



Machine learning analysis reveals relationship between pomacentrid calls and environmental cues

Jill E. Munger^{1,8,*}, Daniel P. Herrera², Samara M. Haver^{1,3}, Lynn Waterhouse⁴, Megan F. McKenna^{5,9}, Robert P. Dziak⁶, Jason Gedamke⁷, Scott A. Heppell³, Joseph H. Haxel^{1,10}

¹Cooperative Institute for Marine Ecosystem and Resources Studies, NOAA Pacific Marine Environmental Laboratory and Oregon State University, Hatfield Marine Science Center, Newport, OR 97365, USA

²Portland, OR 97218, USA

³Department of Fisheries, Wildlife, and Conservation Sciences, Oregon State University, Corvallis, OR 97333, USA

⁴Daniel P. Haerther Center for Conservation and Research, John G. Shedd Aquarium, Chicago, IL 60605, USA

⁵Natural Sounds and Night Skies Division, National Park Service, Fort Collins, CO 80525, USA

⁶NOAA, Pacific Marine Environmental Laboratory, Newport, OR 97365, USA

⁷Office of Science and Technology, NOAA Fisheries, Silver Spring, MD 20910, USA

⁸Present address: Department of Fisheries, Wildlife, and Conservation Sciences, Oregon State University, Corvallis, OR 97333, USA

⁹Present address: Hopkins Marine Station, Stanford University, Pacific Grove, CA 93950, USA

¹⁰Present address: Pacific Northwest National Laboratory, Sequim, WA 98382, USA

ABSTRACT: Sound production rates of fishes can be used as an indicator for coral reef health, providing an opportunity to utilize long-term acoustic recordings to assess environmental change. As acoustic datasets become more common, computational techniques need to be developed to facilitate analysis of the massive data files produced by long-term monitoring. Machine learning techniques demonstrate an advantage in the identification of fish sounds over manual sampling approaches. Here we evaluated the ability of convolutional neural networks to identify and monitor call patterns for pomacentrids (damselfishes) in a tropical reef region of the western Pacific. A stationary hydrophone was deployed for 39 mo (2014–2018) in the National Park of American Samoa to continuously record the local marine acoustic environment. A neural network was trained—achieving 94% identification accuracy of pomacentrids—to demonstrate the applicability of machine learning in fish acoustics and ecology. The distribution of sound production was found to vary on diel and interannual timescales. Additionally, the distribution of sound production was correlated with wind speed, water temperature, tidal amplitude, and sound pressure level. This research has broad implications for state-of-the-art acoustic analysis and promises to be an efficient, scalable asset for ecological research, environmental monitoring, and conservation planning.

KEY WORDS: Convolutional neural network · Passive acoustic monitoring · Fish acoustics · Damselfishes · American Samoa

1. INTRODUCTION

Marine environments are notoriously difficult to study over extended time frames, but long-term monitoring is an effective means to evaluate changes in ecosystem status. Passive acoustic monitoring

(PAM) allows for cost-effective, protracted environmental sampling that can be used to monitor ecosystem health and species diversity. PAM complements other techniques by sampling marine environments at night, in low-visibility conditions, and over longer temporal scales. Specifically, PAM provides some

*Corresponding author: jillemunger@gmail.com

advantages over visual assessments and can provide information to inform more holistic conservation efforts and marine park management.

Within the marine environment, the teleosts (bony fishes) are excellent candidates for using PAM to understand behavior and evaluate the effects of environmental change. Teleosts have the most diverse mechanisms for sound generation among all vertebrates (Ladich et al. 2006), and the abundance of sound production presents an opportunity to examine how sound production behaviors respond to external pressures (Mann & Lobel 1998). An important step in PAM analysis for teleost sound production is the development of automated analytical techniques capable of efficiently processing the vast quantity of acoustic data produced during long-term PAM deployments. Starting with a well-studied, acoustically active group of fishes, like pomacentrids, is the best way to facilitate the development of these techniques.

Pomacentrids play an important ecological role in coral reef systems, and healthy populations are often associated with healthy reefs that have high coral cover, habitat complexity, and functional diversity (Noonan et al. 2012, Darling et al. 2017). This group of fishes is important in reef trophic dynamics as prey for meso- and top predators (Emslie et al. 2019). The pomacentrids are also one of the most thoroughly studied groups of acoustically active fishes (Cole 2010). By grinding their pharyngeal teeth, pomacentrids create pops, clicks, and chirps that are generally associated with aggressive behavior and nest defense, or a series of pulses, called a pulse train, that is associated with the courtship display of males (Weimann et al. 2018). Sound generation is associated with these different behaviors and can vary between species. There can even be distinctive variation in sound production at the individual level (Vieira et al. 2015). The combination of an abundance of previous acoustic studies, their ecological importance, and the diversity of sounds produced make pomacentrids ideal candidates for the development of automated PAM analytical techniques.

Machine learning techniques can be used to automate processing of the extensive amount of data collected by PAM recorders, which can continuously sample for multiple years (Allen et al. 2021). Machine learning techniques have been used to detect acoustic signals from birds and bats, with more recent applications to marine mammals (Bergler et al. 2019, Berman et al. 2019, Shiu et al. 2020, Zhong et al. 2020). Some initial research with machine learning and fish sound production has also been done (Vieira et al. 2015, Noda et al. 2016, Sattar et al. 2016, Lin et al. 2018, Malfante et al. 2018).

A convolutional neural network (CNN) is a class of deep neural networks commonly applied to image recognition tasks. CNNs offer a useful methodology for datasets by leveraging state-of-the-art image recognition architectures and training techniques such as transfer learning. Transfer learning uses a given dataset to fine-tune the pretrained weights of an existing model, as opposed to training a model from scratch. Transfer learning accelerates the model training process and improves performance, but more importantly, it allows researchers to train from a much smaller dataset. In the speech recognition realm, CNNs have outperformed more traditional machine learning systems that use Gaussian mixture and hidden Markov models (Hinton et al. 2012). Combining passive acoustic data collection with CNN techniques offers minimally invasive, inexpensive solutions to help understand how the marine environment is changing over time due to climate change and other anthropogenic influences.

In 2015, the National Oceanic and Atmospheric Administration (NOAA) and the National Park Service (NPS) collaborated to establish the Ocean Noise Reference Station (NRS) Network. The NRS is a US-wide long-term PAM array specifically designed to monitor baseline conditions and long-term trends in underwater soundscapes in marine environments, including anthropogenic and biological drivers of underwater sound (Haver et al. 2018). This multi-year effort to collect underwater passive acoustic data includes monitoring in the National Park of American Samoa (NPSA), which presents a unique opportunity to develop bioacoustic detection and analytical techniques used to survey ecologically important indicator species (such as pomacentrids) and explore the relationships between acoustic conditions and animal behavior in a remote ocean area. Haver et al. (2019) examined the underwater acoustic environment at NPSA, focusing on ambient sound levels and humpback whale calls to document acoustic conditions, and compared those features to other marine protected area monitoring sites in waters less than 100 m depth. The findings suggest that acoustic metrics shift with time, reflecting environmental factors. Fish bioacoustics in American Samoa remain unexplored, but these data can be leveraged with minimal additional effort to characterize the relationship between the sound production behaviors of fish and environmental variates, some of which, like sea surface temperature (SST), are closely tied to climate change (Doney et al. 2012).

The goal of our work was to (1) determine the viability of training a CNN to detect pomacentrid sounds,

(2) utilize the CNN to determine the temporal distribution of sound production behaviors in NPSA over a 4 yr period, and (3) explore potential connections between pomacentrid sounds and environmental cues such as wind speed, water temperature, tidal amplitude, and sound pressure level (SPL). Our research demonstrates the applicability of CNN machine learning techniques to PAM data to detect and classify pomacentrid sounds. Determining a baseline archive of pomacentrid sound production contributes to the understanding of changes in occurrence related to shifts in the marine environment and the possible effects of changing acoustic environments on fishes.

