






MARCH 15 2023

SoundScape learning: An automatic method for separating fish chorus in marine soundscapes^{a)}

Ella B. Kim ; Kaitlin E. Frasier ; Megan F. McKenna ; Annebelle C. M. Kok; Lindsey E. Peavey Reeves; William K. Oestreich; Gabrielle Arrieta; Sean Wiggins ; Simone Baumann-Pickering 



J. Acoust. Soc. Am. 153, 1710–1722 (2023)

<https://doi.org/10.1121/10.0017432>



View
Online



Export
Citation

CrossMark








 **ASA**

Advance your science and career as a member of the
Acoustical Society of America

[LEARN MORE](#)

SoundScape learning: An automatic method for separating fish chorus in marine soundscapes^{a)}

Ella B. Kim,^{1,b)}  Kaitlin E. Frasier,¹  Megan F. McKenna,²  Annebelle C. M. Kok,¹ Lindsey E. Peavey Reeves,³ William K. Oestreich,⁴ Gabrielle Arrieta,¹ Sean Wiggins,¹  and Simone Baumann-Pickering¹ 

¹*Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92037, USA*

²*National Marine Sanctuary Foundation-Contracted, Silver Spring, Maryland 20910, USA*

³*National Marine Sanctuary Foundation, Silver Spring, Maryland 20910, USA*

⁴*Monterey Bay Aquarium Research Institute, Moss Landing, California 95039, USA*

ABSTRACT:

Marine soundscapes provide the opportunity to non-invasively learn about, monitor, and conserve ecosystems. Some fishes produce sound in chorus, often in association with mating, and there is much to learn about fish choruses and the species producing them. Manually analyzing years of acoustic data is increasingly unfeasible, and is especially challenging with fish chorus, as multiple fish choruses can co-occur in time and frequency and can overlap with vessel noise and other transient sounds. This study proposes an unsupervised automated method, called SoundScape Learning (SSL), to separate fish chorus from soundscape using an integrated technique that makes use of randomized robust principal component analysis (RRPCA), unsupervised clustering, and a neural network. SSL was applied to 14 recording locations off southern and central California and was able to detect a single fish chorus of interest in 5.3 yrs of acoustically diverse soundscapes. Through application of SSL, the chorus of interest was found to be nocturnal, increased in intensity at sunset and sunrise, and was seasonally present from late Spring to late Fall. Further application of SSL will improve understanding of fish behavior, essential habitat, species distribution, and potential human and climate change impacts, and thus allow for protection of vulnerable fish species.

© 2023 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1121/10.0017432>

(Received 5 October 2022; revised 19 January 2023; accepted 13 February 2023; published online 15 March 2023)

[Editor: Allen Mensinger]

Pages: 1710–1722

I. INTRODUCTION

Soundscapes provide a unique opportunity to non-invasively learn about, monitor, and conserve ecosystems. In the ocean, where space is vast and light reduces rapidly with depth, sound attenuates slowly, so many organisms primarily use sound to interact with their environment and with others (Kasumyan, 2008). Biotic, anthropogenic, and abiotic sounds all contribute to the marine soundscape, and our understanding of how organisms utilize marine soundscapes continues to expand (McKenna *et al.*, 2021). Passive acoustic monitoring (PAM) is a cost-effective tool to record and study soundscapes (Lindseth and Lobel, 2018).

Fish are important contributors to marine soundscapes, and sound production in fishes is likely far more widespread than is currently known. Globally, of the more than 34 000 fish species, at least 989 are currently known to produce sounds, usually while defending territory, feeding, and spawning, and likely many more are soniferous (Winn *et al.*, 1964; Bass and Ladich, 2008; Looby *et al.*, 2022). While there is a large and growing body of literature on fish sound

production, 96% of the 34 000 extant fish species lack published examinations of sound production (Looby *et al.*, 2023; Rice *et al.*, 2022). Sound production evolved approximately 33 times in Actinopterygii and is ancestral for radiations that comprise nearly 29 000 species, and thus sound production in actinopterygians is likely far more widespread than currently known (Rice *et al.*, 2022). Individual fish calls are typically low frequency (generally 40–1000 Hz), short duration, consist of broadband pulses or tones, often with multiple-frequency harmonics, are usually produced at night, dawn, and dusk, and show great diversity among species (Kasumyan, 2008). This diversity allows for the discrimination of sounds among fish species, and from sounds made by other marine organisms (Carrico *et al.*, 2019). While aggregating, some fish produce sounds together in a “chorus,” continuously increasing sound levels in specific frequency bands with few, if any, distinguishable individual calls (Greenfield and Shaw, 1983; Pagniello *et al.*, 2019). Fish chorusing can reach high enough sound levels to dominate the local soundscape (McKenna *et al.*, 2021), can last for multiple hours, and for some species (i.e., oyster toadfish, *Opsanus tau*) can even be heard from land (Kasumyan, 2008). Chorusing is not unique to fish, as birds, frogs, insects, and Baleen whales are also known to chorus to attract mates and to intimidate competitors (Lobel, 1992;

^{a)}This paper is part of a special on Fish Bioacoustics: Hearing and Sound Communication.

^{b)}Electronic mail: ebkim@ucsd.edu

Gerhardt, 1994; Au *et al.*, 2000; Dawson *et al.*, 2001; Thomas *et al.*, 2002; Catchpole and Slater, 2003; Bruni *et al.*, 2014; Party *et al.*, 2014; Greenfield *et al.*, 2017). Similarly to other chorusing organisms, fish chorus is thought to be related to reproduction, specifically mating (Brantley and Bass, 1994; Koenig *et al.*, 2017). Studying fish chorusing has allowed for the mapping of spawning areas and identification of spawning season, which has aided in effective management of fish species, and assessment of phenological shifts in chorusing due to climate change (Luczkovich *et al.*, 1997; Aalbers, 2008; Tellechea *et al.*, 2011; Rowell *et al.*, 2012; Zemeckis *et al.*, 2014; Borie *et al.*, 2021; Siddagangaiah *et al.*, 2022). Therefore, characterizing fish choruses within soundscapes is important as it can help identify mating seasons, essential habitats, species interactions, and distributions (Gannon *et al.*, 2008; Luczkovich *et al.*, 2008).

Historically, many analyses of fish sounds have utilized manual classification, which has become increasingly unfeasible. Fish choruses are difficult to identify manually due to their often-diffuse acoustic characteristics as well as co-occurrence with other signals. In our increasingly noisy global ocean (Hildebrand, 2009; Duarte *et al.*, 2021), since fish choruses are low frequency, they are often intertwined with vessel noise (Slabbekoom *et al.*, 2010; Popper and Hawkins, 2019). Also, the co-occurrence of multiple choruses in time and frequency makes manual analysis challenging. Additionally, manual classification of large acoustic datasets requires an expert human analyst with in-depth knowledge of the acoustic context of fish chorus at each recording location. Moreover, manually analyzing years of acoustic data for fish choruses is labor intensive, costly, and somewhat subjective. Thus, an unsupervised approach would greatly improve reliability, efficiency, and capacity of fish chorus analysis in large acoustically diverse datasets. While some automated methods exist (Sattar *et al.*, 2016; Lin *et al.*, 2017; Lin *et al.*, 2018; Lin, 2020; Butler *et al.*, 2021), there is a need for an unsupervised automated technique that works well at multiple sites and over long temporal scales, when multiple choruses and vessel noise are present, and when chorusing is faint.

Building on current methods, we have developed a new unsupervised automated method, called SoundScape Learning (SSL), to separate fish chorus from the soundscape through integration of randomized robust principal component analysis (RRPCA), unsupervised clustering, and a neural network. RRPCA uses randomized matrix decomposition to produce low rank (chronic) and sparse (transient) events. Data were essentially “denoised” of transient events using the RRPCA process: fish chorus and other chronic sources were aggregated into a low rank matrix, while less common signals like sonar, vessel cavitation, whales, and other biologic sounds were separated into a sparse matrix. Rather than relying on a human analyst to assign class-types manually, with potentially high error and inconsistency, an unsupervised clustering algorithm was used to identify multiple signal types (Frasier *et al.*, 2017). Deep machine learning algorithms have proven successful in classifying large passive acoustic datasets for marine mammals and fish (Bittle and

Duncan, 2013; Frasier *et al.*, 2017; Gibb *et al.*, 2019; Lin *et al.*, 2019). Thus, the clusters were used to train a neural net classifier (Frasier, 2021), to classify novel data from 14 unique sites representing a total of 5.3 yrs of acoustic data.

By applying SSL to a spatially and temporally extensive dataset, we were able to evaluate its ability to separate fish chorus from soundscape and uncover new biologically relevant insights. SSL was applied to a fish chorus present throughout southern and central California. The fish chorus has not been identified yet to species level; however, it is the same chorus reported by Pagniello *et al.* (2019). This chorus could be from pelagic and/or diel vertically migrating fish, given its temporal alignment with diel vertical migration (McCauley and Cato, 2016). SSL was also applied to a complex soundscape within Monterey Bay National Marine Sanctuary (MBNMS), which had multiple co-occurring choruses and vessel noise. Complex and simpler soundscapes both have a mixture of biotic, anthropogenic, and abiotic sounds, but here, complex soundscapes are specifically defined as those with multiple co-occurring sound sources overlapping in time and frequency, while simpler soundscapes also have multiple sound sources present, but they do not overlap in time and frequency. Through applying SSL across sites and varying soundscape conditions, we evaluate the utility of SSL under familiar and novel conditions. While this study focused on fish chorus, SSL is widely applicable to signal processing tasks which require separation and distinction of transient and chronic signals.

II. METHODS

A. Overall workflow

The SSL workflow broadly includes feature preparation, denoising, partitioning, and classification phases (Fig. 1). This workflow consists of five main steps: (1) calculating a matrix of sound levels from long-term spectral averages, (2) denoising and decomposition using RRPCA, (3) unsupervised clustering of denoised features to identify distinct classes, (4) deep network training, and (5) classification of novel data. Below, we provide a general description of each step and its utility, followed by specific parameterization details used in this application.

