Transfer learning for efficient classification of grouper sound

Ali K. Ibrahim, Hanqi Zhuang, Laurent M. Chérubin, Michelle T. Schärer-Umpierre, Richard S. Nemeth, Nurgun Erdol, and Ali Muhamed Ali

Citation: The Journal of the Acoustical Society of America **148**, EL260 (2020); doi: 10.1121/10.0001943
View online: https://doi.org/10.1121/10.0001943
View Table of Contents: https://asa.scitation.org/toc/jas/148/3
Published by the Acoustical Society of America

JASA EXPRESS LETTERS

https://doi.org/10.1121/10.0001943

Transfer learning for efficient classification of grouper sound

Ali K. Ibrahim,[1,a] Hanqi Zhuang,[2] Laurent M. Chérubin,[1] Michelle T. Schärer-Umpierre,[3] Richard S. Nemeth,[4] Nurgun Erdol,[2] and Ali Muhamed Ali[2]

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

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

[3]*HJR Reefscaping, P.O. Box 1442, Boquerón, Puerto Rico 00622, USA*

[4]*Center for Marine and Environmental Studies, University of Virgin Islands, 2 John Brewers Bay, St. Thomas, US Virgin Islands 00802, USA*

aibrahim2014@fau.edu, zhuang@fau.edu, lcherubin@fau.edu, michelle.scharer@upr.edu, amuhamedali2014@fau.edu, rnemeth@uvi.edu, erdol@fau.edu

Abstract: A transfer learning approach is proposed to classify grouper species by their courtship-associated sounds produced during spawning aggregations. Vessel sounds are also included in order to potentially identify human interaction with spawning fish. Grouper sounds recorded during spawning aggregations were first converted to time-frequency representations. Two types of time frequency representations were used in this study: spectrograms and scalograms. These were converted to images, and then fed to pretrained deep neural network models: VGG16, VGG19, Google Net, and MobileNet. The experimental results revealed that transfer learning significantly outperformed the manually identified features approach for grouper sound classification. In addition, both time-frequency representations produced almost identical results in terms of classification accuracy.

[Editor: Paul Gendron] Pages: EL260–EL266

Received: 6 April 2020 **Accepted:** 24 August 2020 **Published Online:** 14 September 2020

1. Introduction

Many fish species swim long distances and gather in high densities for mass spawning at precise locations and times. This widespread reproductive strategy of forming spawning aggregations is typically shared among many of the species-rich family *Epinephelidae*, or groupers (Ma and Craig, 2018; Sadovy *et al.*, 1994). Declines in grouper populations due to overfishing at such aggregations have been documented widely (Sadovy and Domeier, 2005). Studying these spawning aggregations is vital to conservation efforts aimed at sustainable fisheries management and reversing worldwide depletion of endangered fishes. Some of these groupers are known to produce sounds for communication specially during courtship behaviors (Rowell *et al.*, 2012; Schärer *et al.*, 2012a, 2012b, 2014). Specifically, the acoustic characteristics and related courtship behavior of call types of four epinephelids common in the Caribbean Sea, the Nassau grouper (*Epinephelus striatus*), red hind (*E. guttatus*), black grouper (*Mycteroperca bonaci*), and yellowfin grouper (*M. venenosa*), have been well studied (Mann *et al.*, 2010; Rowell *et al.*, 2011, 2015; Schärer *et al.*, 2012a, 2012b, 2014; Zayas, 2019). These calls can be generally characterized by their low frequency band, which ranges between 20 and 400 Hz (Rowell *et al.*, 2012) and relatively low sound levels ranging between 121 and 165 dB re 1 μPa @ 1 m (Wilson *et al.*, 2020). The application of passive acoustic methods to locate and monitor the occurrence and abundance of spawning aggregations is key to understanding the importance of fish communications and developing indices of abundance that can be used in fisheries management and for the conservation of marine biodiversity (Jublier *et al.*, 2019, Chérubin *et al.*, 2020).

