Cataloging fish sounds in the wild using combined acoustic and video recordings

Xavier Mouy, Rodney Rountree, Francis Juanes, and Stan E. Dosso

Citation: The Journal of the Acoustical Society of America **143**, EL333 (2018); doi: 10.1121/1.5037359
View online: https://doi.org/10.1121/1.5037359
View Table of Contents: https://asa.scitation.org/toc/jas/143/5
Published by the Acoustical Society of America

Cataloging fish sounds in the wild using combined acoustic and video recordings

Xavier Mouy[a)]
School of Earth and Ocean Sciences, University of Victoria, 3800 Finnerty Road, Victoria, British Columbia, V8P 5C2, Canada
xaviermouy@uvic.ca

Rodney Rountree[b)] **and Francis Juanes**
Department of Biology, University of Victoria, 3800 Finnerty Road, Victoria, British Columbia, V8P 5C2, Canada
rrountree@fishecology.org, juanes@uvic.ca

Stan E. Dosso
School of Earth and Ocean Sciences, University of Victoria, 3800 Finnerty Road, Victoria, British Columbia, V8P 5C2, Canada
sdosso@uvic.ca

Abstract: Although many fish are soniferous, few of their sounds have been identified, making passive acoustic monitoring (PAM) ineffective. To start addressing this issue, a portable 6-hydrophone array combined with a video camera was assembled to catalog fish sounds in the wild. Sounds are detected automatically in the acoustic recordings and localized in three dimensions using time-difference of arrivals and linearized inversion. Localizations are then combined with the video to identify the species producing the sounds. Uncertainty analyses show that fish are localized near the array with uncertainties < 50 cm. The proposed system was deployed off Cape Cod, MA and used to identify sounds produced by tautog (*Tautoga onitis*), demonstrating that the methodology can be used to build up a catalog of fish sounds that could be used for PAM and fisheries management.

[WJL]
Date Received: January 11, 2018 **Date Accepted:** April 19, 2018

1. Introduction

Passive acoustic monitoring (PAM) of fish (i.e., monitoring fish in the wild by listening to the sound they produce) is a research field of growing interest and importance (Rountree *et al.*, 2006). The types of sounds fish produce vary among species and regions but consist typically of low frequency (<1 kHz) pulses and amplitude-modulated grunts or croaks lasting from a few hundreds of milliseconds to several seconds (Kasumyan, 2008). As is the case for marine mammal vocalizations, fish sounds can typically be associated with specific species and behaviors (Kasumyan, 2008). Consequently, temporal and spectral characteristics of these sounds in underwater recordings could identify, non-intrusively, which species are present in a particular habitat, deduce their behavior, and thus characterize critical habitats. Unfortunately, many fish sounds have not been identified which reduces the usefulness of PAM. Many studies carried out in laboratory settings attempt to catalog fish sounds (e.g., Širović and Demer, 2009; Hawkins and Amorim, 2000). However, behavior-related sounds produced in natural habitats are often difficult or impossible to induce in captivity (e.g., spawning or interaction with conspecifics; Rountree *et al.*, 2006). Consequently, there is a need to record and identify fish sounds in their natural habitat. Because there is no control over biological and environmental variables (e.g., number of fish vocalizing), *in situ* measurements are challenging and require accurate localization of the soniferous fish, both acoustically and visually (Rountree, 2008). Although numerous methods have been developed for the large-scale localization of marine mammals based on their vocalizations (see review in Zimmer, 2011), only a handful of studies have been published to date on the fine-scale localization of individual fish (Parsons *et al.*, 2009; Parsons *et al.*, 2010; Locascio and Mann, 2011). To our

[a)]Also at: JASCO Applied Sciences, 2305–4464 Markham Street, Victoria, British Columbia, V8Z 7X8, Canada. Author to whom correspondence should be addressed.
[b)]Also at: The Fish Listener, 23 Joshua Lane, Waquoit, MA 02536, USA.

knowledge, no studies combining underwater acoustic localization and video recording to catalog fish sounds have been published. This letter develops and demonstrates the use of a compact hydrophone and video-camera array designed to record fish sounds, localize the source (acoustically), and identify the species (visually).