1. Data collection

Acoustic data were collected using high frequency acoustic recording packages (HARPs) at 14 sites (some with multiple deployments) throughout southern California and one SoundTrap in Monterey Bay National Marine Sanctuary (MBNMS) at various depths (Fig. 2). Sites were named after location, in which San Diego Trough was abbreviated to “SDT,” Southern California to “SOCAL,” and Monterey Bay to “MB,” and acronyms following the underscore allow for further differentiation between sites. HARPs and SoundTraps (SoundTrap ST500, Ocean Instruments, Auckland, NZ) are long-term, seafloor-mounted acoustic recorders that consist of a hydrophone, recording equipment, batteries, flotation, and release (Wiggins and Hildebrand, 2007; SoundTrap, 2022). HARPs are custom acoustic

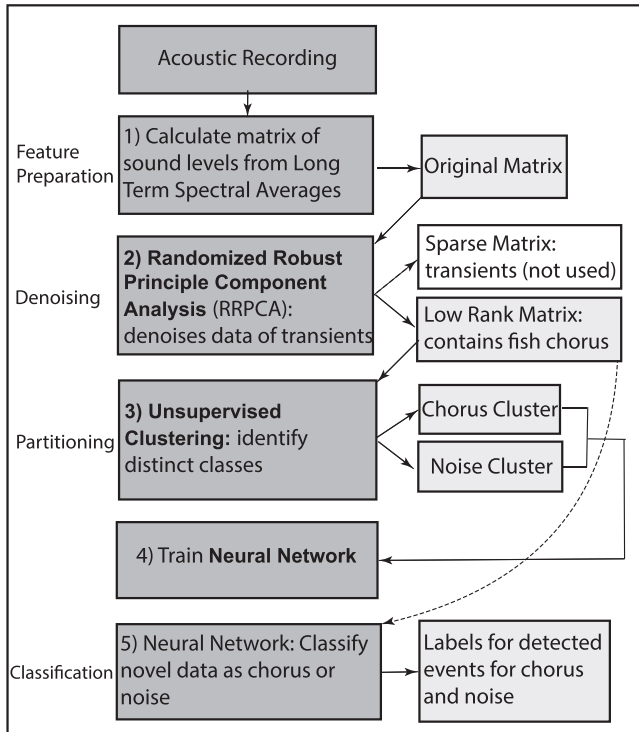


FIG. 1. Overview of the SSL workflow including data preparation (calculating matrix of sound levels from long-term spectral averages), RRPCA (denoising of transients), unsupervised clustering (identifies distinct classes), neural network training, and classification on novel data. Light gray boxes on the right side of the schematic represent output, and white box represents sparse matrix which is not used in later analysis.

recording devices designed and built at Scripps Institution of Oceanography and consist of a high frequency stage (ITC-1042, 2022) and a low frequency AQ1 (AQ1, 2022) bundle (Wiggins and Hildebrand, 2007). Instruments were moored 1–3.5 m from the ocean floor with a subsurface float and an acoustic release. The SoundTrap500 sampled continuously at 48 kHz and HARPs at 20 or 200 kHz, at various depths between ~60 to 1000 m between the years 2006 and 2019 (Table I). HARP hydrophones have an approximate sensitivity of -202.5 dB re $V/\mu\text{Pa}$, ~ 50 dB of gain, and 5 V of dynamic range. SoundTrap500s have an approximate full system sensitivity of -175 dB re $V/\mu\text{Pa}$, maximum clip level around 173 dB re $1 \mu\text{Pa}$, and 2 V of dynamic range. HARPs and SoundTraps were calibrated in the laboratory to provide frequency-dependent sensitivity (Wiggins and Morris, 2019). Representative data loggers and hydrophones were also calibrated at the U.S. Navy’s Transducer Evaluation Center facility to verify the laboratory calibrations.

2. Preparing data for analysis: Sound level matrices

Analyzing 5.3 yrs of raw XWAV (Triton Github, 2023) time series files was not practical, so data were compressed for overview in long-term spectral averages (LTSAs). Instead of using short duration spectrograms, successive spectra were calculated and averaged together, and then arranged sequentially to provide a time series of the spectra. LTSAs of each acoustic deployment were created, using

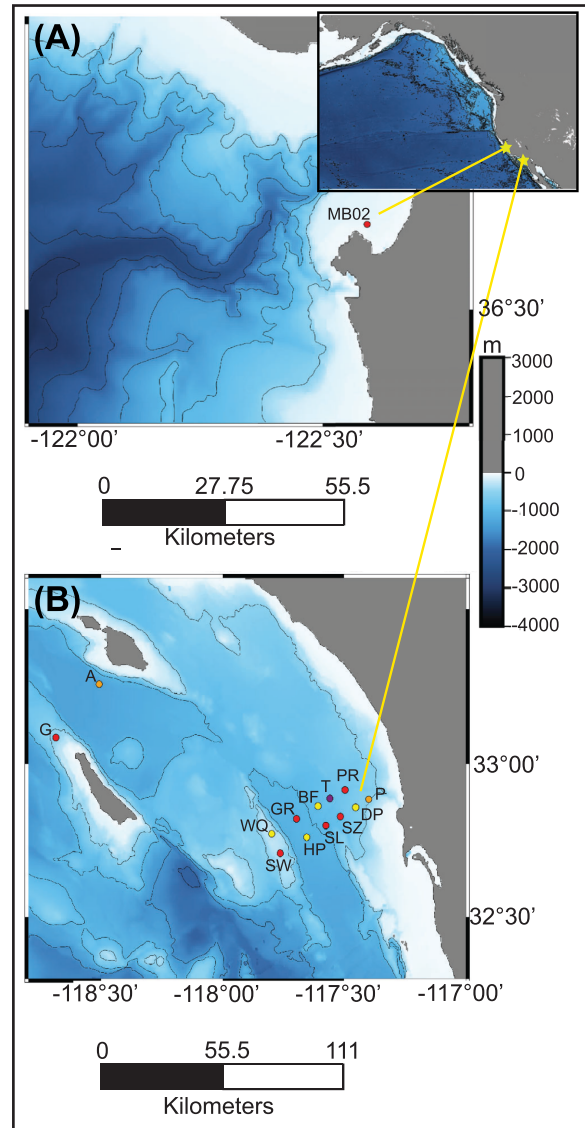


FIG. 2. (Color online) Site maps including (A) SoundTrap500 deployed in Monterey Bay National Marine Sanctuary, (B) HARPs deployed throughout Southern California. Yellow, neural net development sites; red, novel sites; orange, sites used for neural net development and novel classification; purple, long-term site. The range of distances between the 11 San Diego Trough southern-most sites and their nearest neighboring sites within the array was 3.16–14.24 km, and the SDT array itself was 98.7 km from the western most hydrophone (site G).

1000 point fast Fourier transforms (FFTs), Hanning windows, no overlap, and 1 s and 1 Hz resolution, using Triton’s “Soundscape LTSA package,” a custom MATLAB (MathWorks, Natick, MA) software (Triton Github, 2023; Wiggins and Hildebrand, 2007). Using the LTSAs, power spectral density (PSD) median values were computed for 20 Hz bins from 20–1000 Hz, for successive 20 min bins using the same Soundscape LTSA package. The output of the data preparation process was a sound level (PSD) matrix describing the soundscape for each deployment, considered the “original matrix.” Data were separated into neural net development sets, and novel datasets with appropriate spatial spread to adequately train and apply the network (detailed in Table I).

TABLE I. Site information for acoustic recordings. Note that some sites have multiple deployments which are reflected in “recording date.”

| Neural net development/novel sets | Site | Sampling rate (kHz) | Recording date | Depth (m) | Location |
|-----------------------------------|---------------------------|---------------------|-----------------------|-----------|-------------------------|
| Development | SDT_BF | 20 | 07/27/2017-11/13/2017 | 915 | 32°51.72 N, 117°34.51 W |
| | SDT_DP | 20 | 08/03/2017-11/15/2017 | 609 | 32°51.47 N, 117°27.20 W |
| | SDT_HP | 20 | 07/31/2017-11/13/2017 | 1051 | 32°45.64 N, 117°39.29 W |
| | SDT_WQ | 20 | 07/31/2017-11/13/2017 | 285 | 32°46.32 N, 117°46.91 W |
| | SOCAL_P | 200 | 09/23/2009-12/03/2009 | 479 | 32°53.60 N, 117°22.71 W |
| | SOCAL_A | 200 | 07/10/2007-10/21/2007 | 1112 | 33°15.03 N, 118°14.99 W |
| | SOCAL_T | 200 | 07/08/2017-01/17/2018 | 814 | 32°53.20 N, 117°33.50 W |
| Novel | SDT_PR | 20 | 08/03/2017-11/15/2017 | 725 | 32°54.87 N, 117°29.81 W |
| | SDT_SL | 20 | 07/27/2017-11/15/2017 | 943 | 32°47.93 N, 117°34.51 W |
| | SDT_SW | 20 | 07/31/2017-11/13/2017 | 285 | 32°42.51 N, 117°45.81 W |
| | SDT_SZ | 20 | 08/03/2017-11/15/2017 | 823 | 32°49.68 N, 117°30.95 W |
| | SDT_GR | 20 | 07/31/2017-11/13/2017 | 1068 | 32°49.23 N, 117°41.80 W |
| | SOCAL_A | 200 | 09/01/2006-11/07/2006 | 335 | 33°15.10 N, 118°15.14 W |
| | LJ_P | 200 | 05/30/2017-09/29/2017 | 517 | 32°53.05 N, 117°23.95 W |
| | LJ_P | 200 | 05/27/2013-09/05/2013 | 521 | 32°53.49 N, 117°24.82 W |
| | SOCAL_G | 200 | 07/23/2007-10/22/2007 | 480 | 32°55.61 N, 118°37.25 W |
| | SOCAL_T | 200 | 09/28/2016-12/13/2016 | 900 | 32°89.86 N, 117°60.98 W |
| | SOCAL_T | 200 | 03/03/2017-07/06/2017 | 825 | 32°53.21 N, 117°33.36 W |
| | MB02 (<i>SoundTrap</i>) | 48 | 04/08/2019-08/11/2019 | 68 | 36°38.97 N, 121°54.50 W |

3. RRPCA: Denoising the data of transients

RRPCA was utilized to decompose the original matrix into low rank and sparse matrices, using Rstudio’s ([Rstudio Team, 2022](#)) *rrpca* package ([Erichson et al., 2019](#)). RRPCA was applied to each original matrix between the frequency range for most fish choruses (and the target chorus) from 60–800 Hz to avoid low frequency noise and variation in sound level roll-off at ~850 Hz in some deployments due to data decimation. The sparse matrix was visually scanned to make sure it did not include the fish chorus of interest, and to generally understand types of transient signals included. Chronic fish chorus was separated into the low rank matrix while transient acoustic signals like vessel cavitation noise, whales, sonar, etc., were separated into the sparse matrix.