2. MATERIALS AND METHODS

2.1. Location and site description

American Samoa is a small archipelago in the South Pacific Ocean region of Polynesia. It is geographically isolated and not in proximity to any major ports. NPSA is a 54 km² national park in the US territory of American Samoa established in 1988 and distributed across 3 islands: Tutuila, Ofu, and Ta'ū. The park preserves and protects coral reefs, tropical rainforests, fruit bats, and the Samoan culture. The

Coral Reef Ecosystem Division (CRED) of NOAA's Pacific Islands Fisheries Science Center (PIFSC) has previously monitored fish abundance at the study site via towed-diver and rapid ecological assessment techniques. These studies were conducted biannually from 2002 to 2010 (PIFSC 2011). CRED was also involved in the PAM of humpback whales in American Samoa (Munger et al. 2012). The NRS autonomous underwater hydrophone (AUH) recorder was deployed within a marine protected area managed by NPSA to monitor the acoustic environment and assess the impacts of noise on marine mammals, fish, and other marine resources in the park. The hydrophone recorded more than 18 000 h of acoustic data (616 GB) over 39 mo.

2.2. Instrumentation

The AUH was designed and built by the NOAA Pacific Marine Environmental Laboratory, and deployed by NPSA-based NPS dive staff in sandy substrate at a depth of 33 m off the northern coast of Tutuila Island (-14.27° , -170.72°), approximately 60–100 m from the nearest coral reef (Fig. 1). The AUH was recovered annually to download data and refurbish equipment. The in-water deployment dates were from 14 June 2015 to 5 April 2016; 18 August to

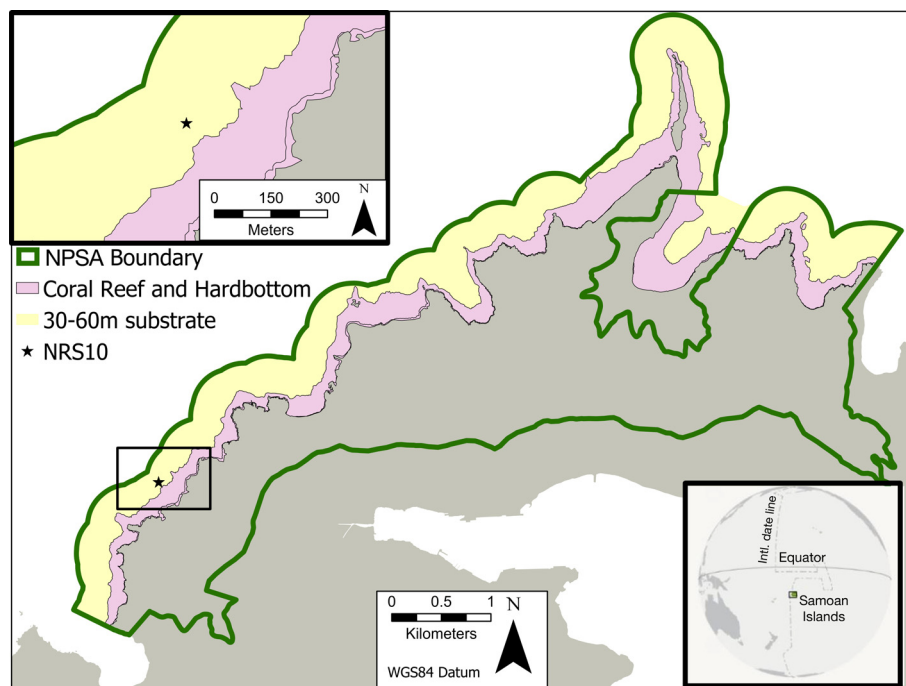


Fig. 1. Location of the autonomous underwater hydrophone (NRS10) in the National Park of American Samoa (-14.27° , -170.72°) in relation to nearby coral reef and hardbottom substrate

12 November 2016; and 28 July 2017 to 1 September 2018. The AUH contained a single hydrophone with a sensitivity of -192 dB re 1 V/m Pa and it was programmed to continuously sample at a rate of 5 kHz. Raw binary data were conditioned by a calibrated pre-amplifier and pre-whitening filter before conversion to .WAV format for analysis (Haver et al. 2019).

2.3. Analysis of patterns in fish sound production

2.3.1. Temporal variation and patterns in sound production

We wanted to examine the variation in pomacentrid sound production across years. Because the numbers of pomacentrid calls were zero-inflated and non-normal (Fig. S1 in the Supplement at www.int-res.com/articles/suppl/681p197_supp.pdf), a Kruskal-Wallis non-parametric test was performed to determine if there was a statistically significant difference in the number of subsamples that contained pomacentrid calls across all 4 study years. Additionally, to evaluate the 6 pairwise differences between the years, a pairwise Wilcoxon rank sum test was performed with a Bonferroni correction to account for multiple comparisons. All computations were run in R (version 4.0.4; R Core Team 2021).

2.3.2. Relationship of sound production to environmental parameters

We investigated the relationship between wind speed, water temperature, tidal amplitude, and SPL and the number of subsamples where pomacentrid sounds were present using a zero-inflated negative binomial (ZINB) regression analysis with the 'pscl' package in R (Zeileis et al. 2008). The ZINB model accounts for an excess of zero values and non-Gaussian data (Blasco-Moreno et al. 2019). The results for ZINB models are reported in 2 parts. The zero-inflated portion indicates if the covariates are significant predictors of false zeros. A false zero is defined as the absence of calls due to experimental design or observer error. The count portion of the model indicates if the covariates are significant predictors of the number of subsamples containing pomacentrid sounds. Each part of the ZINB model is reported as an odds ratio, a measure of association between an exposure (to the environmental variables) and an outcome (a pomacentrid call). It is not possible to calculate probability in a case-control study. Both the zero-inflated and

count portions of the model included all environmental variables (tidal amplitude, wind speed, SST, and SPL) with year, month, and hour as factors (Table S1).

SST and wind speed data were obtained from NOAA National Ocean Service Buoy Station NS TP6 - 1770000 - Pago Pago, American Samoa (-14.28° , -170.69° ; NOAA National Ocean Service 1971). Tidal amplitude was calculated in MATLAB[®] (version 2018b; MATLAB 2018) software using Oregon State University's Tidal Model Driver model. Ambient SPLs were calculated in 5 min segments with MATLAB[®] software from the original 10 Hz to 2 kHz data files (.DAT binary format) using long-term spectral averages summarized in 1 Hz/5 min bins. We chose ambient SPL measured as the root mean square (RMS) of the time series within the 50 Hz to 1.5 kHz frequency band ($SPL_{RMS\ 50\ Hz - 1.5\ kHz}$) because it contains the dominant frequency range of pomacentrid sound production (Frédérich & Parmentier 2016).

2.3.3. Estimating detection range

Sound propagation is influenced by a suite of environmental factors, so we used conservative propagation estimates in the absence of specifics about how particular types of substrate and environmental conditions would impact the sound propagation. To estimate the detection range of pomacentrid calls for the acoustic sensor, we assumed that the study environment (i.e. water temperature, salinity, and benthic substrate) was consistent at the stationary AUH. We also assumed simple spherical spreading of underwater sound (i.e. sound propagates uniformly in all directions) with transmission (energy) loss, relating only to the distance from the sources [$10 \times \log(\text{range}^2)$]. Transmission loss of pomacentrid sounds between the reef and the AUH (approximately 100 m minimum at 33 m depth) was calculated in the MATLAB[®] Phased Array System Toolbox assuming a fundamental frequency of 400 Hz (Frédérich & Parmentier 2016) and sound production source levels of green damselfish *Abudefduf adominalis* at 105–130 dB_{RMS} re 1 μ Pa at 1 m (Maruska et al. 2007). In general, source levels of reef fish are reported to be 144–157 dB_{RMS} re 1 μ Pa at 1 m (Tolimieri et al. 2004), indicating that the source levels of *A. abdominalis* are lower and represent a conservative estimate. To approximate the range of distances that the fish calls could be detected at the acoustic sensor, we subtracted the transmission loss and daily mean ambient noise (dB_{RMS} re 1 μ Pa 355 Hz [center frequency 400 Hz]; Haver et al. 2019) from the source levels of the fish calls.