Passive acoustic monitoring generates large amount of data, which until recently was mostly analyzed manually by listening to the audio and visualizing the spectrograms. In addition to being a labor-intensive method, it is affected by differences in human perception, background noise interference, and the validation of descriptions of new sound types produced by the same species (Zayas, 2019). With the advent of machine learning, new automatic classification methods have been

[a]Also at: Department of Computer and Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, Florida 33431, USA. Author to whom correspondence should be addressed.

https://doi.org/10.1121/10.0001943

developed. For example, several manually identified acoustic features such as Mel-frequency cepstral coefficients (MFCC) and Multi resolution acoustic features (MRAF) were used by Ibrahim *et al.* (2018a) to classify grouper sounds of four species, namely *E. striatus*, *E. guttatus*, *M. bonaci*, and *M. venenosa*. In a following study, a deep learning method was applied by Ibrahim *et al.* (2018b) to classify the spectrograms of the same sounds. The latter method was shown to have over 90% accuracy detecting calls and outperformed the one based on manually identified acoustic features. While spectrograms, which are computed with short-time Fourier transformation, are a fixed-time and -frequency representation of signals, wavelet transform can incorporate multiple scales, and for this reason can locally reach the optimal time-frequency resolution with regards to the Heisenberg uncertainty (Daubechies, 1990). A scalogram is the time-frequency representation of a signal by wavelet transformation, where brightness or color can be used to indicate coefficient values at corresponding time-frequency locations. The effectiveness of both representations is considered herein.

Transfer learning is an active research area in machine learning that exploits the reusability of features learned by deep models such as GoogleNet, VGG16, VGG19, and AlexNet (Krizhevsky *et al.*, 2012; Simonyan and Zisserman, 2014; Szegedy *et al.*, 2015; Zhong *et al.*, 2020). Here the representation learned for a given input distribution or task is transferred to a different distribution or task. This study aims to apply the concept of transfer learning to effectively extract informative features from grouper sound spectrograms and scalograms. Our approach is conducted in three phases: 1) Scalograms and spectrograms are used to represent grouper calls. 2) Discriminative features are extracted from both types of time-frequency representations with pretrained Convolutional Neural Networks (CNNs), which are more efficient than manually designed CNNs. 3) An experimental study is conducted to evaluate the effectiveness of different pretrained networks using different time-frequency representations (scalograms versus spectrograms) for classifying grouper call types.

This paper is organized as follows. In Sec. 2, the sound transformation methods used to create spectrograms and scalograms are briefly introduced. In Sec. 3, the concept of transfer learning and popular pretrained Deep Neural Networks (DNNs) are introduced. Experimental results of classifying grouper call types according to each species with the transfer learning approach are presented in Sec. 4. Concluding remarks are provided in Sec. 5.

2. Sound transformation methods

In this study, spectrograms and scalograms were used to extract time-frequency representations of grouper sounds, which were then converted to RGB images. In this section, the wavelet transform, which is used to calculate scalograms, is introduced first, followed by scalograms and spectrograms of calls from different grouper species.

2.1 Wavelet transform

Wavelet transform is superior to Fourier transform and short time Fourier transform (STFT) for numerous applications because of the time-frequency resolution versatility it provides to measure time-frequency variations in a signal. Fourier transform contains globally averaged spectral information. Thus, the transient spectral information is lost. The Heisenberg boxes in the time-frequency domain illustrate the multiscale zooming property of wavelet transform, where coefficients of the higher frequency components of the signal have a shorter time step between them than those of the lower frequency components (Carmona *et al.*, 1998; Mallat, 1999).