4. Unsupervised clustering: Creating distinct classes of chorus and noise

Utilizing the transient-denoised low rank matrix from development deployments, a MATLAB-based unsupervised clustering toolkit called “Cluster Tool” within Triton was utilized to identify distinct classes of chorus and noise ([Frasier, 2021](#)). Each development dataset was analyzed independently, and similar classes were manually pooled across datasets to form the neural network development set. A Euclidean distance score was computed between all possible pairs of the 20 min median PSD vectors in the development dataset utilizing the MATLAB function *pdist* [as computed by Eq. (1)]:

$$D(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}. \tag{1}$$

The distance between each pair of 20 min median PSD vectors was converted into a similarity metric *S*, such that

$$S = \exp(-D). \tag{2}$$

This resulted in a distance matrix, which can be interpreted as a network in which each PSD estimate is a “node.” and connections between nodes (edges) are assigned a length based on the nodes’ similarity. Similar nodes connected by short edges cluster together in this network while dissimilar nodes are pushed apart. After similarities were calculated between all nodes, edge pruning was utilized to reduce the size of the distance matrix input into the clustering algorithm, in which only the highest similarity scores were retained for clustering. We used the Chinese Whispers (CW) clustering algorithm ([Biemann, 2006](#)) to automatically identify groupings within the network. This algorithm starts by assuming that each node is its own cluster, and iteratively reassigns each node to the cluster to which it is most strongly connected until reassignments cease. This process partitioned the dataset into multiple categories which was used to train a neural network to recognize these categories in novel datasets.

Distance metrics were computed by comparing PSD vectors over a frequency range from 60–800 Hz in 20 Hz bins. During clustering, PSD values were normalized between values of 0–1 for each 20 min bin, where the lowest spectrum level bin was set to 0, and the highest spectrum level bin was set to 1. Cluster normalization resulted in larger and cleaner clusters. Edge pruning thresholds varied from 80%–90% as needed to isolate one or more clusters containing chorus. Clusters containing fewer than 30 nodes were discarded as they generally contained low quality,

highly variable events that were deemed unsuitable for classifier training. All chorus clusters from each of the seven training deployments were pooled into one chorus class and all “noise” clusters were pooled into a single noise class for later use in neural network training.

5. Comparing cluster quality of low rank vs original matrices

To evaluate whether RRPCA improved separation of chorus from transient signals, cluster quality of the original vs the low rank matrix was compared using two metrics: the Calinski–Harabasz index and silhouette scores. Calinski–Harabasz (CH) cluster evaluation (Calinski and Harabasz, 1974) measures the sum of inter- and intra-cluster dispersion for all clusters using the formula

$$CH = \frac{tr(Bk)}{tr(Wk)} \times \frac{nE - k}{k - 1}. \tag{3}$$

For a set of data, E , where nE is the number of data points, k is the number of clusters, $tr(Bk)$ is trace of the between group dispersion matrix and $tr(Wk)$, of the within-cluster dispersion matrix, defined by

$$W_k = \sum_{q=1}^k \sum_{x \in C_q} (x - c_q)(x - c_q)^T, \tag{4}$$

$$B_k = \sum_{q=1}^k n_q(c_q - c_E)(c_q - c_E)^T. \tag{5}$$

In which n_q is the number of points in cluster q , c_q is the center of cluster q , c_E is the center of E , and T is the number of iterations. Larger CH values indicate increased density within clusters and stronger separation between clusters. Additionally, the CH metric finds the ideal number of clusters. CH scores were calculated for the low rank and original matrix of SDT_BF using the *evalclusters* function in MATLAB.

Silhouette plots were used to visualize differences in original matrix and low rank matrix cluster quality. Silhouette scores range from -1 to 1 , where scores close to 1 are best, close to 0 indicate weak separation between clusters, and negative are likely misclassifications. The Silhouette score (SS) was calculated using the mean intra-cluster (i) and near-cluster distance (n) for each sample; the Silhouette score for a sample is defined by

$$SS = \frac{n - i}{\max(i, n)}. \tag{6}$$

Silhouette scores were calculated for original and low rank matrices using the *silhouette* function in MATLAB to create plots that visualize differences in cluster quality.

6. Neural network for classification of novel data

In the final step of this process, a neural network was trained to distinguish between noise and chorus classes as

aggregated by the unsupervised clustering process. The output from the unsupervised clustering algorithm was organized into training, testing, and validation sets using 60% of the development dataset for training, 30% for testing, and 10% for validation, with a maximum training set size of 1000 detections (Frasier, 2021). This 60/30/10 ratio is typical, and the maximum training set size was 1000 detections to utilize all chorus examples without excessive resampling. The total number of examples of each class in each subset were balanced to contain the same number of examples of chorus and noise, respectively, so that the neural network was not biased towards chorus or noise (He and Garcia, 2009). Additionally, 20 min of temporal separation were required between training and testing examples, so that the neural network was not testing on the same examples with which it had been trained (Jones, 2019).

Training the neural network: A binary classification network (yes or no to chorus presence) was trained using the classes identified with the unsupervised clustering process, utilizing a neural network toolbox that draws on MATLAB’s Deep Learning Toolbox (Frasier, 2021). The network consisted of a 512-node input layer and a 2-node softmax output layer, with four fully connected 128-node hidden layers in between, and 50% dropout between layers. Leaky rectified linear (ReLU) unit activations (Maas et al., 2013) were used and the network was trained over 15 epochs with a batch size of 50 events and constant learning rate of 0.0003. This design was utilized as it is straightforward to implement in most neural network frameworks.

Classifying novel data using trained neural network and assessing performance: The trained neural network was applied to low rank matrices computed from novel data. The neural network labels were manually reviewed as overlays on decimated LTSAs (lower resolution for faster manual screening) to assess label accuracy. Automated labels were manually reviewed for two novel deployments: SDT_PR, which had strong chorusing with few overlapping signals (in frequency) and MB02, which had great soundscape complexity with three overlapping choruses (overlapping in time and frequency) and ample noise. True and false positives and negatives were tabulated based on the manual corrections. Accuracy, recall, and precision were then calculated for both deployments using the equations:

$$Accuracy = \frac{true\ positives + true\ negatives}{all\ detections}, \tag{7}$$

$$Recall = \frac{true\ positives}{true\ positives + false\ negatives}, \tag{8}$$

$$Precision = \frac{true\ positives}{true\ positives + false\ positives}. \tag{9}$$

7. Timeseries analysis

To illustrate the potential of this method for long-term monitoring, the neural network was used to label chorus throughout over a year of data at site SOCAL_T.

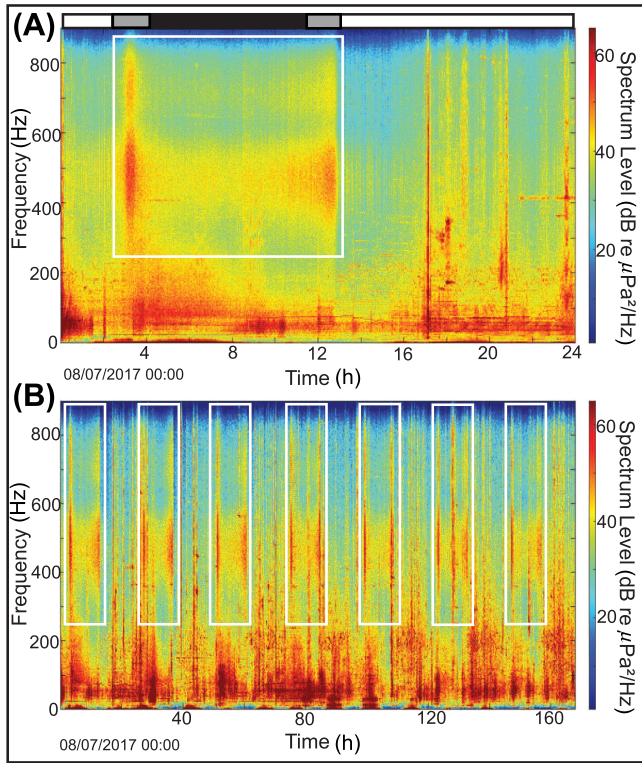


FIG. 3. (Color online) Long-term spectral average (LTSA) at San Diego Trough showing chorus within the larger soundscape over (A) 24 h, (B) 1 week. Chorus is at ~250–850 Hz with increased intensity at sunset and sunrise (white boxes around chorus). Colored bar at top of (A) night as black, astronomical twilight as gray, and day as white. Time in UTC.

Chorus presence was plotted in 20 min bins, and overlaid on local astronomical sunset and sunrise times with the MATLAB *sunrise* package (Beauducel, 2019).