2. Methods

2.1 Array and data collection

The acoustic components of the array developed here consist of six Cetacean Research C55 hydrophones, denoted H1–H6, placed on each vertex of an octahedron constructed from a foldable aluminum frame, as shown in Fig. 1. Each hydrophone is located approximately 1 m from the center of the octahedron (considered the origin of the array coordinate system) and is connected by cable to a TASCAM DR-680mkII multi-track recorder (TEAC Corporation, Japan) to collect data continuously at a sampling frequency of 48 kHz with a quantization of 16 bits. A downward-facing AquaVu II fishcam (Crosslake, MN) underwater video camera is attached to the central pole of the frame below the top hydrophone and records continuously on a Sony portable DVD recorder (model VRD MC6) during the acoustic recordings. A floating light is deployed on top of the frame to improve visibility for the video recordings.

Underwater acoustic and video data were collected with this array during the night of 18 October, 2010, off the Cotuit town dock in Cape Cod, MA (41° 36.969′ N, 70° 26.000′ W). The array was deployed on the sea bottom off the dock in 3 m of water, while both the video and acoustic recorders stayed on the dock. A chum can was also deployed on the sea bottom to attract fish. A total of 7.5 h of continuous acoustic and video data were collected. All data collected were processed after array recovery.

2.2 Automated detection of acoustic events

Acoustic events (transient signals) were detected automatically in recordings from hydrophone H2. First, the spectrogram of the recordings was calculated (4096-sample Blackman window zero-padded to 8192 samples for FFT, with a time step of 480 samples or 10 ms) and normalized from 5 to 2000 Hz using a split-window normalizer to increase the signal to noise ratio of acoustic events in the frequency band of typical fish sounds (Struzinski and Lowe, 1984, 4-s window, 0.5-s notch). Second, the spectrogram was segmented by calculating the local energy variance on a two-dimensional kernel of size 0.01 s by 50 Hz. Events were defined in time and frequency by connecting the adjacent bins of the spectrogram with a local normalized energy variance of 0.5 or higher using the Moore neighborhood algorithm (Moore, 1968). All acoustic events with a frequency bandwidth less than 100 Hz or with a duration less than 0.02 s were discarded. All detection parameters were empirically defined to capture acoustic events whose time and frequency properties correspond to typical fish sounds. An illustration of the detection process can be found in Riera *et al.* (2016).

2.3 Acoustic localization by linearized inversion

The time difference of arrival (TDOA) of acoustic events between hydrophone 2 and each of the other hydrophones was used to localize the sound source in three dimensions (3D). Given their low source levels, fish sounds are typically detectable for distances of a few tens of meters (Amorim *et al.*, 2015). In this case, the problem can be formulated by assuming that the effects of refraction are negligible and propagation

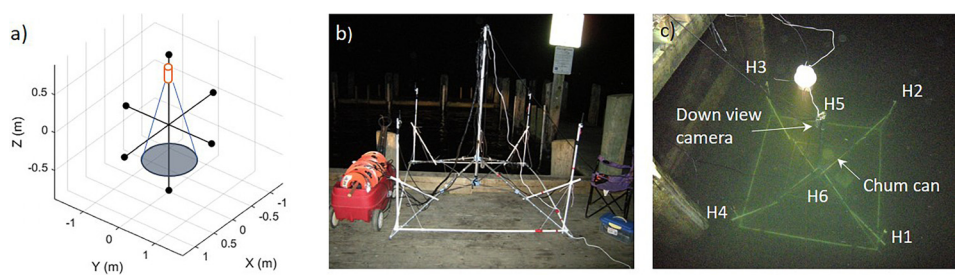

Fig. 1. (Color online) Array configuration. (a) Hydrophones (black dots) with the downward-looking camera (cylinder). (b) System on the dock before deployment. (c) System once deployed. H1–H6 indicate the locations of the hydrophones.