A wavelet is a linear transform that decomposes an arbitrary signal $x(t)$ via basis functions that are simply dilations and translations of a parent wavelet $\psi(t)$:

$$W(a,t) = \frac{1}{|a|^{1/2}} \int_{-\infty}^{\infty} x(\tau)\psi^*\left(\frac{t-\tau}{a}\right) d\tau. \tag{1}$$

Dilation by the scale factor a controls the effective duration of the wavelet as it windows the signal, thus allowing for variations in time localization, which are accompanied by reciprocal variations in frequency localization. By this approach, a scalogram is able to represent a high-resolution signal in both time and frequency. This is in contrast to the use of constant duration windows of the Fourier transform. $W(a,t)$, the scalogram at time t and scale a, represents a measure of the similitude between the signal and the dilated/shifted parent wavelet. A local time-frequency energy density, which measures the energy of x in the Heisenberg box of each wavelet, is known as wavelet scalogram $\psi(t)$:

$$P_W x(\tau,a) = |W_x(a,\tau)|^2. \tag{2}$$

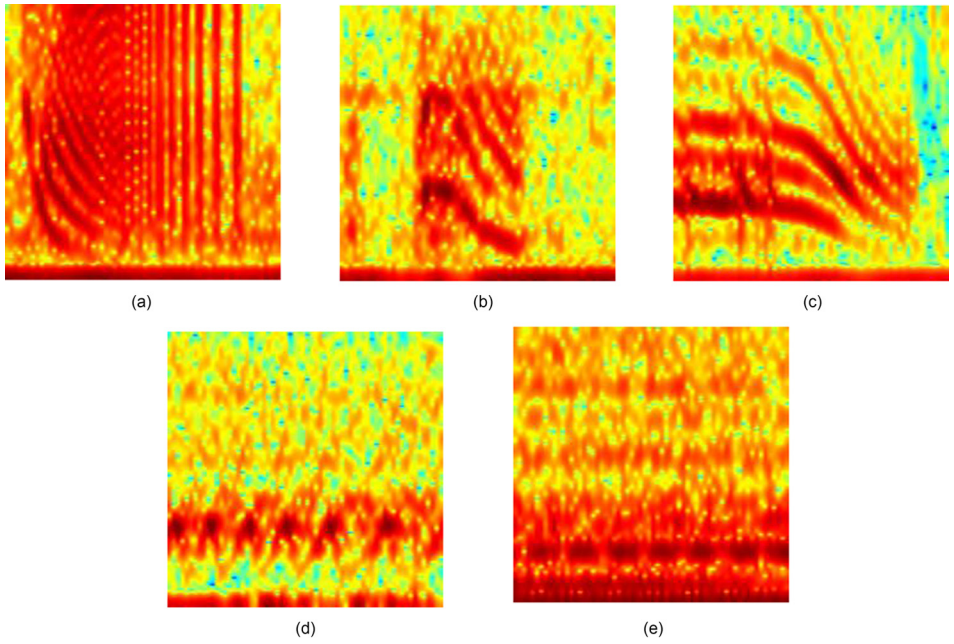

Fig. 1. (Color online) Examples of scalograms of sounds produced by four grouper species and a vessel. (a) *E. guttatus*, (b) *E. striatus*, (c) *M. venenosa*, (d) *M. bonaci*, and (e) vessel.

2.2 Scalograms

A scalogram is therefore the absolute value of the Continuous Wavelet Transform (CWT) of a signal, plotted as a function of time and frequency. This representation of the localized wavelet transform is well suited for the analysis of nonstationary phenomena, revealing the frequency content of the signal at each time step to pinpoint the occurrence of transients, while tracking evolutionary phenomena in both time and frequency.

Scalograms can be more useful than spectrograms for identifying signals with low-frequency components or rapidly changing frequency content. Unlike the spectrogram, which decomposes an input signal into sinusoids of infinite duration, CWT decomposes the signal into wavelets. To create scalograms, a CWT filter bank is pre-computed, which is a preferred method when the same

