|, \quad |w_{j,i}| > \lambda_j, \\ 0, & |w_{j,i}| \le \lambda_j, \end{cases} \tag{2}$$

where $w_{j,i}$ and $\widetilde{w_{j,i}}$ are noisy and denoised wavelet coefficients, respectively, at the jth decomposition level and the ith location of the detail component. Since the signal sampling frequency in this study is 10 kHz and the frequency range of grouper sound is below 500 Hz, the grouper signals are decomposed up to four levels. In our experiment, we tested both threshold methods and found out that the hard threshold method performed slightly better. The resulting coefficients are reconstructed by inverse DWT to obtain the denoised signals.

A typical DWT denoising result is shown in Fig. 3. The left graphs depict a call generated by red hind groupers and its spectrogram, and the right graphs illustrate the result of DWT denoising. It is evident from the graphs that DWT is quite effective in removing ocean ambient noise while preserving grouper calls.

2.2 LSTM

LSTM networks have been proved to be successful in addressing the problem of the vanishing gradients for RNNs.[19] The LSTM architecture consists of a set of recurrently connected subnets, known as memory blocks. Each block contains one or more self-connected memory cells and three multiplicative units—the input, output, and forget gates—which provide continuous analogues of write, read, and reset operations for the cells. A LSTM network is formed exactly like a simple RNN, except that the nonlinear units in the hidden layers are replaced by memory blocks.

A typical LSTM node is shown in Fig. 4.[20] In the figure, $\boldsymbol{x}_t$ is the input at time t, $\boldsymbol{h}_t$ the hidden state, $\boldsymbol{c}_t$ the cell state, and $\boldsymbol{y}_t$ the output. Furthermore, σ denotes a sigmoid function and tanh a hyperbolic tangent function. The multiplicative gates inside the block allow LSTM memory cells to store and access information over long periods of time, thereby avoiding the vanishing gradient problem. For instance, as long as the input gate remains closed (i.e., has an activation close to 0), the activation of the cell will not be overwritten by the new inputs arriving in the network and can therefore be made available to the network much later in the sequence, by opening the output gate.

Given an input data sequence $\boldsymbol{x} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$, a LSTM cell computes the hidden sequence $\boldsymbol{h} = \{\boldsymbol{h}_1, \ldots, \boldsymbol{h}_N\}$ and output sequence $\boldsymbol{y} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N\}$ by iterating the following equations from $t = 1$ to N (refer to Fig. 4):

$$\boldsymbol{f}_t = \sigma(W_f \boldsymbol{x}_t + U_f \boldsymbol{h}_{t-1} + \boldsymbol{b}_f), \tag{3}$$

$$\boldsymbol{i}_t = \sigma(W_i \boldsymbol{x}_t + U_i \boldsymbol{h}_{t-1} + \boldsymbol{b}_i), \tag{4}$$

$$\boldsymbol{o}_t = \sigma(W_o \boldsymbol{x}_t + U_o \boldsymbol{h}_{t-1} + \boldsymbol{b}_o), \tag{5}$$

$$\boldsymbol{p}_t = \tanh(W_c \boldsymbol{x}_t + U_c \boldsymbol{h}_{t-1} + \boldsymbol{b}_c), \tag{6}$$