can be modeled along straight-line paths with a constant sound velocity v. The TDOA Δt_{ij} between hydrophones i and j is then defined by

$$\Delta t_{ij} = \frac{1}{v}\left(\sqrt{(X-x_i)^2+(Y-y_i)^2+(Z-z_i)^2} - \sqrt{(X-xj)^2+(Y-yj)^2+(Z-zj)^2}\right), \tag{1}$$

where x, y, z are the known 3D Cartesian coordinates of hydrophones i and j relative to the array center [Fig. 1(a)], and X, Y, Z are the unknown coordinates of the acoustic source ($M=3$ unknowns). The 6-hydrophone array provides measurements of a maximum of $N=5$ TDOA data, assuming the signal could be identified on all hydrophones. Localizing the acoustic source is a non-linear problem defined by

$$d_k = d_k(\mathbf{m}); \quad k = 1, \ldots, N, \tag{2}$$

where $\mathbf{d} = [\Delta t_{21}, \Delta t_{23}, \Delta t_{24}, \Delta t_{25}, \Delta t_{26}]^T$ represents the measured data and $\mathbf{d}(\mathbf{m})$ the modeled data with $\mathbf{m} = [\mathrm{X}, \mathrm{Y}, \mathrm{Z}]^T$ (in the common convention adopted here bold lower-case symbols represent vectors and bold upper-case symbols represent matrices). The expansion of Eq. (2) in a Taylor series to the first order about an arbitrary starting model $\mathbf{m}_0$ can be written

$$\mathbf{d} - \mathbf{d}(\mathbf{m}_0) = \mathbf{A}(\mathbf{m} - \mathbf{m}_0) \tag{3}$$

or

$$\delta\mathbf{d} = \mathbf{A}\,\delta\mathbf{m}, \tag{4}$$

where $\mathbf{A}$ is the $N \times M$ Jacobian matrix of partial derivatives with elements

$$\mathrm{A}_{ij} = \frac{\partial d_i(\mathbf{m}_0)}{\partial m_j}; \quad i = 1, \ldots, N; j = 1, \ldots, M. \tag{5}$$

This is an over-determined linear problem ($N=5$, $M=3$). Assuming errors in the data are identical and independently Gaussian distributed, the maximum-likelihood solution is

$$\delta\mathbf{m} = [\mathbf{A}^T\mathbf{A}]^{-1}\mathbf{A}^T\delta\mathbf{d}. \tag{6}$$

The location $\mathbf{m}$ of the acoustic source can be estimated by solving for $\delta\mathbf{m}$ and redefining iteratively

$$\mathbf{m}_{l+1} = \mathbf{m}_l + \alpha\,\delta\mathbf{m}; \quad l = 0, \ldots, L; \quad 0 < \alpha \leq 1, \tag{7}$$

until convergence (i.e., appropriate data misfit and stable $|\mathbf{m}|$). In Eq. (7), α is a step-size damping factor and L is the number of iterations until convergence. Localization uncertainties can be estimated from the diagonal elements of the model covariance matrix $\mathbf{C}_m$ about the final solution defined by

$$\mathbf{C}_m = \left[\mathbf{A}^T\mathbf{C}_d^{-1}\mathbf{A}\right]^{-1}, \tag{8}$$

where $\mathbf{C}_d = \sigma^2\mathbf{I}$ is the data covariance matrix with σ^2 the variance of the TDOA measurement errors and $\mathbf{I}$ the identity matrix. The 3D localization uncertainty is defined as the square root of the sum of the variances along each axis (diagonal elements of $\mathbf{C}_m$). All localizations were performed using the starting model $\mathbf{m}_0 = [0, 0, 0]^T$, a constant sound velocity $v = 1484$ m/s, and step size damping factor $\alpha = 0.1$.