2.4. CNN

2.4.1. Background and general approach

The artificial intelligence (AI) strategy is 1 of 6 key areas of the NOAA 2020 Science & Technology Strategy where NOAA plans to expand efforts in transformative advancement (NOAA Research Council no date). NOAA has already been working collaboratively with Google on innovative applications of machine learning such as the acoustic monitoring of humpback whales (Allen 2018, Harvey 2018, Allen et al. 2021). However, automatic detection of the sound production of fishes is a relatively new and under-explored area of science, and one that supports the ambition of the NOAA AI Strategy (NOAA 2020).

Our approach translated the sound detection and classification problem into a supervised image recognition task using spectrograms that allowed us to leverage advances in CNN architectures. We utilized the ResNet-50 neural network (Russakovsky et al. 2015, He et al. 2016) and transfer learning to produce a state-of-the-art binary classification model (Fig. 2).

2.4.2. Training data collection

The training dataset was comprised of known pomacentrid pulse trains and chirps because they are

good indicators of pomacentrid acoustic activity. Pomacentrid sounds were identified at the family level by comparing acoustic attributes such as frequency bandwidth, spectral peak frequency, pulse repetition frequency, pulse number, inter-pulse interval, and call duration (Maruska et al. 2007, Parmentier et al. 2009, Weimann et al. 2018). Field recordings in .WAV format were analyzed visually and aurally using the sound analysis software Raven Pro (The Cornell Lab of Ornithology 2014; Fig. 3).

Training data samples were manually selected using a systematic sampling strategy in Raven Pro by scrolling spectrograms from 1 day from each of the first 10 mo of data, every 2 h in 2 min segments. These samples were used to build the initial pomacentrid sound production training dataset ($n = 223$). An initial dataset with approximately the same number of samples was also collected to approximate noise, including non-pomacentrid fish calls, whales, snapping shrimp, and ambient sound that occurs when no pomacentrid sounds are present (Fig. 2, Step 1).

The timestamps from the Raven Pro tables were used to create 2 s long samples from the original acoustic signal. These samples were then converted into Mel spectrograms (Fig. 2, Step 2) using the 'fastai_audio' library (Blum & Bracco 2019) with the following parameters: $f_{\min} = 200.0$, $f_{\max} = 1000.0$, $\text{hop_length} = 32$, $n_{\text{fft}} = 128$, $n_{\text{mels}} = 64$, $\text{pad} = 0$,

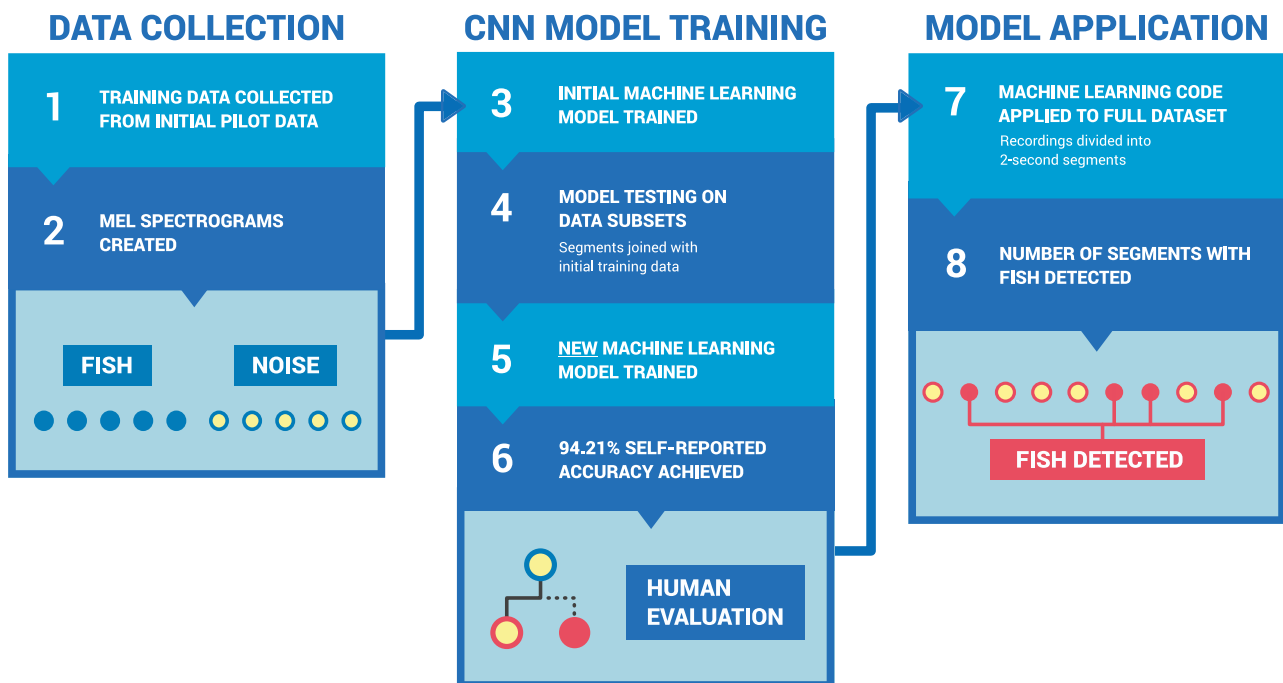


Fig. 2. Machine learning process for detection of pomacentrid sound production from initial collection of data and model training to application over the entire dataset. Details are provided in Section 2. CNN: convolutionary neural network

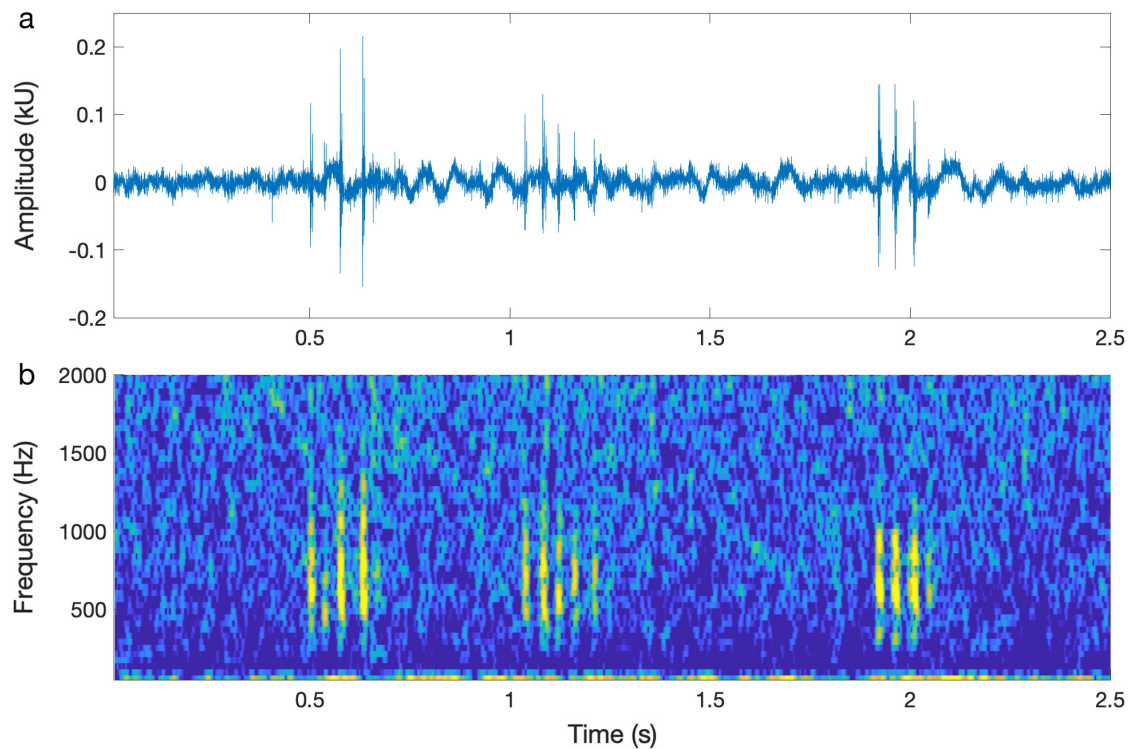


Fig. 3. Representative (a) waveform and (b) spectrogram of pomacentrid sound production (Hanning window Fast Fourier transform [FFT] = 128, overlap = 90 %, Nonequispaced Fast Fourier Transform [NFFT] = 128, sampling frequency = 5000). Color gradient (dark blue to yellow) indicates intensity of the sound in each frequency and time bin. The lowest intensity sound levels are dark blue, while the highest intensity are orange to bright yellow). Sounds typically exhibit short duration (0.1 s), 400–1200 Hz clusters of pulses, with inter-pulse intervals of ~0.1–0.25 s

win_length = None. A Mel spectrogram uses the Mel scale—a logarithmic conversion from hertz based on pitch comparisons that reduces the dynamic range of the audio, making the resulting spectrogram more amenable to generalized image recognition. Mel spectrograms led to higher accuracy rates over linear spectrograms when fine-tuning the model, presumably because they made the frequencies containing pomacentrid calls more visually prominent.