III. RESULTS

A fish chorus of interest was initially identified in the San Diego Trough Soundscape (Fig. 3) during manual review of a subset of data. Manually identified chorus events occurred at night with increased intensity at sunset (~03:00 UTC) and sunrise (~13:00 UTC) (Fig. 3). Within the soundscape, instances of variable broadband noise were routinely present, mostly from close vessel encounters [Fig. 3(A)] (hours 16–24), primarily during daytime, and below 200 Hz (Fig. 3). Additionally, there was a low frequency chorus that often occurred just after sunset, between 20 and 200 Hz [Fig. 3(A)] (hours 3–7).

RRPCA separated the original spectral data at San Diego Trough into low rank and sparse matrices (Fig. 4). Chronic fish chorus was separated into the low rank matrix, and transient events were separated into the sparse matrix (Fig. 4). In the low rank and original spectra, the fish chorus appears as two peaks of increased amplitude between 300 and 800 Hz, with lower variance in the low rank matrix [standard deviation (sd) of spectra = 5.03] than the original matrix (sd = 5.33) (Fig. 4). Sparse matrix visualizations were confirmed to be transient events with the majority of energy below 200 Hz (Fig. 4). Thus, the RRPCA step

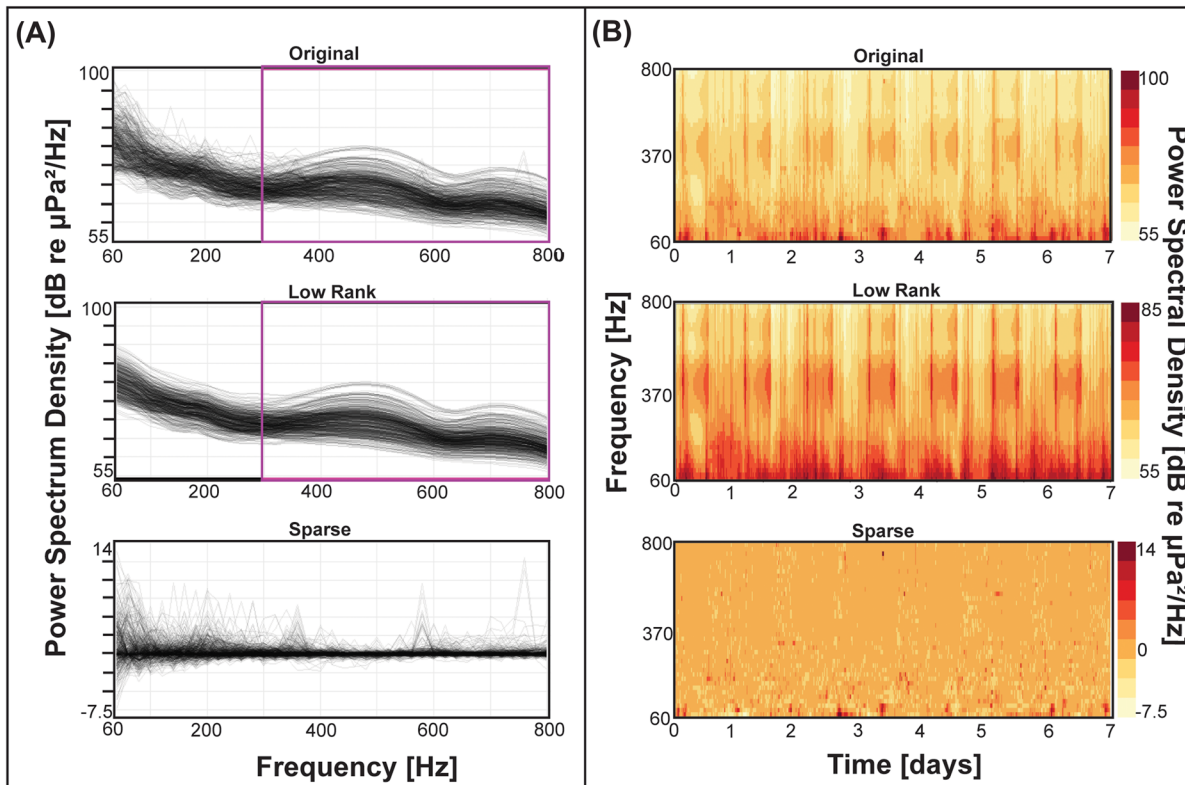


FIG. 4. (A) (Color online) Spectra for San Diego Trough in which each line represents 20 min binned PSD median values of original (sd = 5.33), low rank (sd = 5.03), and sparse matrices (sd = 0.95) for 1 week of the SDT_BF deployment (magenta box around chorus). (B) LTSA of 20 min binned PSD median values of original, low rank, and sparse matrices for the same week of the SDT_BF deployment (hotter color represents stronger PSD values).

denoised the data of transient events, and the low rank matrix was utilized for later analysis.

The unsupervised clustering algorithm identified distinct classes of chorus and noise. The total number of clustered chorus detections increased by a factor of two when using the low rank matrices as input rather than the original matrix [Fig. 5(A)]. Additionally, there were less clustered noise detections when using the low rank matrix as input rather than the original matrix [Fig. 5(B)]. Silhouette plots illustrated that low rank matrix chorus clusters included a larger number of chorus-positive bins than those of the original matrix, and Calinski–Harabasz index indicated that low rank matrix clusters resulted in denser and better separated clusters [Fig. 5(C)]. Note that for deployment SDT_HP, the noise cluster appeared to include some chorus, so the noise cluster was omitted from the training set. This may have been due to the low PSD levels of the chorus relative to other high PSD level noise at this site, and the lack of strong chorus examples to initiate cluster formation. For SOCAL_35_P, a

cluster of blue (*Balaenoptera musculus*) and fin whale (*Balaenoptera physalus*) calls (dominant energy <100 Hz) was formed when fish chorusing was absent. This cluster was omitted from the training set. For SOCAL_15_A, no chorus was detected, so the entire deployment contributed to the noise class.

The neural network classified chorus and noise on testing data with an overall 94.6% accuracy, in which signal intensity impacted classification accuracy (Fig. 6). The neural network assigned higher predicted probability values to chorus labels when the chorus magnitude was stronger [Fig. 6(A)]. Noise at and below 200 Hz was a notable deciding factor for classification, with predicted label probability decreasing as <200 Hz noise magnitude decreased [Fig. 6(A)]. Some low predicted probability classifications labeled as noise appear to be misclassified chorus [Fig. 6(A), right side of concatenated spectrum). There were more misclassifications of chorus (5%) than misclassifications of noise (0.4%), but those misclassifications generally were rare [Figs. 6(B) and 6(C)].

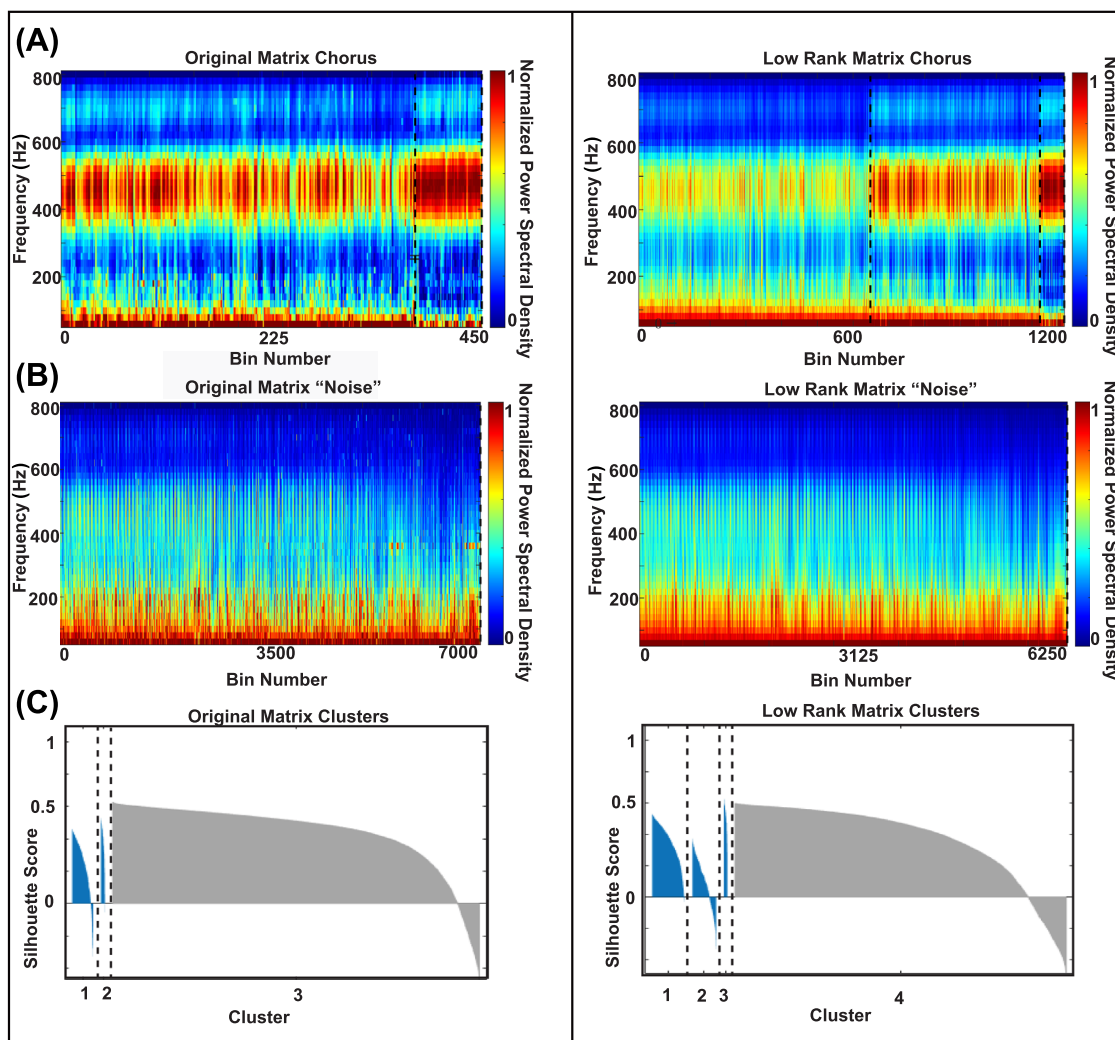


FIG. 5. (Color online) Spectra of concatenated clusters for original and low rank matrices showing number of 20 min binned detections vs frequency for (A) original and low rank matrix chorus clusters, (B) original and low rank matrix noise clusters, (C) silhouette plots for original and low rank matrix clusters. Blue, chorus clusters; gray, noise clusters. Scores close to 1: densest clusters with best separation, 0: overlapping clusters, -1: potential misclassifications. (CH index, original matrix: 1.27×10^3 , 3 clusters ideal vs low rank matrix 2.31×10^3 , 4 clusters ideal). Dashed vertical lines indicate separation between distinct clusters.