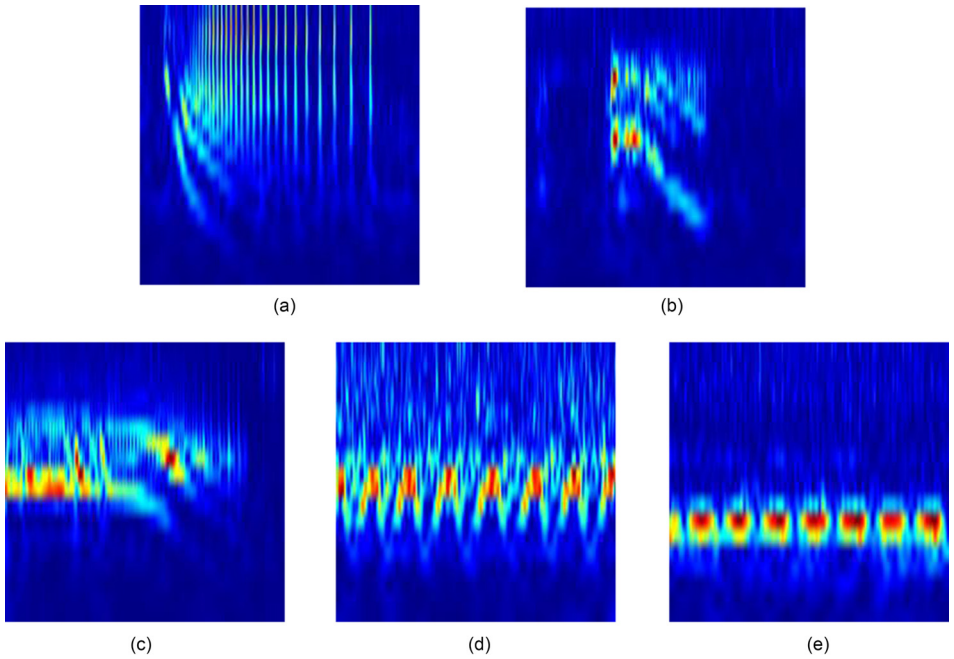

Fig. 2. (Color online) Examples of spectrograms of sounds produced by four grouper species and a vessel. (a) *E. guttatus*, (b) *E. striatus*, (c) *M. venenosa*, (d) *M. bonaci*, and (e) vessel.

https://doi.org/10.1121/10.0001943

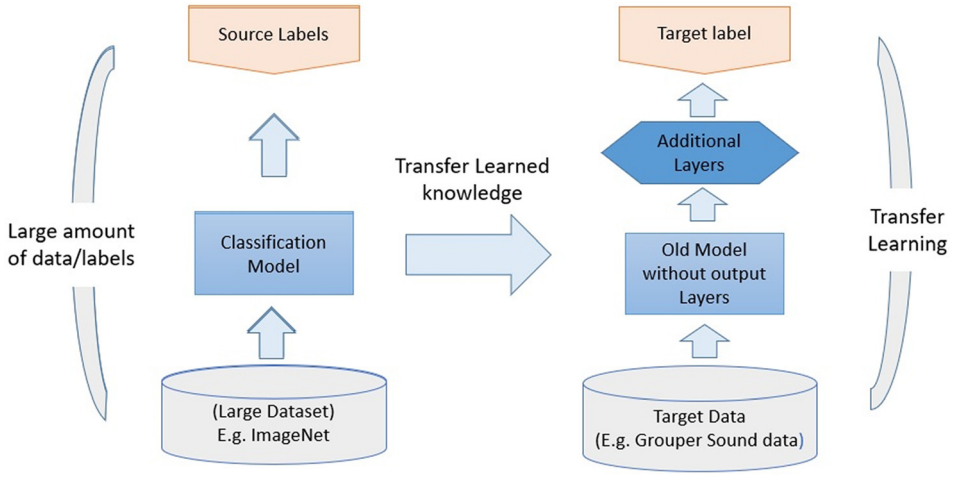

Fig. 3. (Color online) Illustration of the concept of transfer learning.

parameters are used to process different sound signals. The scalogram for each grouper sound is then computed by taking the magnitude of the wavelet transform with frequency range 10–400 Hz. Since all pretrained deep learning models used in this study originally took color images as input type, the magnitude spectra were converted to images using a MATLAB function ind2rgb (Fig. 1).