2.4.3. Training process

Using the pre-trained weights from the ResNet-50 architecture, we fine-tuned an image classifier with our initial training dataset using the 'fastai_audio'

library (Fig. 2, Step 3). The initial model was deployed to infer candidate call segments within 4 h recording segments that were not previously used to populate the initial training set. In machine learning, inference refers to the process of making predictions based on your trained model. Candidate segments containing pomacentrid sounds were evaluated by a human and joined with the initial training dataset (Fig. 2, Step 4). The training dataset was approximately doubled using this technique, improving the accuracy of the model. Total samples collected for training included 465 damselfish sounds and 472 examples of noise (Supplement, [https://github.com/whoahbot/sonumator/tree/master/training_set] for training and test datasets). The 'fastai' library uses 20 % of the training dataset for validation (Fig. 4).

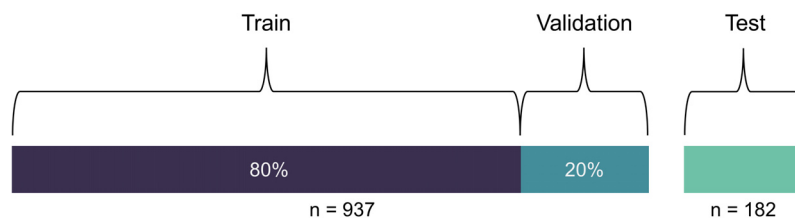


Fig. 4. Split of the data used to train the model (Train), evaluate the model during training (Validation), and evaluate the model with previously unseen data (Test)

We began the training process again with the initial Resnet-50 weights and our expanded training corpus. The 'fastai' weight decay default was increased to 0.1 to avoid overfitting of the model. A learning rate of 0.02 and 20 epochs yielded the highest accuracy (Fig. 2, Step 5). The newly trained model was tested with novel recording segments, and classifications were both visually and aurally inspected for accuracy (Fig. 2, Step 6).

2.4.4. Model application and inference

All recordings were binned into 2 s segments and the model was used in a binary fashion to detect the presence of sound production in each segment (Fig. 2, Steps 7 & 8). One-Zero scoring methods were used to determine the number of segments where pomacentrid sounds were detected by the model. The area under the curve (AUC) was calculated at the default 'fastai_audio' library decision threshold of 0.5 using the 'scikit-learn' module (Pedregosa et al. 2011) in Python.

3. RESULTS

3.1. Model performance

The accuracy of the trained CNN model based on the validation dataset was 0.9421, suggesting that 94% of all 2 s segments were correctly classified as containing either a pomacentrid sound or noise. Cross-validation was measured using a previously unseen test dataset, separate from training and validation data, of $n = 182$ recording segments (Fig. 4). Accuracy of the model based on the test dataset was 0.8736. Sensitivity or recall, measured as the ability of the model to correctly identify sound production, was 1.00. Precision, the metric reporting the proportion of correct positive identifications, was 0.7553. The specificity (probability of no detection given that there was no sound production) of the model was 0.7928, and with the decision threshold of the fastai library set at 0.5, we calculated the AUC at 0.9 (Fig. S2). Further explanation of the use of these metrics can be found in Metz (1978).

3.2. Detection range estimate

We calculated the approximate received sound level of pomacentrid calls to be in excess of daily

mean ambient levels at ranges up to 200 m, which is greater than the estimated distance (60–100 m, Fig. 1) between the AUH and nearest reef system.

3.3. Temporal distribution and patterns

At least 1 year was identified as having statistically different call rates from the other years (2015–2018; $H(4) = 4386$, 3 df, $p < 0.0001$). Upon further analysis, each year was found to be significantly different from all other years ($p < 0.0001$) with a Bonferroni corrected significance level of $\alpha = 0.0083$. The percent of 2 s segments where sound production was detected declined from 3.3% in 2015 to 1.1% in 2017 (Table 1). The interannual variation results suggest heterogeneous occurrence of sound production behaviors between study years and a non-monotonic decline in the presence of sounds over the first 3 yr of the study (Table 1, Fig. 5). There was a slight increase in the median number of segments where sound production was detected from 2017 to 2018.

We also found notable variability in pomacentrid sound production rate. For example, there was an increase in the 1 wk rolling average number of segments in February 2018 and another peak in April 2018 (Fig. 6). Additionally, pomacentrids are more sonically active during the day, following a diel cycle of increased call activity between 06:00 and 18:00 h local time (Fig. 7).

3.4. Environmental parameters

3.4.1. Zero-inflated portion of the model

Tidal amplitude and $SPL_{RMS\ 50\ Hz - 1.5\ kHz}$ were both significant predictors in the logit model predicting

Table 1. Summary statistics for the number of 2 s segments where sound production was detected in each 15 min interval

| | 2015 | 2016 | 2017 | 2018 |
|--|-------|-------|-------|-------|
| Number of 2 s segments with sounds detected | 19242 | 17280 | 14976 | 23401 |
| Percent of segments where sounds were detected | 3.3 | 2.7 | 1.1 | 1.8 |
| Median number of segments where sounds were detected | 15 | 12 | 5 | 8 |
| Standard deviation from median | 31.50 | 33.37 | 19.15 | 29.07 |

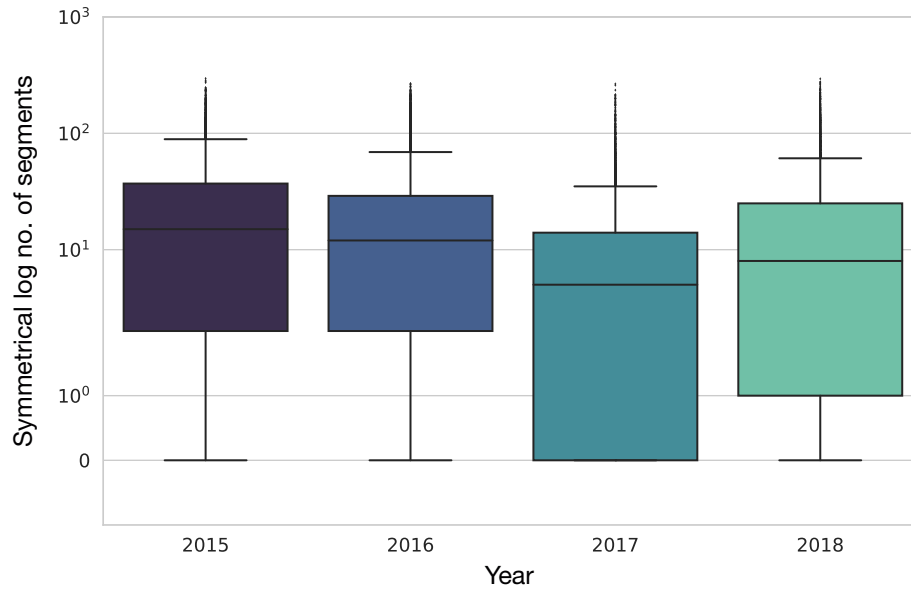


Fig. 5. Symmetrical log number of 2 s segments where sound production was detected in each 15 min interval by year for pomacentrids in the National Park of American Samoa. Median values, represented by central tendency lines, are provided in Table 1. Interquartile ranges are represented by the box limits. Error bars extend to the smallest and largest observations that are not outliers. Data (represented by points) outside of the error bars are considered outliers

false zeros (resulting from observer or experimental design error). The odds of a sound not being present in each 15 min recording interval decreased by 0.80 for each unit increase in tidal amplitude and increased by 1.10 for each unit increase (1 dB) in $SPL_{RMS\ 50\ Hz - 1.5\ kHz}$, holding other variables constant. Wind speed and water temperature were not significant predictors of the absence of sound production (Table 2). Occurrence of increased tidal amplitude was associated with fewer false zeros, while increased $SPL_{RMS\ 50\ Hz - 1.5\ kHz}$ was associated with more false zeros.