The TDOAs in $\mathbf{d}$ were obtained by cross-correlating acoustic events detected on the recording from hydrophone 2 with the recordings from the other 5 hydrophones (search window: ± 2.5 ms). Before performing the cross-correlation, each recording was band-pass filtered in the frequency band determined by the detector using an eighth order zero-phase Butterworth filter (FILTFILT function in MATLAB, MathWorks, Inc., Natick, MA). Only detections with a sharp maximum peak in the normalized cross-correlation were considered for localization (peak correlation amplitude > 0.3, kurtosis > 14). The TDOA measurement errors were estimated by subtracting the measured TDOAs $\mathbf{d}$ at each hydrophone pair ($N=5$) from the predicted TDOAs $\mathbf{d}(\mathbf{m})$ for the estimated source location $\mathbf{m}$ using Eq. (1). The variance of the measurement errors σ^2 was then estimated as

$$\sigma^2 = \frac{1}{Q(N-3)}\sum_{i=1}^{Q}\sum_{j=1}^{N}\left(d_j^{(i)} - d(m)_j^{(i)}\right)^2, \tag{9}$$

where Q is the total number of acoustic events that were localized.

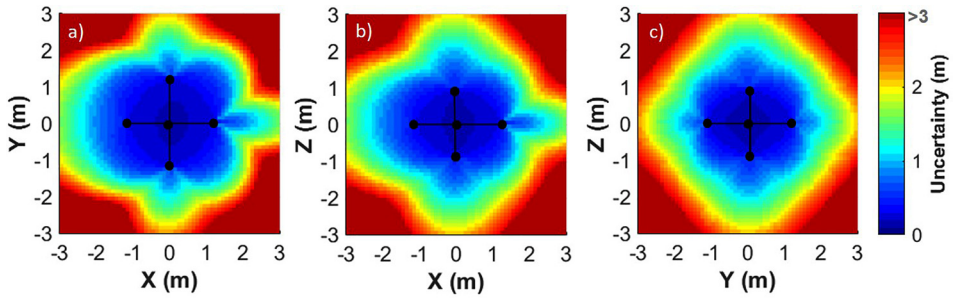

Fig. 2. (Color online) Localization uncertainties of the hydrophone array in the (a) XY, (b) XZ, and (c) YZ plane.

2.4 Video processing

To facilitate the visualization of fish in the video data, the recordings were processed to detect any movements that occurred in the camera's field of view. Each frame of the video recording was converted to a gray scale and normalized to a maximum of 1. An image representing the background scene was defined as the median of each pixel over a 5-min recording and was subtracted from each frame of the video. Finally, temporal smoothing was performed using a moving average of pixel values over 10 consecutive frames. Pixels with values greater than 0.6 were set to 1, and the others were set to zero. Each binarized image was overlaid in red on the original video image. All the processing of the acoustic and video data was performed using MATLAB 2017a (MathWorks, Inc., Natick, MA).

3. Results

This paper shows results from one 8-min data file. Out of the 185 acoustic events detected in this recording from hydrophone 2, 9 had a high enough cross-correlation peak with the other hydrophones to be localized. Other detections not selected for the localization stage were most often due to mechanical sounds from crabs crawling on the array frame or from sounds that were too faint to be received on all hydrophones. The standard deviation of the TDOA measurement errors was estimated to $\sigma = 0.12$ ms [Eq. (9), $Q = 9$]. The localization capabilities of the hydrophone array were assessed by calculating $\mathbf{C}_m$ and mapping the localization uncertainties of hypothetical sound sources located every 10 cm of a $3 \times 3 \times 3$ m cubic volume centered at [0,0,0] m. Figure 2 shows the localization uncertainties of the hydrophone array calculated for a 3D grid around the array using Eq. (8). The localization uncertainty in the middle of the water volume spanned by the arms of the array is less than 50 cm and increases progressively for sound sources farther from the center (Fig. 2). A 3D visualization of Fig. 2 is shown in Mm. 1 Localization uncertainties for sources outside the hydrophone array are generally greater than 1 m.