2.3 Spectrograms

A spectrogram is an indicator of a signal's time-varying spectral content over a range of frequencies. In the work reported here, spectrograms are generated by framing the grouper sounds with 0.1 s Hanning windows, or 1000 samples at a sample rate of 10 KHz, and 80% overlap. A 4096-point Fast Fourier transform (FFT) is applied to each frame. The spectrogram is estimated by taking the magnitude of each FFT frame with frequency range 10–400 Hz and is converted to an image (Fig. 2).

3. Pretrained deep neural network and transfer learning

Transfer learning relies on pretrained DNN models, which were trained in advance using a large number of images, and the trained models were distributed by their inventors for adoption. The application concept is described in Fig. 3. When any of these models is applied to a particular problem, additional layers of neurons are added to the existing model, and these additional weights are trained using application specific data points. In this study, the weights of these additional layers are trained using either spectrograms or scalograms of ocean sound segments. In essence, the pretrained DNN models serve as feature extractors of either spectrograms or scalograms of ocean sounds. The pretrained DNN plus the additional layers constitutes a complete classifier. Popular pretrained models include "Alexnet" (Krizhevsky *et al.*, 2012), VGG16 and VGG19 (Simonyan and Zisserman, 2014), Google Net (Szegedy *et al.*, 2015), ResNet, MobileNet (Howard *et al.*, 2017), and ShuffleNet. The structures of some of these DNN models are described next.

AlexNet is a CNN, which is trained on an ImageNet dataset. A CNN consists of an input layer, a number of hidden layers, and an output layer. Each hidden layer is usually made up of a convolutional layer with an activation function, a pooling layer, and a normalization layer. AlexNet has five convolutional layers. The kernel sizes of the convolutional layers are 11×11, 5×5, and 3×3. This network was initially trained to classify images into one thousand categories. AlexNet has 61 million weights.

Table 1. Number of layers and convolutional layers for each of pretrained model.

Method	Total Number of Layers	Number of Convolutional Layers	Number of Inception Layers	Number of Parameters (Millions)
GoogleNet	144	22	9	7
VGG16	41	16	0	138
VGG19	47	19	0	144
AlexNet	25	5	0	61
Inception V3	316	7	11	23.9
MobileNet	155	10	0	3.5

Table 2. Training and testing data for the call types of four groupers *E. guttatus* (EGUT), *E. striatus* (ESTRI), *M. venenosa* (MVEN), *M. bonaci* (MBON) and vessels.

Sound Type	Training Size	Testing Size
EGUT	7220	1805
ESTRI	5495	1373
MVEN	7571	1892
MBON	4000	1000
Vessel	5000	1000

VGG16, a type of CNN, has sixteen convolutional and five max-pooling layers. The size of the kernel filter is 3×3. The convolution stride is fixed to one pixel, and spatial padding is used to ensure that the spatial resolution is preserved after convolution. Spatial pooling is carried out by five max-pooling layers, which follow some of the convolutional layers. Max-pooling is performed over a 2×2 pixel window, and the stride is two.

VGG16 has 138 million weights. VGG19 has a similar structure as VGG16, except that it has 19 convolutional layers instead of 16. GoogleNet, also a type of CNN, uses a novel element called an inception module, which is designed to drastically reduce the number of parameters by concatenating the results of four different sizes of convolution and max-pooling operations and by adding auxiliary classifiers to combat the vanishing gradient problem that exists in a deep network. GoogleNet consists of 22 deep convolution layers with seven million parameters. Inception V3 evolves from the GoogleNet to improve further its classification performance by decomposing filter kernels and adding batch normalization, among other things. MobileNet, also introduced by Google, is a type of CNN that adopts depth-wise separable convolution to facilitate mobile applications. It reduces the number of parameters significantly without compromising much of the classification performance. The number of parameters for each of those pretrained DNNs is shown in Table 1.