3.4.2. Count portion of the model

The odds of a pomacentrid sound being present in each 15 min recording interval increased by 1.05 for each unit increase in tidal amplitude (m), by 1.04 for each unit increase in wind speed ($m\ s^{-1}$), and by 1.08 for each unit increase (1 dB) in $SPL_{RMS\ 50\ Hz - 1.5\ kHz}$, holding other variables constant. The odds of pomacentrid sound being present decreased by 0.77 for each unit increase in SST ($^{\circ}C$), holding other variables constant (Table 2). Increased tidal amplitude, wind speed, and

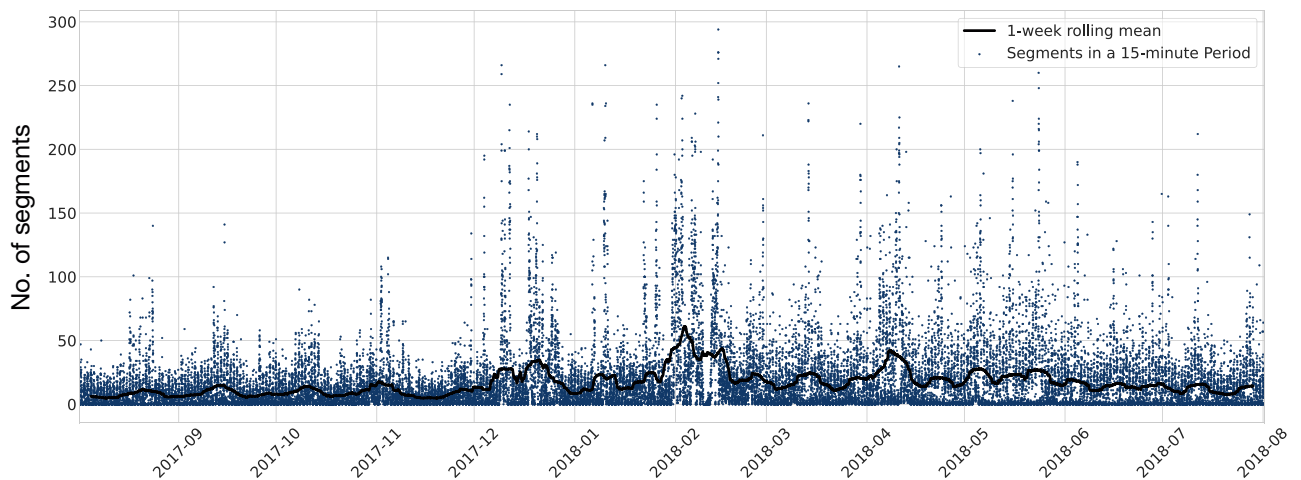


Fig. 6. Variability in the number of pomacentrid sounds produced over time at the National Park of American Samoa. Data are displayed as the number of 2 s segments where sound production was detected for each 15 min interval and a 1 wk rolling mean between August 2017 and July 2018

SPL_{RMS 50 Hz - 1.5 kHz} was associated with more pomacentrid sounds being present. Occurrence of increased SST was associated with fewer sounds being present.

4. DISCUSSION

4.1. CNN for pomacentrid sounds

Deep learning techniques continue to advance the image recognition and object recognition domains,

each of which has broad applicability to a range of classification problems in biology. In this study, we demonstrate that the development of CNNs within the context of fish bioacoustics provides a useful tool for the extraction of relevant ecological information from PAM. Our CNN model performed at 94 % accuracy — a promising result considering the variation in acoustic conditions at this site over the 4 yr period.

The methodology presented in this study is an example of efficient means for processing PAM data with relatively little effort put into training a model

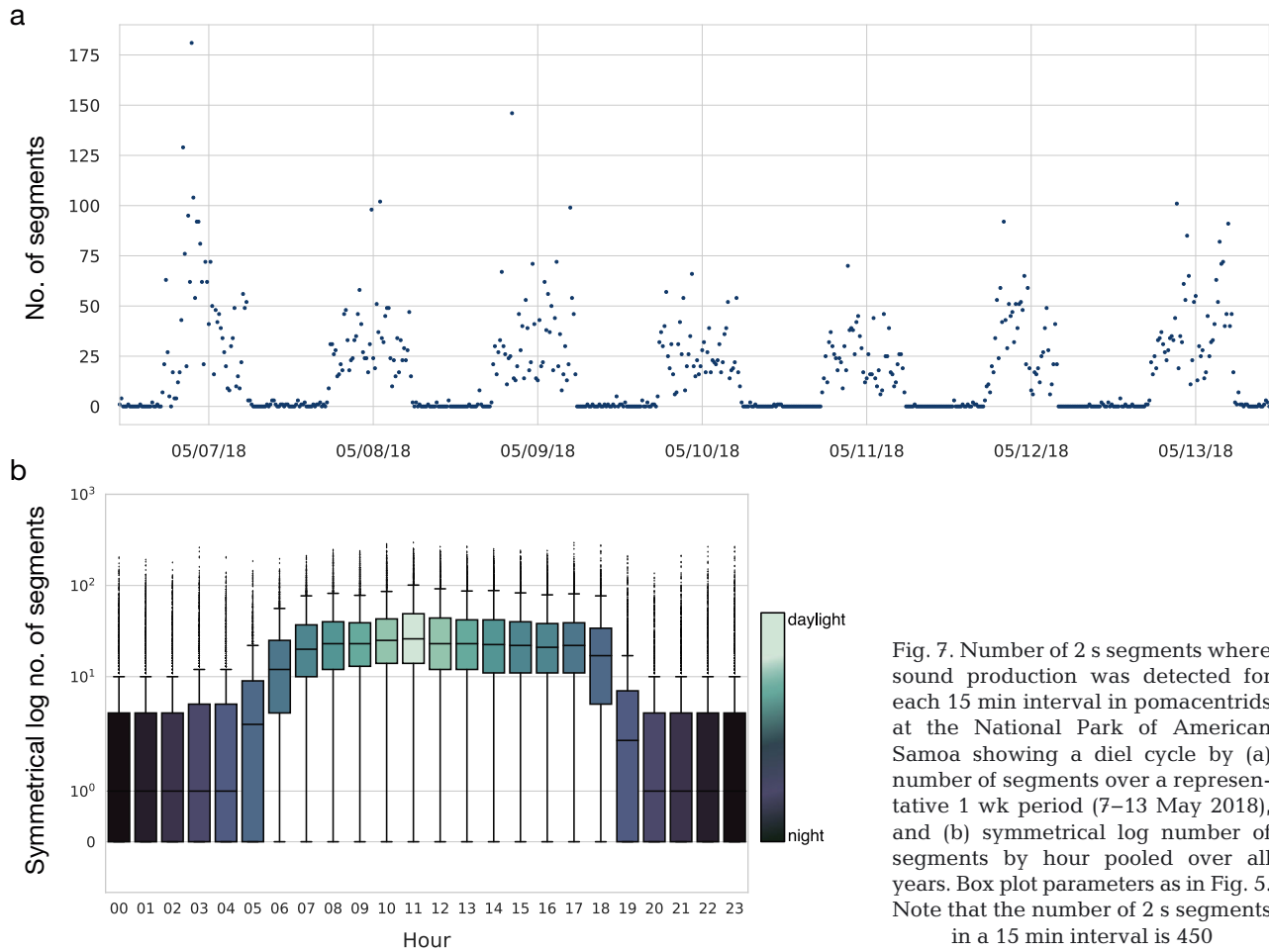


Fig. 7. Number of 2 s segments where sound production was detected for each 15 min interval in pomacentrids at the National Park of American Samoa showing a diel cycle by (a) number of segments over a representative 1 wk period (7–13 May 2018), and (b) symmetrical log number of segments by hour pooled over all years. Box plot parameters as in Fig. 5. Note that the number of 2 s segments in a 15 min interval is 450