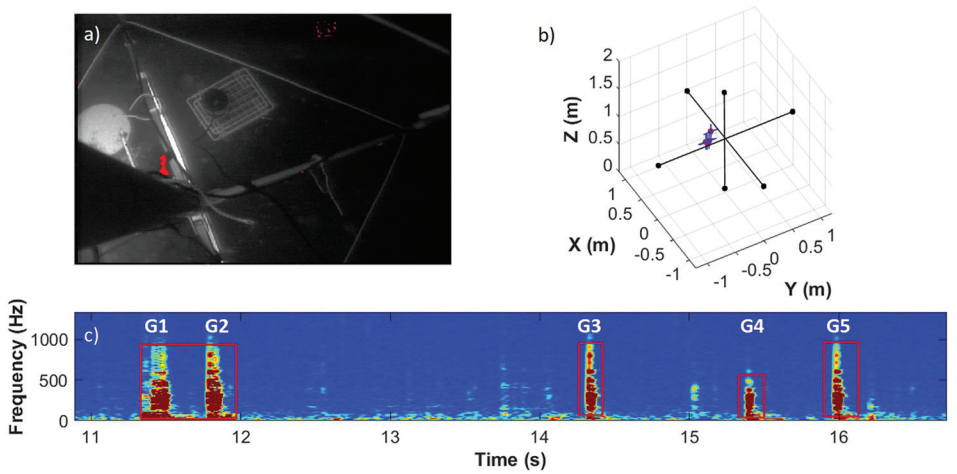

Fig. 3. Identification of sounds produced by a tautog. (a) Image from the video camera showing the tautog swimming in the middle of the array (red pixels). (b) Simultaneous acoustic localization (red dots) with uncertainties on each axis (blue lines). (c) Spectrogram of the sounds recorded on hydrophone 2. Red boxes indicate the sounds automatically detected by the detector that were used for the localization.

Mm. 1. 3D visualization of the localization uncertainties. This is a file of type "avi" (4071 KB).

Figure 3 shows the acoustic localization results when a tautog (*Tautoga onitis*) was swimming in the field of view of the camera. Identification of the species was performed visually from the top camera and from an additional non-recording side-view camera deployed on the side of the array. The location of the tautog from the video [highlighted with red pixels in Fig. 3(a)] coincides with the acoustic localization [Fig. 3(b)] of the five low-frequency grunts detected in the acoustic recording [labeled G1–G5 in Fig. 3(c)]. Grunts G1 and G2 were detected as one acoustic event by the automated detector and were consequently localized at the same time (i.e., one localization for both grunts). Small localization uncertainties [blue lines in Fig. 3(b)] leave