To use pretrained deep neural network models for classification of grouper sound types, the last four layers of the pretrained models are replaced with four new layers that are suitable for our application (Fig. 3). These are two fully connected layers, one drop-out layer and one Softmax layer. The two new fully connected layers are introduced to learn the new features for the application. The dropout layer is used to prevent overfitting. The Softmax layer is served as a classification layer that outputs scores for each of the candidate classes.

4. Grouper call type classification with transfer learning

4.1 Dataset

The proposed classification method of grouper call types was tested with ambient acoustic recordings from different spawning aggregation sites, namely the Red Hind Bank (Nemeth, 2005) and the Grammanik Bank (Nemeth *et al.*, 2020) in the United States Virgin Islands, and at Abrir La Sierra, in Puerto Rico (Rowell *et al.*, 2012). Each recording 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 and memory space during six-month deployments. These recordings were made onto a SD memory card and downloaded as one.wav file for each 20 s recording. This cycle generates 288 files per day, which are stored in folders with 9999 files each. Each file can be heard with noise canceling headphones, and spectrograms can be visualized with an acoustic analysis software such as Adobe Audition or Audacity. Two second audio segments of four grouper courtship associated sound types, one for each species (*E. guttatus*, *E. striatus*, *M. bonaci*, *and M. venenosa*) and vessel sounds were collected from each of the files. Table 2 shows the number of training and testing files for all pretrained models. Because

Table 3. Accuracy per call type using different pretrained models with spectrograms.

Model	EGUT	ESTRI	MVEN	MBON	Vessel
GoogleNet	93.4%	79.9%	88.7%	99.4%	94.2%
VGG16	87.1%	81.6%	76.3%	94.2%	92.5%
VGG19	92.6%	**89.2**%	90.5%	99%	**98.0**%
AlexNet	**95.6**%	88.7%	**91.2**%	99.1%	95.2%
Inception V3	87.5%	84.8%	86.0%	**100**%	93.1%
MobileNet	89.8%	86.7%	88.6%	99%	96.8%

https://doi.org/10.1121/10.0001943

Table 4. Accuracy per call type using different pretrained models with scalograms.

Model	EGUT	ESTRI	MVEN	MBON	Vessel
GoogleNet	92.7%	87.1%	82.3%	**100**%	93%
VGG16	89.4%	80.5%	88.4%	94.6%	91.3%
VGG19	**95.1**%	**91.3**%	**92.9**%	99%	97.3%
AlexNet	93.3%	90.5%	92.2%	98.3%	**98.2**%
Inception V3	88.6%	85.3%	86.0%	**100**%	94.7%
MobileNet	90.4%	85.1%	90.8%	98%	96%

M. bonaci had fewer samples, data augmentation was implemented to increase the number of these samples to 5000 images by rotation and scaling.

4.2 Results and discussion

Training data were obtained by creating spectrograms and scalograms for each call type in the dataset. The epoch number was set to 100 for the training phase with the same mini patch size of 64 for all pretrained models in a NVIDIA DGX workstation. A 5-folder validation procedure was applied to assess the performance of each classifier. In this procedure, 80% of data samples were used to train the classifier and 20% for testing its performance. This procedure was repeated five times to ensure that every data sample was tested once. The performance of each model with spectrograms is presented in Table 3. Comparing the various models, AlexNet outperformed all other pretrained DNN models in classifying *E. guttatus* and *M. venenosa* sounds with accuracies of 95.6% and 91.2%, respectively (Table 3). VGG19 had the highest accuracy for *E. striatus* and vessels. Most models performed at the same level for *M. bonaci*, except for VGG16, which was the least accurate. Nonetheless, among all the call types of grouper species, *M. bonaci* calls received the highest level of accuracy, and *E. striatus* calls received the lowest accuracy.