Table 2. Zero-inflated negative binomial results for environmental parameters. Coefficients have been scaled exponentially. Asterisk (*) indicates statistical significance at $\alpha = 0.05$. SPL: sound pressure level; RMS: root mean square

| | Count model | | | Zero inflation model | | |
|---|-------------|--------|----------|----------------------|--------|----------|
| | Coefficient | SE | p | Coefficient | SE | p |
| Intercept | 4.8129 | 1.6074 | 0.0009* | 0 | 7.4932 | <0.0001* |
| Tidal amplitude (m) | 1.0524 | 1.0155 | 0.0009* | 0.7971 | 1.0664 | <0.0001* |
| Wind speed ($m s^{-1}$) | 1.0420 | 1.0024 | <0.0001* | 0.9994 | 1.0095 | 0.5636 |
| Sea surface temperature ($^{\circ}C$) | 0.7728 | 1.0081 | <0.0001* | 0.9852 | 1.0434 | 0.6572 |
| SPL _{RMS 50 Hz - 1.5 kHz} | 1.0775 | 1.0038 | <0.0001* | 1.0997 | 1.0149 | <0.0001* |

and no need for complex feature engineering. Specifically, leveraging transfer learning techniques and using the pre-trained weights of the ResNet-50 architecture allowed us to train a binary classification model to a high level of accuracy with a relatively small training dataset. When compared to a manual sound identification procedure, this is a considerable improvement in terms of the investment of time to achieve this level of ecological inference. It is important to point out that although we had access to a large dataset for this study, we only needed a small number of samples to train a model that performed well. Once the model was trained, we were then able to apply the model for inference over the entire 4 yr period to gain insight on ecological patterns.

4.2. Temporal distribution and patterns of pomacentrid sound production

Pomacentrids in NPSA are more acoustically active during daylight hours (Fig. 7), which is in line with their diurnal feeding and spawning behaviors, with some variation by species (Robertson 1984, Zemke-White et al. 2002). While the results of this study do not distinguish between feeding and spawning acoustic behaviors, they provide initial insight for how patterns in calling relate to feeding activity. Further, herbivorous pomacentrids are algal mat gardeners, and when feeding they are likely producing territorial call displays and having agonistic interactions with other fishes. Therefore, our diurnal call result suggests that the feeding pattern may be related to changes in the nutritional quality of the algae throughout the day (Randazzo Eisemann et al. 2019), with more agonistic interactions as food becomes scarce.

Longer-term variability in call patterns may provide insight into courtship activities and reproduction (Fig. 6). Reproduction in some pomacentrids is cued by lunar and tidal cycles, presumably to synchronize spawning with water conditions that are favorable for larval survival (Doherty 1983, Foster 1987). The interannual variation in call rates suggests heterogeneous occurrence of sound production behaviors across study years, with a slight decline in sound production rates over the first 3 yr of the study (Fig. 5, Table 1). Understanding the biological significance of these differences requires either separation of calling by behavioral state or integrating PAM methods with observational data. These results offer a promising method for detecting changes in the temporal occurrence of sound production related to

reproduction to ultimately provide key information for conservation management planning and actions.

Pomacentrids have been used as indicator species because of their influence on nutrient cycling and their important role as prey for commercially harvested fishes (Pinnegar 2018). Our work creates a baseline for the presence and seasonal patterns of pomacentrid sound production in NPSA, useful for ongoing monitoring and the detection of trends for the purposes of conservation (Teixeira et al. 2019). Building on the strong relationship between sound production behaviors and environmental variates, further research is recommended to determine potential ties between long-term patterns of sound occurrence and reef conditions, fishing pressure, noise pollution, water quality, and other anthropogenic influences. These data can supplement current monitoring efforts by the NPS for the key indicators of long-term impacts of climate change (National Park Service, date unknown).

4.3. Environmental parameters influencing sound occurrence

The relationship between sound occurrence and all environmental parameters was analyzed using the ZINB regression analysis described in Section 2.3.2. The results for the count portion of the model will be discussed separately for each parameter.

We found that the presence of pomacentrid sound increases as tidal amplitude increases. High tidal amplitude is associated with high flow velocity, which is important to planktivorous species of fishes (Eggertsen et al. 2016). Planktivorous species of pomacentrids may be more likely to elicit territorial calls while feeding on prey being brought in by the current. Additionally, because the lunar cycle influences both tidal amplitude and pomacentrid reproductive activities (Doherty 1983), there could be a more complex, multivariate influence on pomacentrid call rates. Using the methods developed in this study, future studies could explore this relationship in more detail using a hydrophone array for localization and cameras to capture behaviors.

Like tidal amplitude, the presence of pomacentrid sound increases in the recording as wind speed increases. Although ambient sound levels increase as wind speed increases (Wenz 1962, Haver et al. 2019), creating the potential for ambient noise to mask call detections, we found an increase in call presence, suggesting that these fish—similar to other species— increase sound production in noisy conditions (Picci-

ulin et al. 2012, Holt & Johnston 2014). More parsimoniously, the correlation between higher wind speed and increased sound production rates in this study is likely attributable to the co-occurrence of diel patterns of increased wind speed and fish sound production. A continued line of inquiry between high wind speed events and sound production may inform this question and provide a general inference related to other types of noise such as vessel traffic.

The presence of pomacentrid sound decreases as water temperature increases. Temperature affects the physiological process of sound production and behaviors associated with calling activity. In general, increases in water temperatures are correlated with increased frequency of sound production in several species of fishes (Ladich 2018). The exact role played by temperature in fish communication as described by Ladich (2018) is unclear, varies by species, and is confounded by seasonal effects. Tracking changes in sound production frequency may provide a supplemental tool for understanding the long-term impacts of increasing water temperatures on fishes.

The presence of pomacentrid sound increases with increased acoustic energy in $SPL_{RMS\ 50\ Hz - 1.5\ kHz}$. As the sound production behavior of the fish increases, the contribution to the soundscape results in an increase in overall soundscape energy. There are other contributors to the soundscape, including vessel traffic, which also influence the sound production behaviors of fish. This research does not address those relationships directly, but it allows us to establish baselines enabling research on the impacts of vessel noise on fishes.

The absence of sounds in the zero-inflated portion of the model could be due to the lack of fish presence, the receiver or model not detecting a sound, or misclassifying a pomacentrid sound as noise.

4.4. Detection ranges and ecological inference

We found that even using conservative estimates, the origin of pomacentrid sounds could be detected above ambient conditions at ranges of up to 200 m, indicating that PAM can be used to detect teleost sound production across a substantial section of reef tract. With this sound production detection range, the sounds themselves could be generated from planktivorous fish in the water column or residents of the nearby coral reef. However, our sound production analysis was at the family (pomacentrid) level, not at the species level. Therefore, it should be noted that while we have successfully developed a machine

learning approach to the detection of pomacentrid sound production and we provide initial insight into pomacentrid sound production, further ecological inference requires different types of data. Future work to pair behavioral observations of individual species with PAM-derived acoustic recordings will allow us to align PAM analysis with specific behavioral activities and thereby infer ecological significance based on sound type. Finally, it should be noted that overlapping sounds or the masking effects of other acoustic signals (abiotic, biotic, and anthropogenic) could also potentially interfere with the CNN analytic process. Additionally, our binary scoring method is not a true reflection of the sound production rate; rather, it is an estimation of the percentage of time in which the state of sound production is present (Altmann 1974) and a dominant feature in the soundscape.