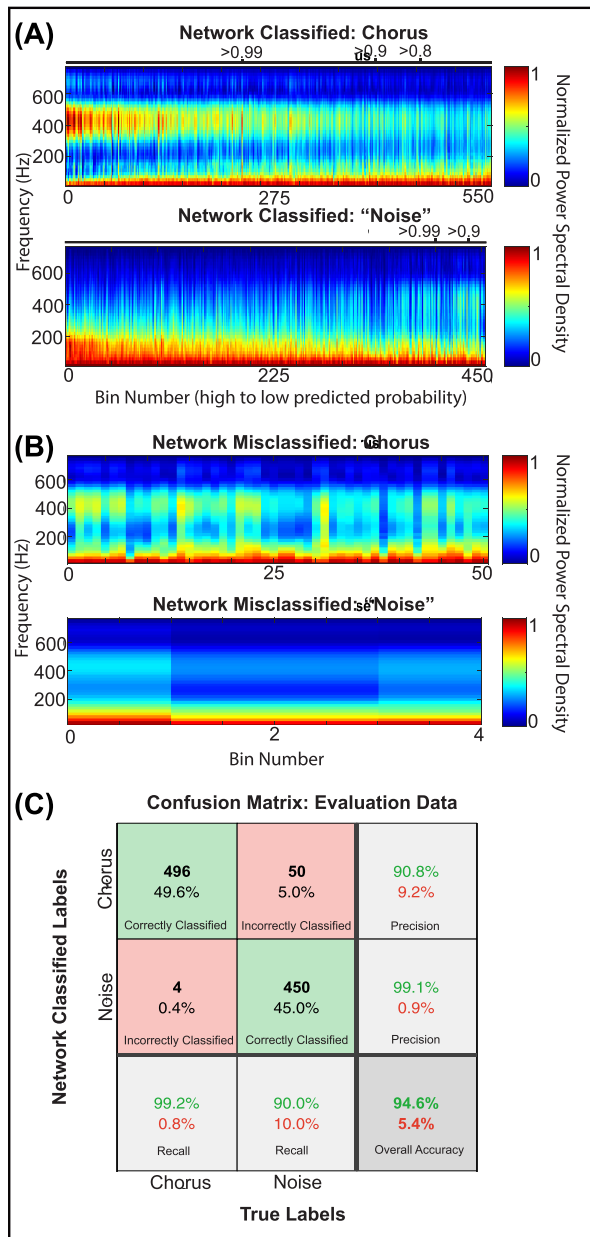


FIG. 6. (Color online) (A) Spectra of neural net classification output on test data, with predicted probability values on top of the figure. (B) Spectra of bins counted as network misclassifications of chorus and noise for test data. (C) Confusion matrix of test data in which the output class are network classified labels and the input class are true labels. Diagonal green cells represent observations that were correctly classified, and off diagonal red cells represent incorrectly classified observations. Both the number of observations and percentages of the total observations are shown in each cell. The far-right column shows precision, or percentages of all examples that the network classified to belong to each class that were correctly (top green percentage) and incorrectly (bottom red percentage) classified. The bottom row shows recall, or percentages of all examples belonging to each class that were correctly (top green percentage) and incorrectly (bottom red percentage) classified. The bottom right cell (dark gray) shows overall accuracy.

This tendency towards chorus “false negatives” rather than chorus “false positives” led to more conservative estimates of chorusing behavior, which was beneficial in this ecological application as it was not likely to include false detections of chorus even if a small number of true chorus detections

were lost. Also, many of the detections that the network thought were misclassifications of chorus, appear to actually be chorus detections which were erroneously included in the noise cluster test set [Fig. 6(B)], and were actually correctly labeled by the network. Essentially, the neural network classifier found inaccuracies in the ground truth. Overall, the neural network’s accuracy of 94.6% on the training set instilled confidence that the neural network was trained adequately and was performing well [Fig. 6(C)].

The neural network successfully classified chorus and noise in novel data (Fig. 7). For the simpler SDT_PR soundscape with just two choruses present (which did not overlap in frequency binning, but did co-occur in time binning), precision, recall, and accuracy metrics were higher [Fig. 7(A)], than the more complex MBNMS soundscape that had multiple choruses present (which co-occurred in time and frequency binnings) and was geographically separated from the deployment with which the neural network was trained [Fig. 7(B)]. The week-long SDT_PR LTSA showed that the neural network consistently labeled the sunset and sunrise choruses, at which time, the chorus magnitude was stronger, as well as daytime noise [Fig. 7(A)]. For the 48 h SDT_PR LTSA, within the nighttime chorus, noise was detected, which was likely from a lower frequency fish chorus of unknown species (~20 to 200 Hz) that started right after the sunset chorus and lasted for approximately 5 h [Fig. 7(A)]. The neural network skipped over broadband instances of noise at the ~17th hour, and between 38–43rd hours, illustrating the network’s ability to bypass broadband noise, and not mistake it for fish chorus [Fig. 7(A)]. For MBNMS, the network labeled sunset chorus and no sunrise chorus, which was consistent with manual review of chorusing behavior at this site [Fig. 7(B)]. A different nighttime chorus appearing as horizontal banding at 100, 200, 300, and 400 Hz, produced by the plainfin midshipman (*Porichthys notatus*) (McIver *et al.*, 2014), together with ample small vessel noise at this location masked potential occurrence of our target chorus at sunrise [Fig. 7(B)]. An additional nighttime chorus occurring at sunset at ~200 Hz produced by bocaccio (*Sebastes paucispinis*) (Sirovic and Demer, 2009) likely decreased labeling precision at sunset [Fig. 7(B)].

Time series analysis at site SOCAL_T elucidated diel and seasonal periodicity. The chorus began in May and ended in November (Fig. 8). Although we did not have two full years of coverage, the chorus likely ended around the same time, albeit slightly later in 2017 as opposed to 2016 (Fig. 8). Chorus presence was predominantly nocturnal, beginning at sunset and ending at sunrise (Fig. 8). In the beginning of the season, chorusing occurred at sunset and sunrise, then became more continuous through the night, and at the end of the season, waned to presence at just sunset and sunrise once again (Fig. 8).

IV. DISCUSSION

RRPCA worked well to “denoise” the matrix of transient events (Fig. 4) for more accurate classification. In the

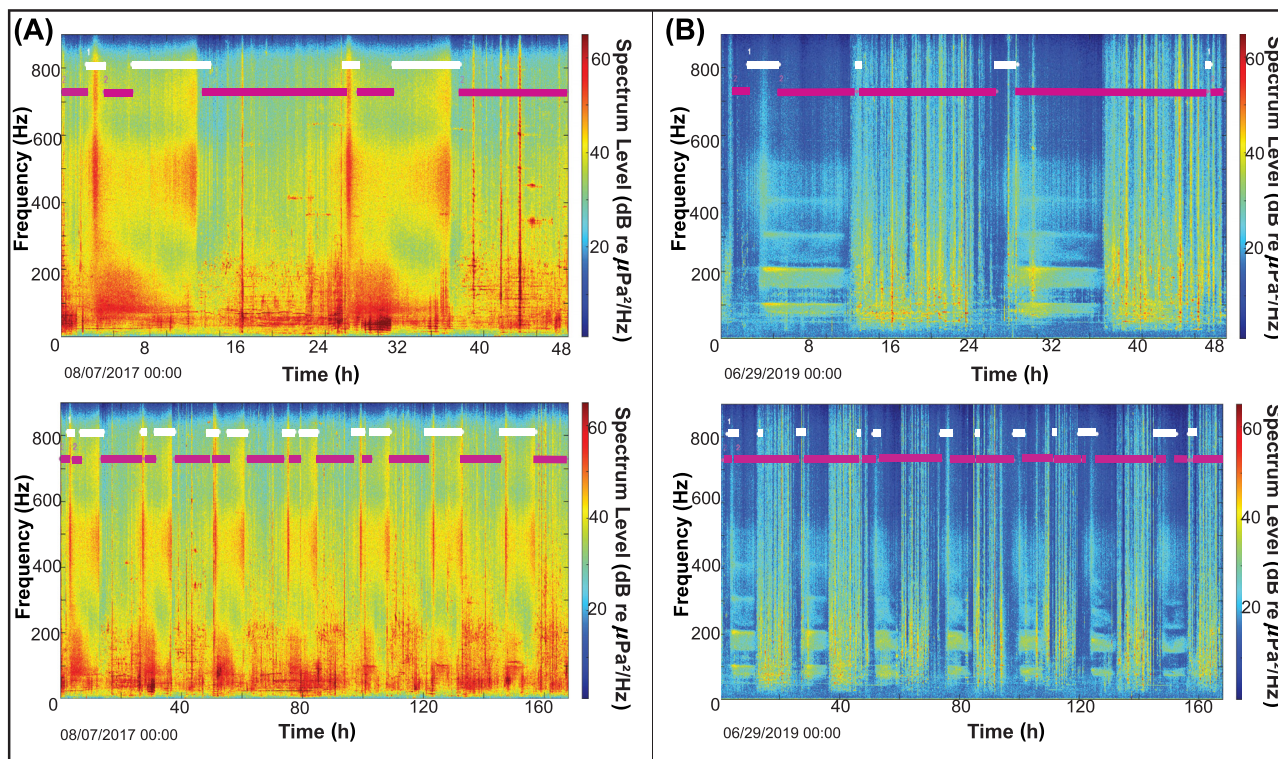


FIG. 7. (Color online) Original LTSAs of (A) San Diego Trough deployment SDT_PR (precision: 97.4%, recall: 99.3%, accuracy: 99.2%), (B) Monterey Bay deployment MB02 (precision: 69.9%, recall: 85.5%, accuracy: 96.5%) with neural network classification labels (20 min bins) for chorus (white), noise (magenta), visualized over (top) 48 h and (bottom) 1 week.

week-long LTSAs, there was no noticeable residual energy from the chorus left in the sparse matrix, which is beneficial for those who might want to use this method to quantify signal magnitude post-separation. In our application, PSD median values computed over 20 Hz and 20 min binning allowed for clean separation of fish chorus from transient events. However, the time and frequency binning (e.g., hourly third octave band levels), and the use of other averaging metrics (e.g., mean values), could be altered to target a signal of interest in other applications. Smaller standard deviation values in the low rank matrix spectra in comparison to the original matrix confirmed that the variance was reduced following the removal of transients by RRPCA. This result was advantageous as it mirrors the common practice of applying standard principal component analysis (PCA) prior to a clustering algorithm, as it is believed that denoising improves clustering results (Ding and He, 2004; Li *et al.*, 2021). Computationally, RRPCA is roughly five times faster than traditional RPCA (Erichson *et al.*, 2019), which was beneficial in this application with large acoustic datasets.