Scalograms were also used to train all pretrained models (Table 4). It was evident that VGG19 outperformed all other pretrained models for classifying sound types for nearly all grouper species but was slightly less accurate for *M. bonaci* and vessel sounds. For *M. bonaci*, both GoogleNet and Inception V3 were the most accurate, and for vessel sounds, AlexNet was most accurate.

The comparison of spectrograms versus scalograms for sound classification suggests that the performance difference in terms of accuracy is not significant. With the same recordings as the ones used herein, the accuracy of the manually identified feature approach introduced in Ibrahim *et al.* (2018a) was only 67% for *M. bonaci*, which was the lowest among all species, and at best 82% for *E. guttatus*.

5. Conclusion

In this study, both the scalogram and spectrogram techniques were applied to extract time-frequency representations of the sounds produced by four grouper species during spawning aggregations. These representations were then used to train pretrained deep learning models with transfer learning. Here pretrained models acted as feature extractors, which computed deep features from spectrograms and scalograms. These models were then applied to classify courtship-associated sounds recorded during spawning aggregations. Vessel sounds were also included in this study to potentially identify sounds of anthropogenic sources interacting with fish aggregated to spawn. Classification results demonstrated higher levels of effectiveness and accuracy of the transfer learning approach compared to the manually identified feature approach. In addition to classification of grouper sounds, vessel sound identification by this type of analysis provides new opportunities to assess the use by boaters of areas designated as critical for fish spawning aggregations and the potential for this vessel engine noise in disturbing courtship-associated sounds and behaviors. This information can ultimately be used by fishery managers to estimate visitation or transit routes through marine protected areas that could be related to fishing activities in order to assess the impact of these activities on ocean resources.

Acknowledgments

The authors acknowledge the Harbor Branch Oceanographic Institute Foundation for supporting part of this research. Ibrahim, Chérubin, Schärer-Umpierre, and Nemeth were also supported in part by NOAA Saltonstall-Kennedy Grant No. NA15NMF4270329. In addition, Zhuang and Chérubin were partially supported by NSF MRI Grant No. 1828181. Passive acoustic data was 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 Puerto

Rico Department of Natural and Environmental Resources agency, and The University of the Virgin Islands' Center for Marine and Environmental Studies. We thank the crew of Orca Too as well as the volunteer divers and students that analyze passive acoustic data, primarily E. Tuohy, T. Rowell, K. Clouse, and C. Zayas. We also thank technical divers and support staff S. Heidmann, E. Kadison, I. Byrne, and S. Prosterman. This is contribution No. 219 of the University of the Virgin Islands' Center for Marine and Environmental Studies.

References and links

Carmona, R., Hwang, W., and Torresani, B. (**1998**). *Wavelet Analysis and Applications: Practical Time Frequency Analysis*, Vol. 9 (Academic Press, Cambridge, MA).

Chérubin, L. M., Dalgleish, F., Ibrahim, A., Schärer, M., Nemeth, R. S., and, Appeldoorn, R. (**2020**). "Fish spawning aggregations dynamics as inferred from a novel, persistent presence robotic approach," Front. Marine Sci. **6**, 779.

Daubechies, I. (**1990**). "The wavelet transform, time frequency localization and signal analysis," IEEE Trans. on Inform. Theor. **36**(5), 961–1005.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (**2017**). "Mobile nets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861.

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

Ibrahim, A. K., Zhuang, H., Chérubin, L. M., Schärer -Umpierre, M. T., and Erdol, N. (**2018b**). "Automatic classification of grouper species by their sounds using deep neural networks," J. Acoust. Soc. Am. **144**(3), EL196–EL202.

Jublier, N., Bertucci, F., Kever, L., Colleye, O., Ballesta, L., Nemeth, R. S., Lecchini, D., Rhodes, K. L., and Parmentier, E. (**2019**). "Passive monitoring of phenological acoustic patterns reveals the sound of the camouflage grouper, *Epinephelus polyphekadion*, at a spawning aggregation site in Fakarava atoll (French Polynesia)," Aquatic Conserv. Marine Freshwater Ecosys. **2019**, 1–11.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (**2012**). "ImageNet classification with deep convolutional neural network," in *Advances in Neural Information Processing Systems*, Vol. 25, pp. 1097–1105.