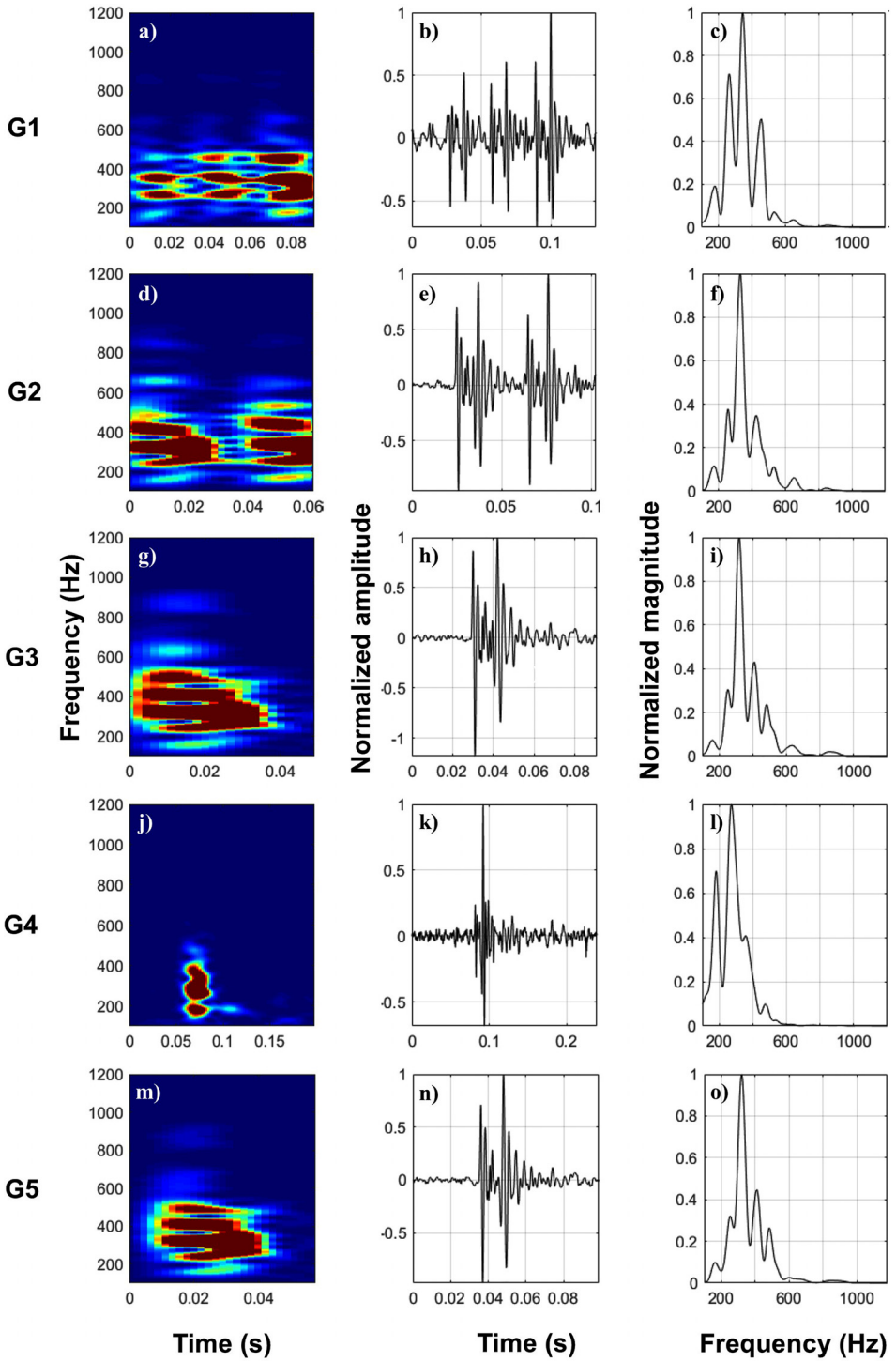

Fig. 4. (Color online) Spectrogram (left column), waveform (middle column), and spectrum (right column) of the five localized tautog grunts G1–G5 (each row corresponds to a tautog grunt).

no ambiguity that these grunts were produced by the tautog. A video of the simultaneous video recording, sound localization, and sound detections while the tautog was swimming inside the array is shown in Mm. 2. Note that the five other sounds that were automatically detected and localized could not be identified to specific fish species because they were outside of the field of view of the camera.

Mm. 2 Video showing simultaneously the video, localization results and sound detection. This is a file of type "avi" (1726 KB).

Figure 4 provides the spectrogram, waveform, and spectrum for each of the identified tautog grunts. All grunts are composed of one (G3–G5), two (G2), or three (G1) double-pulses. The component pulses of a double-pulse are separated by 11.25 ± 0.7 ms ($n = 8$). Grunts have a peak frequency of 317 ± 28 Hz ($n = 8$) and a duration from 22 ms (G3) to 81 ms (G1). Most of the energy for all tautog grunts was below 800 Hz. All time and frequency measurements were performed using the waveform (band-pass filtered between 100 and 1200 Hz with an eighth order zero-phase Butterworth filter, middle column in Fig. 4), and the average periodogram (spectral resolution of 3 Hz, 2048-sample Hanning window zero-padded to 16 384 samples for FFT, with a time step of 102 samples or 2.1 ms; right column in Fig. 4), respectively.

4. Discussion

Compact hydrophone arrays, like the one used in this study, in combination with underwater cameras provide the ability to catalog fish sounds non-intrusively in the wild. Their small footprint allows such systems to be portable and easily deployable. The system described here is cabled to the surface which does not allow deployment in remote areas for extended periods. An autonomous system that can record acoustic and video data for several weeks is currently being developed. In addition to cataloging fish sounds, and use in soniferous behavior research, such an array can be used to document source levels of fish sounds, which is critical information required for assessing the impact of anthropogenic noise on fish communication.