Implementation of RRPCA prior to clustering improved the size and quality of chorus clusters (Fig. 5). Overall, in the spectra of concatenated clusters [Figs. 5(A) and 5(B)] and the silhouette plots, which show number, size, density, and separation of clusters [Fig. 5(C)], the low rank matrix produced more detections of fish chorus, and higher-density clusters with improved separation [Fig. 5(C)], in comparison to the original matrix. Chorus detections were doubled in

the low rank matrix (vs original) due to RRPCA [Figs. 5(A) and 5(C)]. This was likely because the unsupervised clustering was able to detect fainter chorus [Fig. 5(A)] with less noise in the soundscape [Fig. 5(B)]. If the RRPCA step was skipped, and the original matrix was utilized instead, more instances of chorus would be pulled into the noise cluster, likely due to the inclusion of transient noise.

The neural network made classification decisions on the test set based on the intensity of chorus and noise, and overall showed strong accuracy (Fig. 6). The neural network found instances of chorus that were incorrectly clustered in the training set as noise (which were likely nodes at the edges of the clustering network) [Fig. 6(B)]. Thus, the few mistakes that were included in the training set did not disrupt the neural network’s classification, illustrating that the deep learning algorithm can recognize general patterns across many examples. The neural network also detected chorus in novel data across a diverse set of soundscapes of varying complexity. In this case, site SDT_PR was considered less complex because it had overlap of multiple choruses in time but not frequency. The MBNMS site was considered to be more complex, due to overlap of multiple choruses in time and frequency (Fig. 7). The network achieved high precision, recall, and accuracy values for the SDT_PR deployment, with less vessel noise and minimal other fish chorusing (which overlapped temporally but not in frequency). The neural network likely labeled instances of noise within the nighttime chorus due to decreased magnitude of the chorus of interest during the time of the <200 Hz

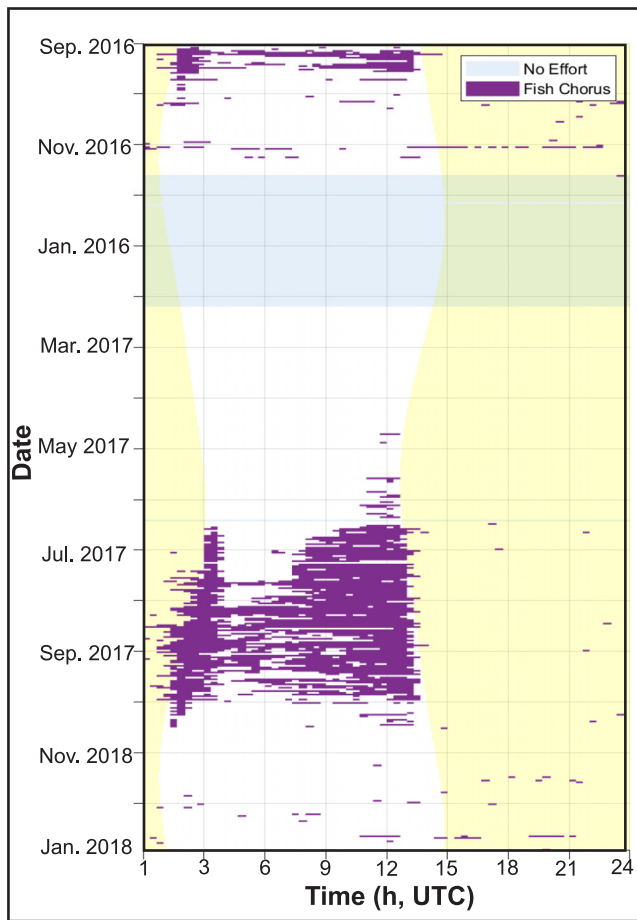


FIG. 8. (Color online) Diel presence of fish chorus (purple) as detected using SSL approach in UTC at site SOCAL_T in 20 min bins. Yellow shading represents daytime; blue shading represents “no effort,” when hydrophones were not deployed.

low frequency chorus (just after sunset) [Fig. 7(A)]. No unique cluster was formed during the training set development step for this chorus, likely because the clustering metric ignored frequency bins below 60 Hz to avoid low frequency noise (fragmenting this signal), or because of vessel noise dominance at the same frequency range. Nonetheless, it was beneficial that this low frequency chorus was labeled as noise as it was not the chorus of interest. In future studies, additional chorus classes could be added to the analysis process. The waxing and waning of various chorus intensities, and differing frequencies of these choruses were presumably due to acoustic niche partitioning in time and frequency (Krause, 1993). Acoustic niche partitioning is the result of various species in acoustic communities sharing limited soundscape bandwidth to limit competition and effectively communicate (Weiss *et al.*, 2021).

The neural network worked fairly well for the Monterey Bay National Marine Sanctuary (MBNMS) soundscape, which was more complex and considerably outside the range of southern California deployments with which the network was trained. The MBNMS deployment was considered more complex due to multiple overlapping fish choruses (in time and frequency) and higher occurrence and

amplitude of vessel noise [Fig. 7(B)]. Decreased labeling precision was likely the result of chorus misclassification as noise when bocaccio chorus (~200 Hz) occurred at the same time as the sunset chorus [Fig. 7(B)]. This was because the low frequency bocaccio chorus would have increased median PSD values at low frequencies, appearing far different from trained chorus examples in which intensity was strongest between 300 and 600 Hz. Bocaccio chorus was not present in the training set, and performance could likely be improved by adding Bocaccio examples during classifier training. Additionally, plainfin midshipman chorus and ample small vessel noise at this location masked potential occurrence of our target chorus at sunrise [Fig. 7(B)]. Additionally, differences between the HARPs and SoundTrap500 hydrophones (especially in gain) could have impacted neural network performance. For the MBNMS deployment, recall was notably higher than precision, as the neural network performed better at labeling all instances of true chorus (low false negative rate), but had a higher false positive rate (instances of noise labeled as chorus). Future work could consider inclusion of multiple chorus classes, and could explore the use of multilabel overlapping clustering analysis (Xia *et al.*, 2016; Peng and Liu, 2018) to increase the neural network’s precision and accuracy for soundscapes in which multiple choruses all occur simultaneously in time and frequency.

Through this automated method, we were able to gain insight on the temporal nature of this fish chorus at a long-term monitoring site SOCAL_T within the San Diego Trough. We found that the fish chorus occurred at night, with increased intensity at sunset and sunrise [Fig. 8]. Note that the few detections during the daytime were often the result of misclassifications of the neural network. The nocturnal nature of this chorus was consistent with other fish species (Helfman, 1986; Locascio and Mann 2011; McIver *et al.*, 2014; Staaterman *et al.*, 2014; Ruppé *et al.*, 2015), and the increased intensity at sunset and sunrise has been noted for Bocaccio rockfish (Sirovic and Demer, 2009) as well as for various bird species (Thomas *et al.*, 2002; Bruni *et al.*, 2014). There was no chorus detected from March–May, which is consistent with manual review of those time periods in the LTSA, and the neural network was confirmed to be working through labeling these times as “noise”, with no misclassifications. The chorus was present from May–November, which could indicate that the mating period of this fish species begins in late Spring and ends in late Fall (Fig. 8). The connection between fish calling and spawning has been noted in goliath grouper (*Epinephelus itajara*) and plainfin midshipman, in studies in which eggs were collected on nights of calling, and not collected on nights without calling (Brantley and Bass, 1994; Koenig *et al.*, 2017). The pattern of non-continuous nighttime chorus in the beginning of the season, with chorus at sunset and sunrise, more continuous chorusing mid-season, and then discontinuous chorusing at the end of the chorusing season, could indicate times of peak spawning during August–September. The chorusing season lines up with

known distribution, nighttime feeding, summer mating season, and reverse diel vertical migration habits of queenfish, *Seriphus politus*, making this species a possible candidate (D'Spain *et al.*, 2013). Future work might apply these methods across a wider range of recording locations to learn more about the spatial nature of this chorus (coastal and off-shore), and whether this chorus is indeed from queenfish, or from another croaker, and/or pelagic or diel vertically migrating fish.