5. CONCLUSIONS

This research provides an efficient methodology for characterizing the temporal distribution of pomacentrid sound production over a 4 yr period through the utilization of CNN techniques. Further, the results provide a baseline of pomacentrid sound production in a protected area (NPSA)—invaluable information for future opportunities to connect sound production occurrence with behavioral state and the detection of changes in pomacentrid acoustic activity related to environmental change. By creating the ability to rapidly and efficiently sample complete datasets collected from multiple years of monitoring, we can now generate a thorough and accurate picture of the natural scales of variation in sound production behaviors. The creation of a bioacoustic archive for pomacentrids, as we have done here, is an important step in identifying patterns of communication that shed light on the various behaviors and ecological implications related to sound production.

The connection between pomacentrids and reef health means that our CNN analytical approach is a viable monitoring tool with conservation and management implications. Healthy coral reefs have intrinsic value, but also provide services such as coastal protection and food resources. Coral reefs are also a source of rich biodiversity, making it for them easier to recover from repeated disturbances (Worm et al. 2006), but they are still susceptible to change. Our CNN machine learning technique, in concert with environmental monitoring, has the potential to provide an early warning system for perturbations in the

reef ecosystem, giving managers time to react to shifts in the environment.

As the adoption rate of machine learning grows in the field of marine science (Chérubin et al. 2020, Ibrahim et al. 2020, NOAA 2020), this momentum should be harnessed to exploit large volumes of data while improving the technology to increase efficiency and accuracy. Researchers will need to create a new training dataset for each application, which should mirror the unique conditions of each novel research problem to address the challenge of acoustic variability at different locations in the ocean. Once created, the CNN can efficiently map the acoustic profiles for that location. Similarly, reuse of models as acoustic conditions change over time is discouraged. However, transfer learning techniques eliminate the need for collecting extensive training datasets, reducing the investment of time, and making it easier to train new models with minimal effort. This research provides new opportunities for the exploration and monitoring of acoustic conditions, supporting the collaborative efforts of NOAA and NPS.

Supplementing these results with visual data — collected from camera traps, for example — would strengthen the inference drawn between sound production patterns and fish behavior, thereby enhancing the use of this technology as an ecological assessment and monitoring tool. By expanding this adaptable methodology to detect of other soniferous fishes (especially rare and cryptic species), mammals, invertebrates, and vessel noise signatures present in the acoustic data, PAM holds potential to understand interactions of acoustic features and changes in conditions. Machine learning can benefit from cooperation between the private and public sectors, leaving it in the hands of scientists to build on the preexisting framework and use the momentum to exploit big data in marine science. Innovative solutions such as the use of CNNs to detect fish sounds are urgently needed to supplement the efforts of NOAA and NPS to understand how conditions are changing over time in response to climate change and shifting practices in resource management.

Acknowledgements. This work was supported by the National Park Service (award P16AC00349 to the Cooperative Institute for Marine Ecosystem and Resource Studies, Oregon State University) as well as NOAA Pacific Marine Environmental Laboratory (OAR), Office of Science and Technology (NMFS), and Office of National Marine Sanctuaries (NOS). Funding for J.E.M. was also provided by the Cooperative Institute for Marine Ecosystem and Resources Studies (CIMERS) Fellowship for Undergraduate Excellence in Coastal and Ocean Studies. Publication of this paper was supported, in part, by the Henry Mastin Graduate Student

Fund. We are grateful to the team from the National Park of American Samoa, in particular Eric Brown, Burt Fuiava, and Ian Moffitt, for coral reef mapping and ecological insight, as well as Eva DiDonato, Brent Lignell, Katie Nuessly, Tammy Cook, Zandy Hillis-Starr, Tracy Ziegler, and Kurt Frstrup for their support of soundscape research in the National Park of American Samoa. We thank Peter Boyd, Taal Levi, and Chenxiao Hu for statistical expertise, Mick Thomure for collaboration on machine learning concepts, Ginger Ngo for illustration support, Lilli Wessling Hart for figure support, and Mary Nagle Wessling for editing and revision assistance. We also thank Noise Reference Station Network collaborators Leila Hatch, Sofie Van Parijs, Haru Matsumoto, T. K. Andy Lau, Lauren Roche, Brian Kahn, David K. Mellinger, Michael Craig, Christian Meinig, Carrie Wall Bell, and Charles Anderson. This paper is NOAA/PMEL contribution number 5185.

LITERATURE CITED

- ✦ Allen A (2018) OK Google: Find the humpback whales. NOAA Fisheries. <https://www.fisheries.noaa.gov/science-blog/ok-google-find-humpback-whales> (accessed 19 October 2020)
- ✦ Allen AN, Harvey M, Harrell L, Jansen A and others (2021) A convolutional neural network for automated detection of humpback whale song in a diverse, long-term passive acoustic dataset. *Front Mar Sci* 8:607321
- ✦ Altmann J (1974) Observational study of behavior: sampling methods. *Behaviour* 49:227–266
- ✦ Bergler C, Schröter H, Cheng RX, Barth V and others (2019) ORCA-SPOT: an automatic killer whale sound detection toolkit using deep learning. *Sci Rep* 9:10997
- ✦ Bermant PC, Bronstein MM, Wood RJ, Gero S, Gruber DF (2019) Deep machine learning techniques for the detection and classification of sperm whale bioacoustics. *Sci Rep* 9:12588
- ✦ Blasco-Moreno A, Pérez-Casany M, Puig P, Morante M, Castells E (2019) What does a zero mean? Understanding false, random and structural zeros in ecology. *Methods Ecol Evol* 10:949–959
- ✦ Blum HC, Bracco R (2019) Fastai_audio. https://github.com/fastaudio/fastai_audio
- ✦ Chérubin LM, Dalgleish F, Ibrahim AK, Schärer-Umpierre M, Nemeth RS, Matthews A, Appeldoorn R (2020) Fish spawning aggregations dynamics as inferred from a novel, persistent presence robotic approach. *Front Mar Sci* 6:779
- ✦ Cole KS (2010) Reproduction and sexuality in marine fishes: patterns and processes. University of California Press, Berkeley, CA
- ✦ Darling ES, Graham NAJ, Januchowski-Hartley FA, Nash KL, Pratchett MS, Wilson SK (2017) Relationships between structural complexity, coral traits, and reef fish assemblages. *Coral Reefs* 36:561–575
- ✦ Doherty PJ (1983) Diel, lunar and seasonal rhythms in the reproduction of two tropical damselfishes: *Pomacentrus flavicauda* and *P. wardi*. *Mar Biol* 75:215–224
- ✦ Doney SC, Ruckelshaus M, Duffy JE, Barry JP and others (2012) Climate change impacts on marine ecosystems. *Annu Rev Mar Sci* 4:11–37
- ✦ Eggertsen L, Hammar L, Gullström M (2016) Effects of tidal current-induced flow on reef fish behaviour and function on a subtropical rocky reef. *Mar Ecol Prog Ser* 559:175–192