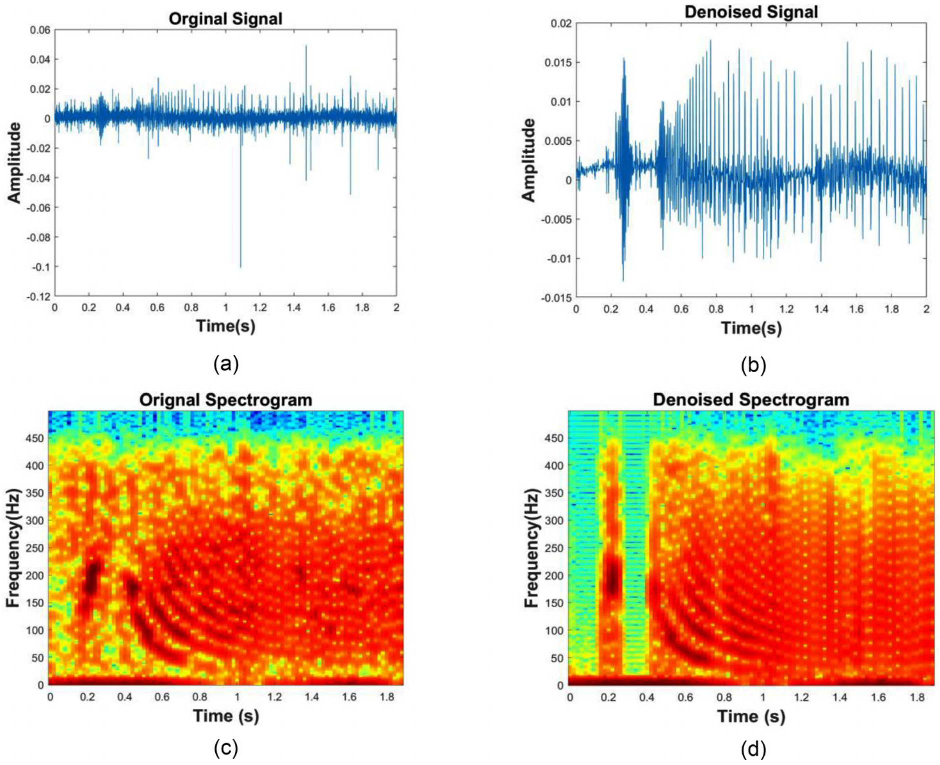


Fig. 3. (Color online) (a) Red hind (*E. guttatus*) sound; (b) red hind sound after DWT denoising; (c) red hind sound spectrogram; (d) denoised red hind sound spectrogram.

$$\boldsymbol{c}_t = \boldsymbol{f}_t \circ \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \circ \boldsymbol{p}_t, \quad (7)$$

$$\boldsymbol{y}_t = \boldsymbol{h}_t = \boldsymbol{o}_t \circ \tanh(\boldsymbol{c}_t), \quad (8)$$

where $\boldsymbol{i}$, $\boldsymbol{f}$, $\boldsymbol{o}$, and $\boldsymbol{p}$ are, respectively, the input gate, forget gate, output gate, and cell input activation vectors, $\boldsymbol{c}$ is a self-connected state vector, and $\circ$ denotes the Hadamard product.[20] Furthermore, the W and U terms denote weight matrices and the $\boldsymbol{b}$ terms bias vectors.

After the LSTM layers, a fully connected layer and a SoftMax layer are then used to determine the class of the input signal.

2.3 Convolutional neural networks

CNNs have established themselves as one of the best tools for classification. CNNs are made up of a number of CNN blocks. Each CNN block consists of a convolutional layer, an activation layer and a pooling layer. The convolution layer has a constant convolution window with small size that strides across the previous layer. After that, an activation function such as a ReLu is used to remove any negative values. And a pooling layer runs across the resulting vector to reduce the data size and improve its discrimination and representative capability. After a sequence of CNN blocks, the

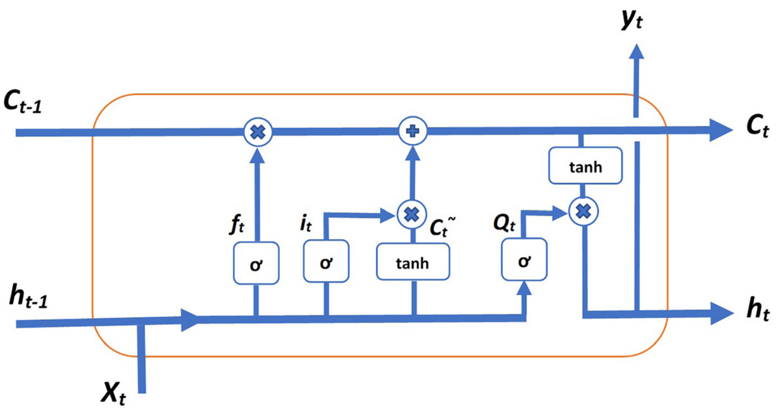


Fig. 4. (Color online) A LSTM network.