While this study focused on fish chorus, this method is widely applicable to separate other signals when there is a chronic signal present, regardless of whether the chronic signal is or is not of interest. For instance, one could analyze the sparse matrix to learn about transient marine signals, like explosions, vessel noise, sonar, and other biologics. In one deployment, a small cluster of blue and fin whales was formed, and through simply altering time/frequency binning, one could better target separation of these whale calls or other biological calls of interest. SSL is a methodological advance that is a key step towards advancing marine soundscape analysis more broadly (as outlined by McKenna *et al.*, 2021), allowing for better autonomous monitoring of the health of the ecosystem and species. SSL could also be applied to terrestrial PAM sites. Applying SSL to bird, frog, and bat acoustics would likely be fruitful, especially for frogs, in which there is a current need for machine learning tools (Kitzes *et al.*, 2021; Larsen *et al.*, 2021). Outside of acoustics, any other ecological time series studies in which data can be represented as a large matrix (i.e., imagery, video) could apply these methods to easily separate signals of interest over time.

V. CONCLUSION

We successfully produced SSL, a novel unsupervised automated method to separate chronic fish chorus from other chronic (vessel noise) and transient acoustic signals. SSL was successfully applied across long temporal scales (5.3 yrs) and in diverse soundscapes (14 locations off California coast). In sum, RRPCA was utilized to separate the original matrix into low rank (chronic) and sparse (transient) matrices, and by extension, eliminate transient sounds. The low rank matrix was then clustered using an unsupervised clustering algorithm, which created unique chorus and noise classes. RRPCA was shown to significantly improve the size and quality of the clusters of interest. The clusters were then utilized to train a neural network for automatic classification on novel and diverse soundscapes. Through this application, we learned that the fish chorus was largely nocturnal in nature, with distinct seasonality. While this example was focused on separating fish chorus from soundscape, SSL is widely applicable to other large datasets across marine and terrestrial ecosystems, in which there is a need to automatically separate, detect, and classify signals. In the acoustic realm, manually analyzing data is becoming increasingly untenable with the collection of decades of data. It is our hope that this method will aid others to automatically

separate and detect signals with increased ease, with special appreciation for how much we can learn from marine soundscapes.

ACKNOWLEDGMENTS

Thank you to the science staff, vessel crews, and coordinators for their assistance with data collection and archiving. HARP data collection was made possible through Cooperative Ecosystems Study Unit Cooperative Agreement (Contract No. N62473-18-2-0016) with the U.S. Navy Pacific Fleet, with special thanks to Chip Johnson. SoundTrap data collection was made possible through NOAA's Sanctuary Soundscape project, which was a collaboration between NOAA and the U.S. Navy to better understand underwater sound within the National Marine Sanctuary System. A sub-award was issued to S.B.-P. at Scripps Institution of Oceanography (Grant Nos. N00244-19-2-0002 and N000244-20-2-0003) through the Naval Postgraduate School with special thanks to John Joseph. Much gratitude to the Dr. Nancy Foster Scholarship Program for funding doctoral studies of E.B.K. There are no conflicts of interest to declare. Acoustic data for MBNMS are available via the SanctSound Data Portal (SanctSound, <https://sanctsound.portal.axds.co/>). An overview guide to SSL, example data, and RRPCA code can be obtained through Dryad (<https://doi.org/10.5061/dryad.vq83bk3xs>). Code for calculating matrix of sound levels, clustering, and the neural network is available on GitHub (<https://github.com/MarineBioAcousticsRC/Triton>).

- Aalbers, S. A. (2008). "Seasonal, diel, and lunar spawning periodicities and associated sound production of white seabass (*Atractoscion nobilis*)," *Fish Bull.* **106**(2), 143–151.
- AQ1 (2022). Teledyne Marine Benthos, North Falmouth, MA, <http://www.teledynemarine.com/benthos/> (Last viewed February 27, 2023).
- Au, W., Mobley, J., Burgess, C., and Lammers, M. (2000). "Seasonal and diurnal trends of chorusing humpback whales wintering in waters off Western Maui," *Mar. Mammal Sci.* **16**(3), 530–544.
- Bass, A. H., and Ladich, F. (2008). "Vocal-acoustic communication: From neurons to brain," in *Fish Bioacoustics*, edited by J. F. Webb, R. R. Fay, and A. N. Popper (Springer Science and Business Media, LLC, New York), pp. 253–278.
- Beauducel, F. (2019). SUNRISE: compute sunset and sunrise time. Matlab/ GNU Octave function, <https://github.com/beaudu/sunrise>, BSD License.
- Biemann, C. (2006). "Chinese Whispers—An efficient graph clustering algorithm and its application to natural language processing problems," in *Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing*, pp. 73–80.
- Bittle, M., and Duncan, A. (2013). "A review of current marine mammal detection and classification algorithms for use in automated passive acoustic monitoring," in *Annual Conference of the Australian Acoustical Society 2013*, pp. 208–215.
- Borie, A., Rezende, S., Ferreira, B. P., Maida, M., Radford, C., and Travassos, P. (2021). "Soundscape of protected and unprotected tropical Atlantic coastal coral reefs," *Sci. Mar.* **85**(1), 5–14.
- Brantley, R. K., and Bass, A. H. (1994). "Alternative male spawning tactics and acoustic signals in the plainfin midshipman fish *Porichthys notatus girard* (Teleostei, Batrachoididae)," *Ethology* **96**(3), 213–232.
- Bruni, A., Mennill, D. J., and Foote, J. R. (2014). "Dawn chorus start time variation in a temperate bird community: Relationships with seasonality, weather, and ambient light," *J. Ornithol.* **155**(4), 877–890.