The tautog is an important fisheries species whose stock is overfished (ASMFC, 2017). Their sounds had previously only been reported by Fish and Mowbray (1970). Unfortunately, their description of the calls provides insufficient details to positively identify tautog sounds in acoustic recordings. While more measurements are needed to fully characterize the vocal repertoire of the tautog, this paper shows that the proposed combination of instruments and automated processing methods provides a systematic and efficient way to identify fish sounds from large datasets. The methodology described here promises to become a valuable tool to aid in developing fish and invertebrate sound libraries, as well as for *in situ* observations of soniferous behavior. This will help to continue the cataloging effort initiated by Fish and Mombray (1970) and make PAM a more viable tool for fish monitoring and fisheries management.

Acknowledgments

This research is supported by the NSERC Canadian Healthy Oceans Network and its Partners: Department of Fisheries and Oceans Canada and INREST (representing the Port of Sept-Îles and City of Sept-Îles), JASCO Applied Sciences, the Natural Sciences and Engineering Research Council (NSERC) Postgraduate Scholarships-Doctoral Program, and MITACS. The data collection was funded by the MIT Sea Grant College Program Grant No. 2010-R/RC-119 to R.R. and F.J.

References and links

Amorim, M. C. P., Vasconcelos, R. O., and Fonseca, P. J. (**2015**). *Fish Sounds and Mate Choice. Sound Communication in Fishes*, edited by F. Ladich (Springer-Verlag, Wien), pp. 1–33.

ASMFC (**2017**). "Amendment 1 to the Interstate Fishery Management Plan for Tautog," Atlantic States Marine Fisheries Commission report, http://www.asmfc.org/uploads/file/5a0477c3TautogAmendment1_Oct2017.pdf (Last viewed May 2, 2018).

Fish, M. P., and Mowbray, W. H. (**1970**). *Sounds of Western North Atlantic Fishes—A Reference File of Biological Underwater Sounds* (The John Hopkins Press, Baltimore).

Hawkins, A. D., and Amorim, M. C. P. (**2000**). "Spawning sounds of the male haddock, *Melanogrammus aeglefinus*," Environ. Biol. Fishes **59**, 29–41.

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

Locascio, J. V., and Mann D. A. (**2011**). "Localization and source level estimates of black drum (*Pogonias cromis*) calls," J. Acoust. Soc. Am. **130**, 1868–1879.

Moore, G. A. (**1968**). "Automatic scanning and computer processes for the quantitative analysis of micrographs and equivalent subjects," Pattern Recog.: Pict. Patt. Recog. **1**, 275–326.

Parsons, M. J., McCauley, R. D., Mackie, M. C., and Duncan A. J. (**2010**). "A comparison of techniques for ranging close-proximity mulloway (*Argyrosomus japonicus*) calls with a single hydrophone," Acoust. Austr. **38**, 145–151.

Parsons, M. J., McCauley, R. D., Mackie, M., Siwabessy, P. J., and Duncan A. J. (**2009**). "Localization of individual mullaway (*Argyrosomus japonicus*) within a spawning aggregation and their behaviour throughout a diel spawning period," ICES J. Mar. Sci. **66**, 1007–1014.

Riera, A., Rountree, R. A., Mouy, X., Ford, J. K., and Juanes, F. (**2016**). "Effects of anthropogenic noise on fishes at the SGaan Kinghlas-Bowie Seamount Marine Protected Area," Proc. Mtgs. Acoust. **27**, 010005.

Rountree, R. A. (**2008**). "Do you hear what I hear? Future technological development—and needs—in passive acoustics underwater observation," Marine Technol. Rep. **51**, 40–46.

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

Širović, A., and Demer, D. A. (**2009**). "Sounds of captive rockfishes," Copeia **2009**, 502–509.

Struzinski, A. W., and Lowe, E. D. (**1984**). "A performance comparison of four noise background normalization schemes proposed for signal detection systems," J. Acoust. Soc. Am. **76**, 1738–1742.

Zimmer, W. M. X. (**2011**). *Passive Acoustic Monitoring of Cetaceans* (Cambridge University Press, Cambridge, UK).