Table 1. Signal to noise ratio: input SNR = −12.44 dB.

Wavelet Type	Output SNR (dB)
haar	–6.00
Db3	−5.71
Db4	***−5.52***
Db5	−6.22
Db6	−6.03
Sym2	−5.73
Sym3	−5.60

result is finally fed to fully-connected layer and a SoftMax layer, from which the class of the input can be determined. For a detailed description of CNN, readers are referred to Ref. 26.

3. Experimental results

The dataset used in this research was recorded off the west coast of Puerto Rico at Abrir La Sierra (ALS),[7] Bajo de Sico (BDS) Bank,[7] and Mona Island.[8] Each one of these datasets contains 60 000 files, and the audio duration in each file is 20 s. The sampling rate for each signal is 10 kHz. Each sound file was labeled by a human operator inspecting the audio and the corresponding video files.[7–9] To implement the scheme depicted in Fig. 2, the first problem was the choice of wavelets for signal denoising. A comparative study was conducted to evaluate the performances of different types of wavelets, including Haar, symlets, and Daubechies. Table 1 indicates that these wavelets had similar results, though Daubechies (db4) produced outputs with the highest SNR therefore these were selected. Another design problem was the selection of the number of LSTM layers and the number of hidden units (nodes) for each LSTM layer in the architecture. The classification scheme, which was implemented in MATLAB in this study, consists of the following layers (after DWT denoising): an input layer, one to four LSTM layers, a fully connected layer, a SoftMax layer, and a classification layer.

Training and testing samples were randomly selected to avoid any bias. A five-fold training scheme was used, which means that 80% of the data samples was used to form a training set and the remaining sets were used for testing. The procedure was repeated five times so each of the data samples was used once for training and once for testing.

The number of nodes in the input layer equals the number DWT approximation coefficients. The number of output nodes in the SoftMax layer equals the number of classes, which is five in this study. The number of LSTM layers and the number of hidden units vary in the experimentation. The number of hidden units (N_h) can initially be set to the average of the number of inputs nodes and output nodes. Also, the number of LSTM layers (N_l) can be set to one initially for a small training set such as those we used in this study. The algorithm increases N_l and change N_h until classification accuracy does not improve significantly. In our experiment, the performance of the algorithm does not change much when N_l is increased from 2 to 4; see Table 2. In addition, the number of hidden units being 200 provides acceptable results. It has to be emphasized that using few parameters in the LSTM network, one can expect better generalizations of the network; on the other hand, with more parameters, the network fits data better while risking an overfit. Therefore, it is recommended that one should use a smaller number of hidden units and few hidden layers if the testing results do not degrade much.

For comparison, we investigated the performance of two CNN structures for this application. A natural choice is 1-D CNN with DWT approximate coefficients as its input. For the CNN models, we varied the number of convolution layers from 2 to 4, and found out that the 2-layer network performs as well as the CNNs with more

Table 2. Comparison of grouper classification accuracy.

Method	EGUT	MVEN	ESTRI	MBON
WMFCCs	86%	78%	84%	67%
LSTM networks	**96**%	**95**%	**93**%	**90**%
CNNs	94%	**95**%	91%	86%

Table 3. Grouper classification accuracy after balancing the classes.