- Butler, J., Pagniello, C. M. L. S., Jaffe, J. S., Parnell, P. E., and Širović, A. (2021). "Diel and seasonal variability in kelp forest soundscapes off the Southern California Coast," *Front. Mar. Sci.* **8**, 629643.
- Calinski, T., and Harabasz, J. (1974). "A dendrite method for cluster analysis," *Commun. Stat.* **3**, 1–27.
- Carriço, R., Silva, M. A., Menezes, G. M., Fonseca, P. J., and Amorim, M. C. P. (2019). "Characterization of the acoustic community of vocal fishes in the Azores," *PeerJ*. **11**, 1–28.
- Catchpole, C., and Slater, P. J. B. (2003). *Bird Song: Biological Themes and Variations* (Cambridge University Press, Cambridge, UK).
- Dawson, A., King, V. M., Bentley, G. E., and Ball, G. F. (2001). "Photoperiodic control of seasonality in birds," *J. Biol. Rhythms* **16**(4), 365–380.
- Ding, C., and He, X. (2004). "K-means clustering via principal component analysis," in *Proceedings of the 21st International Conference on Machine Learning* (Banff, Canada).
- D'Spain, G., Batchelor, H., Helble, T. A., and McCarty, P. (2013). "New observations and modeling of an unusual spatiotemporal pattern of fish chorusing off the southern California coast," *Proc. Mtgs. Acoust.* **19**, 1–7.
- Duarte, C. M., Chapuis, L., Collin, S. P., Costa, D. P., Devassy, R. P., Eguiluz, V. M., Erbe, C., Gordon, T. A. C., Halpern, B. S., Harding, H. R., Havlik, M. N., Meekan, M., Merchant, N. D., Miksis-Olds, J. L., Parsons, M., Predragovic, M., Radford, A. N., Radford, C. A., Simpson, S. D., and Juanes, F. (2021). "The soundscape of the Anthropocene ocean," *Science* **371**(6529), eaba4658.
- Erichson, N. B., Voronin, S., Brunton, S. L., and Kutz, J. N. (2019). "Randomized matrix decompositions using R," *J. Stat. Softw.* **89**(11), 1–48.
- Frasier, K. E. (2021). "A machine learning pipeline for classification of cetacean echolocation clicks in large underwater acoustic datasets," *PLoS Comput. Biol.* **17**(12), 1–26.
- Frasier, K. E., Roch, M. A., Soldevilla, M. S., Wiggins, S. M., Garrison, L. P., and Hildebrand, J. A. (2017). "Automated classification of dolphin echolocation click types from the Gulf of Mexico," *PLoS Comput. Biol.* **13**(12), 1–23.
- Gannon, D. P. (2008). "Passive acoustic techniques in fisheries science: A review and prospectus," *Trans. Am. Fish. Soc.* **137**(2), 638–656.
- Gerhardt, H. C. (1994). "The evolution of vocalization in frogs and toads," *Annu. Rev. Ecol. Syst.* **25**, 293–324.
- Gibb, R., Browning, E., Glover-Kapfer, P., and Jones, K. E. (2019). "Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring," *Methods Ecol. Evol.* **10**(2), 169–185.
- Greenfield, M. D., Marin-Cudraz, T., and Party, V. (2017). "Evolution of synchronies in insect choruses," *Biol. J. Linn. Soc.* **122**(122), 487–504.
- Greenfield, M. D., and Shaw, K. C. (1983). "Adaptive significance of chorusing with special reference to the orthoptera," in *Orthopteran Mating Systems: Sexual Competition in Diverse Group of Insects* (Westview Press, Boulder, CO), pp. 1–27.
- He, H., and Garcia, E. A. (2009). "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.* **21**(9), 1263–1284.
- Helfman, G. (1986). *Fish Behaviour by Day, Night and Twilight* (Springer, Boston, MA), pp. 366–387.
- Hildebrand, J. A. (2009). "Anthropogenic and natural sources of ambient noise in the ocean," *Mar. Ecol. Prog. Ser.* **395**, 5–20.
- ITC-1042 (2022). Gavial ITC, Santa Barbara, CA, <https://www.gavial.com/itc-products> (Last viewed February 27, 2023).
- Jones, D. T. (2019). "Setting the standards for machine learning in biology," *Nat. Rev. Mol. Cell Biol.* **20**(11), 659–660.
- Kasumyan, A. O. (2008). "Sounds and sound production in fishes," *J. Ichthyol.* **48**(11), 981–1030.
- Kitzes, J., Blake, R., Bombaci, S., Chapman, M., Duran, S. M., Huang, T., Joseph, M. B., Lapp, S., Marconi, S., Oestreich, W. K., Rhinehart, T. A., Schweiger, A. K., Song, Y., Suringhe, T., Yang, D., and Yule, K. (2021). "Expanding NEON biodiversity surveys with new instrumentation and machine learning approaches," *Ecosphere* **12**(11), e03795.
- Koenig, C. C., Bueno, L. S., Coleman, F. C., Cusick, J. A., Ellis, R. D., Kingon, K., Locascio, J. V., Malinowski, C., Murie, D. J., and Stallings, C. D. (2017). "Diel, lunar, and seasonal spawning patterns of the Atlantic goliath grouper," *Bull. Mar. Sci.* **93**(2), 391–406.
- Krause, B. L. (1993). "The niche hypothesis: A virtual symphony of animal sounds, the origins of musical expression and the health of habitats," *Soundscape Newsl.* **6**, 6–10.
- Larsen, A. S., Schmidt, J. H., Stapleton, H., Kristenson, H., Betchkal, D., and McKenna, M. F. (2021). "Monitoring the phenology of the wood frog breeding season using bioacoustic methods," *Ecol. Indic.* **131**, 1–11.
- Li, P., Wang, W., Li, X., and ZHANG, C. (2021). "An image denoising algorithm based on adaptive clustering and singular value decomposition," *IET Image Process.* **15**, 598–614.
- Lin, T. H., Fang, S. H., and Tsao, Y. (2017). "Improving biodiversity assessment via unsupervised separation of biological sounds from long-duration recordings," *Sci. Rep.* **7**(1), 4547.
- Lin, T. H., Huang, J.-M., Yao, C.-J., Lien, Y.-S., Wang, P.-J., and Hu, Y. (2019). "Evaluating changes in the marine soundscape of an offshore wind farm via the machine learning-based source separation," in *2019 IEEE Underwater Technology (UT)*, Kaohsiung, Taiwan, pp. 1–6.
- Lin, T. H., and Tsao, Y. (2020). "Source separation in ecoacoustics: A roadmap towards versatile soundscape information retrieval," *Remote Sens. Ecol. Conserv.* **6**(3), 236–247.
- Lin, T. H., Tsao, Y., and Akamatsu, T. (2018). "Comparison of passive acoustic soniferous fish monitoring with supervised and unsupervised approaches," *J. Acoust. Soc. Am.* **143**(4), 278–284.
- Lindseth, A. V., and Lobel, P. S. (2018). "Underwater soundscape monitoring and fish bioacoustics: A review," *Fishes* **3**, 36.
- Lobel, P. S. (1992). "Sounds produced by spawning fishes," *Environ. Biol. Fishes* **33**(4), 351–358.
- Locascio, J. V., and Mann, D. A. (2011). "Diel and seasonal timing of sound production by black drum (*Pogonias cromis*)," *Fish Bull.* **109**(3), 327–338.
- Looby, A., Cox, K., Bravo, S., Rountree, R., Juanes, F., Reynolds, L. K., and Martin, C. W. (2022). "A quantitative inventory of global soniferous fish diversity," *Rev. Fish. Biol. Fish.* **32**(2):581–595.
- Looby, A., Vela, S., Cox, K., Riera, A., Bravo, S., Davies, H. L., Rountree, R., Reynolds, L. K., Martin, C. W., Matwin, S., and Juanes, F. (2023). "FishSounds Version 1.0: A website for the compilation of fish sound production information and recordings," *Ecol. Inf.* **74**, 101953.
- Luczkovich, J. J., Johnson, S. E., and Sprague, M. W. (1997). "Using sound to map fish spawning: Determining the seasonality and location of spawning for weakfish and red drum (Family Scaenidae) within Pamlico Sound, NC," *J. Acoust. Soc. Am.* **103**(5), 3000.
- Luczkovich, J. J., Mann, D. A., and Rountree, R. A. (2008). "Passive acoustics as a tool in fisheries science," *Trans. Am. Fish. Soc.* **137**, 533–541.
- Maas, A. L., Hannun, A. Y., and Ng, A. Y. (2013). "Rectifier nonlinearities improve neural network acoustic models," in *Proceedings of the 30th International Conference on Machine Learning*, Atlanta, GA, Vol. 28.
- McCauley, R. D., and Cato, D. H. (2016). "Evening choruses in the Perth Canyon and their potential link with Myctophidae fishes," *J. Acoust. Soc. Am.* **140**(4), 2384–2398.
- McIver, E. L., Marchaterre, M. A., Rice, A. N., and Bass, A. H. (2014). "Novel underwater soundscape: Acoustic repertoire of plainfin midshipman fish," *J. Exp. Biol.* **217**(13), 2377–2389.
- McKenna, M. F., Baumann-Pickering, S., Kok, A. C. M., Oestreich, W. K., Adams, J. D., Barkowski, J., Fristrup, K. M., Goldbogen, J. A., Joseph, J., Kim, E. B., Kügler, A., Lammers, M. O., Margolina, T., Peavey Reeves, L. E., Rowell, T. J., Stanley, J. A., Stimpert, A. K., Zang, E. J., Southall, B. L., Wall, C., Van Parijs, S., and Hatch, L. T. (2021). "Advancing the interpretation of shallow water marine soundscapes," *Front. Mar. Sci.* **8**, 719258.
- Pagniello, C. M. L. S., Cimino, M. A., and Terrill, E. (2019). "Mapping fish chorus distributions in southern California using an autonomous wave glider," *Front. Mar. Sci.* **6**, 526.
- Party, V., Brunel-Pons, O., and Greenfield, M. D. (2014). "Priority of precedence: Receiver psychology, female preference for leading calls and sexual selection in insect choruses," *Anim. Behav.* **87**, 175–185.
- Peng, L., and Liu, Y. (2018). "Feature selection and overlapping clustering-based multilabel classification model," *Math. Probl. Eng.* **2814897**.
- Popper, A. N., and Hawkins, A. D. (2019). "An overview of fish bioacoustics and the impacts of anthropogenic sounds on fishes," *J. Fish Biol.* **94**(5), 692–713.
- Rice, A. N., Farina, S. C., Makowski, A. J., Kaatz, I. M., Lobel, P. S., Bemis, W. E., and Bass, A. H. (2022). "Evolutionary patterns in sound production across fishes," *Ichthyol. Herpetol.* **110**(1), 1–12.
- Rowell, T. J., Schäfer, 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. Ecology Prog. Ser.* **462**, 241–250.

- RStudio Team (2022). "Integrated Development Environment for R. RStudio," <http://www.rstudio.com/> (Last viewed September 2022).
- Ruppé, L., Clément, G., Herrel, A., Ballesta, L., Décamps, T., Kéver, L., and Parmentier, E. (2015). "Environmental constraints drive the partitioning of the soundscape in fishes," *Proc. Natl. Acad. Sci. U.S.A.* **112**(19), 6092–6097.
- Sattar, F., Cullis-Suzuki, S., and Jin, F. (2016). "Acoustic analysis of big ocean data to monitor fish sounds," *Ecol. Inf.* **34**, 102–107.
- Siddagangaiyah, S., Chen, C. F., Hu, W. C., and Farina, A. (2022). "The dynamical complexity of seasonal soundscapes is governed by fish chorusing," *Commun. Earth Environ.* **3**(1), 109.
- Sirovic, A., and Demer, D. A. (2009). "Sounds of captive rockfishes," *Copeia* **3**, 502–509.
- Slabbekoorn, H., Bouton, N., van Opzeeland, I., Coers, A., ten Cate, C., and Popper, A. N. (2010). "A noisy spring: The impact of globally rising underwater sound levels on fish," *Trends Ecol. Evol.* **25**(7), 419–427.
- SoundTrap (2022). <https://www.oceaninstruments.co.nz/> (Last viewed January 17, 2023).
- Staaterman, E., Paris, C. B., DeFerrari, H. A., Mann, D. A., Rice, A. N., and D'Alessandro, E. K. (2014). "Celestial patterns in marine soundscapes," *Mar. Ecol. Prog. Ser.* **508**, 17–32.
- Tellechea, J. S., Bouvier, D., and Norbis, W. (2011). "Spawning sounds in whitemouth croaker (Sciaenidae): Seasonal and daily cycles," *Bioacoustics* **20**, 159–168.
- Thomas, R. J., Székely, T., Cuthill, I. C., Harper, D. G. C., Newson, S. E., Frayling, T. D., and Wallis, P. D. (2002). "Eye size in birds and the timing of song at dawn," *Proc. Biol. Sci.* **269**(1493), 831–837.
- Weiss, S. G., Cholewiak, D., Frasier, K. E., Trickey, J. S., Baumann-Pickering, S., Hildebrand, J. A., and van Parijs, S. M. (2021). "Monitoring the acoustic ecology of the shelf break of Georges Bank, Northwestern Atlantic Ocean: New approaches to visualizing complex acoustic data," *Mar. Policy* **130**, 104570.
- Wiggins, S. M., and Hildebrand, J. A. (2007). "High-frequency acoustic recording package (HARP) for broad-band, long-term marine mammal monitoring," in *2007 Symposium on Underwater Technology and Workshop on Scientific Use of Submarine Cables and Related Technologies*, Tokyo, Japan (April 17–20, 2007), pp. 551–557.
- Wiggins, S. M., and Morris, M. A. (2019). "SoundTrap ST500 calibration at the Transducer Evaluation Test Center (TRANSDEC), Marine Physical Laboratory Technical Memorandum 645" (Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA).
- Winn, H. E., Marshall, J. A., and Hazlett, B. (1964). "Behaviour, diel activities, and stimuli that elicit sound production and reactions to sounds in the longspine squirrelfish," *Copeia* **2**, 413–425.
- Xia, Y., Nie, L., Zhang, L., Yang, Y., Hong, R., and Li, X. (2016). "Weakly supervised multilabel clustering and its applications in computer vision," *IEEE Trans. Cybern.* **46**(7), 3220–3232.
- Zemeckis, D. R., Hoffman, W. S., Dean, M. J., Armstrong, M. P., and Cadrin, S. X. (2014). "Spawning site fidelity by Atlantic cod (*Gadus morhua*) in the Gulf of Maine: Implications for population structure and rebuilding," *ICES J. Mar. Sci.* **71**(6), 1356–1365.