- ✦ Emslie MJ, Logan M, Cheal AJ (2019) The distribution of planktivorous damselfishes (Pomacentridae) on the Great Barrier Reef and the relative influences of habitat and predation. *Diversity* 11:33
- ✦ Foster SA (1987) Diel and lunar patterns of reproduction in the Caribbean and Pacific sergeant major damselfishes *Abudefduf saxatilis* and *A. troschelii*. *Mar Biol* 95: 333–343
- ✦ Frédéric B, Parmentier E (eds) (2016) *Biology of damselfishes*. CRC Press, Boca Raton, FL
- ✦ Harvey M (2018) Acoustic detection of humpback whales using a convolutional neural network. <http://ai.googleblog.com/2018/10/acoustic-detection-of-humpback-whales.html> (accessed 6 November 2020)
- ✦ Haver SM, Gedamke J, Hatch LT, Dziak RP and others (2018) Monitoring long-term soundscape trends in US waters: the NOAA/NPS Ocean Noise Reference Station Network. *Mar Policy* 90:6–13
- ✦ Haver SM, Fournet MEH, Dziak RP, Gabriele C and others (2019) Comparing the underwater soundscapes of four US national parks and marine sanctuaries. *Front Mar Sci* 6:500
- ✦ He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 27–30 June, 2016, IEEE, Las Vegas, NV, p 770–778
- ✦ Hinton G, Deng L, Yu D, Dahl GE and others (2012) Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Process Mag* 29:82–97
- ✦ Holt DE, Johnston CE (2014) Evidence of the Lombard effect in fishes. *Behav Ecol* 25:819–826
- ✦ Ibrahim AK, Zhuang H, Chérubin LM, Schärer-Umpierre MT, Nemeth RS, Erdol N, Ali AM (2020) Transfer learning for efficient classification of grouper sound. *J Acoust Soc Am* 148:EL260–EL266
- ✦ Ladich F (2018) Acoustic communication in fishes: temperature plays a role. *Fish Fish* 19:598–612
- ✦ Ladich F, Fine ML (2006) Sound-generating mechanisms in fishes: a unique diversity in vertebrates. In: Ladich F, Collin SP, Moller P, Kapoor BG (eds) *Communication in fishes*. Science Publishers, Enfield, NH, p 3–43
- ✦ Lin TH, Tsao Y, Akamatsu T (2018) Comparison of passive acoustic soniferous fish monitoring with supervised and unsupervised approaches. *J Acoust Soc Am* 143:EL278
- ✦ Malfante M, Mars JI, Dalla Mura M, Gervaise C (2018) Automatic fish sounds classification. *J Acoust Soc Am* 143:2834–2846
- ✦ Mann DA, Lobel PS (1998) Acoustic behavior of the damselfish *Dascyllus albisella*: behavioral and geographic variation. *Environ Biol Fishes* 51:421–428
- ✦ Maruska KP, Boyle KS, Dewan LR, Tricas TC (2007) Sound production and spectral hearing sensitivity in the Hawaiian sergeant damselfish, *Abudefduf abdominalis*. *J Exp Biol* 210:3990–4004
- ✦ MATLAB (2018) Version 9.5.0 (R2018b). The MathWorks Inc., Natick, MA
- ✦ Metz CE (1978) Basic principles of ROC analysis. *Semin Nucl Med* 8:283–298
- ✦ Munger LM, Lammers MO, Fisher-Pool P, Wong K (2012) Humpback whale (*Megaptera novaeangliae*) song occurrence at American Samoa in long-term passive acoustic recordings, 2008–2009. *J Acoust Soc Am* 132:2265–2272
- ✦ National Park Service (date unknown) National Park of American Samoa [foundation document]. National Park Service, US Department of the Interior, Pago Pago (PDF document created and modified 2017, accessed 26 June 2021). <http://npshistory.com/publications/foundation-documents/npsa-fd-overview.pdf>
- ✦ NOAA (2020) NOAA Artificial Intelligence Strategy. <https://sciencecouncil.noaa.gov/Portals/0/2020%20AI%20Strategy.pdf?ver=2020-09-17-150016-857> (accessed 10 October 2020)
- ✦ NOAA National Ocean Service (1971) Station NSTP6 - 1770000 - Pago Pago, American Samoa. National data buoy center. https://www.ndbc.noaa.gov/station_page.php?station=nstp6 (accessed 1 December 2019)
- ✦ NOAA Science Council (date unknown) NOAA Science & Technology Focus Areas. <https://sciencecouncil.noaa.gov/NOAA-Science-Technology-Focus-Areas> (accessed 6 November 2020)
- ✦ Noda JJ, Travieso CM, Sanchez-Rodríguez D (2016) Automatic taxonomic classification of fish based on their acoustic signals. *Appl Sci* 6:443
- ✦ Noonan SHC, Jones GP, Pratchett MS (2012) Coral size, health and structural complexity: effects on the ecology of a coral reef damselfish. *Mar Ecol Prog Ser* 456:127–137
- ✦ Parmentier E, Lecchini D, Frederich B, Brie C, Mann D (2009) Sound production in four damselfish (*Dascyllus*) species: phyletic relationships? *Biol J Linn Soc* 97: 928–940
- ✦ Pedregosa F, Varoquaux G, Gramfort A, Michel V and others (2011) Scikit-learn: machine learning in Python. *J Mach Learn Res* 12:2825–2830
- ✦ Picciulin M, Sebastianutto L, Codarin A, Calcagno G, Ferrero EA (2012) Brown meagre vocalization rate increases during repetitive boat noise exposures: a possible case of vocal compensation. *J Acoust Soc Am* 132:3118–3124
- ✦ PIFSC (Pacific Islands Fisheries Science Center) (2011) Coral reef ecosystems of American Samoa: a 2002–2010 overview. PIFSC Special Publication. NOAA Fisheries Pacific Islands Fisheries Science Center, Honolulu, HI
- ✦ Pinnegar JK (2018) Why the damselfish *Chromis chromis* is a key species in the Mediterranean rocky littoral—a quantitative perspective. *J Fish Biol* 92:851–872
- ✦ R Core Team (2021) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
- ✦ Randazzo Eiseemann Á, Montero Muñoz JL, McField M, Myton J, Arias-González JE (2019) The effect of algal-gardening damselfish on the resilience of the Mesoamerican reef. *Front Mar Sci* 6:414
- ✦ Robertson DR (1984) Cohabitation of competing territorial damselfishes on a Caribbean coral reef. *Ecology* 65: 1121–1135
- ✦ Russakovsky O, Deng J, Su H, Krause J and others (2015) ImageNet large scale visual recognition challenge. *Int J Comput Vis* 115:211–252
- ✦ Sattar F, Cullis-Suzuki S, Jin F (2016) Identification of fish vocalizations from ocean acoustic data. *Appl Acoust* 110: 248–255
- ✦ Shiu Y, Palmer KJ, Roch MA, Fleishman E and others (2020) Deep neural networks for automated detection of marine mammal species. *Sci Rep* 10:607
- ✦ Teixeira D, Maron M, Rensburg BJ (2019) Bioacoustic monitoring of animal vocal behavior for conservation. *Conserv Sci Pract* 1:e72
- ✦ The Cornell Lab of Ornithology (2014) Raven Pro: Interactive Sound Analysis Software. Center for Conservation Bioacoustics, Ithaca, NY

- ✦ Tolimieri N, Haine O, Jeffs A, Mccauley R, Montgomery J (2004) Directional orientation of pomacentrid larvae to ambient reef sound. *Coral Reefs* 23:184–191
- ✦ Vieira M, Fonseca PJ, Amorim MCP, Teixeira CJC (2015) Call recognition and individual identification of fish vocalizations based on automatic speech recognition: an example with the Lusitanian toadfish. *J Acoust Soc Am* 138:3941–3950
- ✦ Weimann SR, Black AN, Leese J, Richter ML, Itzkowitz M, Burger RM (2018) Territorial vocalization in sympatric damselfish: acoustic characteristics and intruder discrimination. *Bioacoustics* 27:87–102
- ✦ Wenz GM (1962) Acoustic ambient noise in the ocean: spectra and sources. *J Acoust Soc Am* 34:1936
- ✦ Worm B, Barbier EB, Beaumont N, Duffy JE and others (2006) Impacts of biodiversity loss on ocean ecosystem services. *Science* 314:787–790
- ✦ Zeileis A, Kleiber C, Jackman S (2008) Regression models for count data in R. *J Stat Soft* 27:1–25
- ✦ Zemke-White L, Choat J, Clements K (2002) A re-evaluation of the diel feeding hypothesis for marine herbivorous fishes. *Mar Biol* 141:571–579
- ✦ Zhong M, Castellote M, Dodhia R, Lavista Ferres J, Keogh M, Brewer A (2020) Beluga whale acoustic signal classification using deep learning neural network models. *J Acoust Soc Am* 147:1834–1841

*Editorial responsibility: Elliott Hazen,
Pacific Grove, California, USA
Reviewed by: E. Ozanich, J. Butler, M. Lammers*

*Submitted: January 25, 2021
Accepted: September 21, 2021
Proofs received from author(s): November 25, 2021*