Method	EGUT	MVEN	ESTRI	MBON
LSTM networks	87%	**97**%	**94**%	**96**%
CNNs	**88**%	91%	**94**%	95%

layers. In the CNN classifier, the input size is 10 003, and we used 32 filters of length 20×1. We also adopted ReLU as the activation function and applied Max pooling of stride 2.

Table 2 presents the experimental results obtained by both LSTM networks and CNNs. For a comparison study, those obtained by weighted mel-frequency cepstral coefficients (WMFCCs) are also included. In the table, EGUT, MVEN, ESTRI, and MBON stand for the grouper species *E. guttatus*, *M. venenosa*, *E. striatus*, and *M. bonaci*, respectively. Unlike the WMFCC method, for LSTM networks, DWTs were applied for signal denoising. It is evident that LSTM networks, which achieve an accuracy above 90% across the board, outperform WMFCCs significantly in every case. Furthermore, CNNs performs as well as LSTM networks. Note that we also designed a 2-D CNN classifier, in which case the 1-D input signals need to be transformed to 2-D spectrograms. The results obtained by 2-D CNNs are similar to those of 1-D CNNs, therefore these are omitted.

It must be emphasized that the results given in Table 2 are from an imbalanced dataset with 50% of recorded sounds produced by red hind, and less than 10% by black groupers. The rest is split between yellowfin and Nassau groupers. Therefore, the accuracies reported may not be convincing, especially for black groupers. In the next experiment, we applied a class balance procedure as follows. Due to the limited number of samples, we also shortened the signal length from 20 to 2 s. Each 2-s sound wave contains a maximum of one call only. In the training stage, we duplicated sounds of other species to have similar proportions to those of red hind. In the verification stage, we used 100 samples for each species. Table 3 shows the results from both LSTM and CNN classifiers. It is evident that both classifiers produced competitive results, although its efficiency by species varies.

4. Conclusions

In this paper, we have proposed a method for classifying grouper species by their vocalizations. In the proposed method, DWTs are used for signal denoising and deep neural networks are used for classification. We have compared the performances of the new method with those of the WMFCCs.[15] We also tested two deep learning methods: LSTM networks and CNNs. It was shown through experimental studies that the new approach outperforms significantly the previously reported technique. Furthermore, both the LSTM and CNN classifiers perform well in this application. It should be emphasized that the network structures of these classifiers are not optimized, therefore there may be room for further performance improvement. In future studies, we will implement the proposed algorithm in a real-time platform and explore the potential of applying the method to detect and classify other fish calls.

Acknowledgments

The authors acknowledge the Harbor Branch Oceanographic Institute Foundation for supporting this research. A.K.I., L.M.C., and M.T.S.U. were also supported in part by NOAA Saltonstall-Kennedy Grant No. NA15NMF4270329. Passive acoustic data were collected with the aid of the University of Puerto Rico, Mayagüez campus, the Caribbean Fishery Management Council funding for research, the Caribbean SEAMAP program and permits provided by the Department of Natural and Environmental Resources agency. We thank the crew of Orca Too as well as the volunteer divers and students that analyzed passive acoustic data, primarily Tim Rowell, Kimberly Clouse, Jesus Rivera, and Carlos Zayas. The authors would also like to acknowledge the contribution of an anonymous reviewer, who suggested using CNNs for the classification problem and predicted that CNNs would perform as well as LSTM networks for this application.

References and links

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

[2] 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).

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

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

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

[6] D. A. Mann, J. V. Locascio, M. T. Schärer-Umpierre, 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).

[7] M. T. Schärer-Umpierre, 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).

[8] M. T. Schärer-Umpierre, M. I. Nemeth, D. A. Mann, J. V. Locascio, R. S. Appeldoorn, and T. J. Rowell, “Sound production and reproductive behavior of yellow fin grouper, *Mycteroperca venenosa* (Serranidae) at a spawning aggregation,” Copeia **2012**, 135–144 (2012).