Ma, K. Y., and Craig, M. T. (**2018**). "An inconvenient monophyly: An update on the taxonomy of the grouper (*Epinephelide*)," Copeia **106**(3), 443–456.

Mallat, S. (**1999**). *A Wavelet Tour of Signal Processing* (Academic Press, Cambridge, MA).

Mann, D., Locascio, J., Schärer, M., Nemeth, M., and Appeldoorn, R. (**2010**). "Sound production by red hind *Epinephelus guttatus* in spatially segregated spawning aggregations," Aquat. Biol. **10**(2), 149–154.

Nemeth, R. S. (**2005**). "Population characteristics of a recovering US Virgin Islands red hind spawning aggregation following protection," Marine Ecol. Progress Ser. **286**, 81–97.

Nemeth, R. S., Kadison, E., Brown Peterson, N. J., and Blondeau, J. (**2020**). "Reproductive biology and behavior associated with a spawning aggregation of the Yellowfin Grouper *Mycteroperca venenosa*," Bull. Mar. Sci. **96**(1), 31–56.

Rowell, T., Appeldoorn, R., Rivera, J., Mann, D., Kellison, T., Nemeth, M., and Schärer Umpierre, M. (**2011**). "Use of passive acoustics to map grouper spawning aggregations, with emphasis on red hind, *Epinephelus guttatus*, off western Puerto Rico," Proc. Gulf Caribb. Fisheries Inst. **63**, 139142.

Rowell, T., Schärer, M. T., Appeldoorn, R. S., Nemeth, M. I., Mann, D. A., and Rivera, J. A. (**2012**). "Sound production as an indicator of red hind density at a spawning aggregation," Mar. Ecol. Prog. Ser. **462**, 241–250, https://www.int-res.com/abstracts/meps/v462/p241-250/.

Rowell, T., Nemeth, R. S., Schärer, M. T., and Appeldoorn, R. S. (**2015**). "Fish sound production and acoustic telemetry reveal behaviors and spatial patterns associated with spawning aggregations of two Caribbean groupers," Mar. Ecol. Prog. Ser. **518**, 239–254.

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

Sadovy, Y. J., and Domeier, M. L. (**2005**). "Are aggregation-fisheries sustainable? Reef fish fisheries as a case study," Coral Reefs **24**, 254–262.

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

Schärer, M. T., Nemeth, M. I., Rowell, T. J., and Appeldoorn, R. S. (**2014**). "Sounds associated with the reproductive behavior of the black grouper (*Mycteroperca bonaci*)," Marine Biol. **161**(1), 141–147.

Schärer, M. T., Rowell, T. J., Nemeth, M. I., and Appeldoorn, R. S. (**2012b**). "Sound production associated with reproductive behavior of Nassau grouper *Epinephelus striatus* at spawning aggregations," Endang. Species Res. **19**(1), 29–38.

Simonyan, K., and Zisserman, A. (**2014**). "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (**2015**). "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9.

Wilson, K. C., Semmens, B. X., Pattengill-Semmens, C. V., McCoy, C., and Širović, A. (**2020**). "Potential for grouper acoustic competition and partitioning at a multispecies spawning site off Little Cayman, Cayman Islands," Mar. Ecol. Prog. Ser. **634**, 127–146.

Zayas, C. (**2019**). "Red hind *Epinephelus guttatus* vocal repertoire characterization, temporal patterns and call detection with micro accelerometers," MS thesis. Department of Marine Science, University of Puerto Rico, Mayagüez, PR, 68.

Zhong, M., Castellote, M., Dodhia, R., Lavista Ferres, J., Keogh, M., and Brewer, A. (**2020**). "Beluga whale acoustic signal classification using deep learning neural network models," J. Acoust. Soc. Am. **147**(3), 1834–1841.