[9] M. T. Schärer-Umpierre, M. I. Nemeth, T. J. Rowell, and R. S. Appeldoorn, “Sounds associated with the reproductive behavior of the black grouper (*Mycteroperca bonaci*),” Marine Biol. **161**, 141–147 (2014).

[10] T. J. Rowell, R. S. Appeldoorn, J. A. Rivera, D. A. Mann, T. Kellison, M. Nemeth, and M. T. Schaürer-Umpierre, “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).

[11] 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).

[12] J. J. Luczkovich, D. A. Mann, and R. A. Rountree, “Passive acoustics as a tool in fisheries science,” Trans. Am. Fish. Soc. **137**, 533–541 (2008).

[13] T. H. Lin, H. Y. Yu, C. F. Chen, and L. S. Chou, “Passive acoustic monitoring of the temporal variability of odontocete tonal sounds from a long-term marine observatory,” PLoS One **10**, e0123943 (2015).

[14] R. Stolkin, S. Radhakrishnan, A. Sutin, and R. Rountree, “Passive acoustic detection of modulated underwater sounds from biological and anthropogenic sources,” in *OCEANS 2007* (2007), pp. 1–8.

[15] A. K. Ibrahim, L. M. Chérubin, H. Zhuang, M. T. Schärer-Umpierre, F. Dalgleish, N. Erdol, B. Ouynag, and A. Dalgleish, “An approach for automatic classification of grouper vocalizations with passive acoustic monitoring,” J. Acoust. Soc. Am. **143**(2), 666–676 (2018).

[16] C. Zhang, C. Yu, and J. H. L. Hansen, “An Investigation of deep-learning frameworks for speaker verification antispoofing,” IEEE J. Select. Topics Sign. Proc. **11**(4), 684–694 (2017).

[17] H. Sak, A. Senior, and F. Beaufays, “Long short-term memory recurrent neural network architectures for large scale acoustic modeling,” in *15th Annual Conference of the International Speech Communication* (2014), pp. 1–5.

[18] A. Graves, *Supervised Sequence Labelling with Recurrent Neural Networks* (Springer, Berlin, 2012), Vol. 385.

[19] F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, “Learning precise timing with LSTM recurrent networks,” J. Mach. Learn. Res. **3**, 115–143 (2003).

[20] S. Hochreiter and J. S. Huber, “Long short-term memory,” Neural Comput. **9**(8), 1735–1780 (1997).

[21] 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).

[22] S. Arivazhagan, W. S. Jebarani, and G. Kumaran, “Performance comparison of discrete wavelet transform and dual tree discrete wavelet transform for automatic airborne target detection,” in *IEEE International Conference on Computational Intelligence and Multimedia Applications* (2007), pp. 495–500.

[23] C. S. Burrus, R. A. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transform* (Prentice Hall, Upper Saddle River, NJ, 1998).

[24] P. Motlíček, “Feature extraction in speech coding and recognition,” research report, Oregon Graduate Institute of Science and Technology (2002), pp. 1–50.

[25] T. Kalayci, O. Ozdamar, and N. Erdol, “The use of wavelet transform as a preprocessor for the neural network detection of EEG spikes,” in *IEEE Southeast Conference 1994* (1994), pp. 1–4.

[26] Y. LeCun, Y. Bengio, and G. Hinton, “Deep Learning,” Nature **521**, 436–444 (2015).

[27] D. L. Donoho and I. M. Johnstone, “Ideal spatial adaptation via wavelet shrinkage,” Biometrika **81**(3), 425–455 (1994).

[28] I. Johnstone and B. Silverman, “Wavelet threshold estimators for data with correlated noise,” J. R. Stat. Soc., Ser. B **59**, 319–351 (1997).

[29] F. Chang, W. Hong, T. Zhang, J. Jing, and X. Liu, “Research on wavelet denoising for pulse signal based on improved wavelet thresholding,” in *Pervasive Computing Signal Processing and Applications (PCSPA)* (2010), pp. 564–567.