

Comparison of passive acoustic soniferous fish monitoring with supervised and unsupervised approaches

Tzu-Hao Lin,¹ Yu Tsao,² and Tomonari Akamatsu^{3,a)}

¹*Department of Marine Biodiversity Research, Japan Agency for Marine-Earth Science and Technology, 2-15, Natsushima, Yokosuka City, Kanagawa, 237-0061, Japan*

²*Research Center for Information Technology Innovation, Academia Sinica, Number 128, Section 2, Academia Road, Taipei 115, Taiwan, Republic of China*

³*National Research Institute of Fisheries Science, Japan Fisheries Research and Education Agency, 2-12-4 Fukuura, Yokohama, Kanagawa 236-8648, Japan*
schonkopf@gmail.com, yu.tsao@citi.sinica.edu.tw, akamatsu@affrc.go.jp

Abstract: Passive acoustics has been used to investigate behavior and relative abundances of soniferous fish. However, because of noise interferences, it remains challenging to accurately analyze acoustic activities of soniferous fish. This study proposes a multi-method approach, which combines rule-based detector, periodicity-coded non-negative matrix factorization, and Gaussian mixture models. Although the three methods performed well when used to detect croaker choruses in quiet conditions, inconsistent results are observed in noisy conditions. A consistency matrix can provide insights regarding the bias of acoustic monitoring results. The results suggest that the proposed approach can reasonably improve passive acoustic monitoring of soniferous fish.

© 2018 Acoustical Society of America

[WJL]

Date Received: November 19, 2017 **Date Accepted:** April 9, 2018

1. Introduction

Information concerning fish population dynamics is important for conservation management of marine ecosystems (Russ and Alcala, 1996). Visual observation of fish behavior is an infeasible approach because of the limited visibility of underwater environments. Consequently, passive acoustics has been employed to investigate habitat use, spawning areas, and relative abundances of soniferous fish (Rountree *et al.*, 2006; Luczkovich *et al.*, 2008).

In conventional approaches, experts were recruited to manually identify fish sounds by directly listening to audio clips or inspecting spectrograms (Mann *et al.*, 2008; Wall *et al.*, 2013). To facilitate the analysis of long-duration recordings, computational signal processing techniques have also been employed (Sánchez-Gendriz and Padovese, 2017). Most of the detection algorithms used to identify fish sounds were rule-based, where the rules were defined by expert knowledge (Stolkin *et al.*, 2007; Mann *et al.*, 2008). It is noted that misdetections could occur in noisy or other unexpected conditions. Thus, it is necessary to evaluate the variability of performance in different recording conditions to avoid potential biases (Miyajima-Taga *et al.*, 2016).

One way to facilitate the acoustic variability evaluation is to investigate temporal-spectral patterns in data by using an unsupervised clustering approach, such as *k*-means clustering or Gaussian mixture models (GMMs). These models can cluster samples into groups based on the dispersion of temporal-spectral characteristics without a pre-defined rule. These unsupervised clustering approaches have been predominantly applied in the analysis of call repertoire (Terhune *et al.*, 1993; Lin *et al.*, 2015). The clustering technique has also been employed to identify different animal choruses without a recognition database (Lin *et al.*, 2017b). Therefore, the variation of automatic detection precisions in different recording conditions, which show variations in noise levels or spectral features, can be investigated by combining the rule-based and unsupervised clustering approaches.

Another effective way to reduce potential biases is to separate noise by a de-noising algorithm and subsequently detect the sound of soniferous fish. A feasible class of de-noising algorithm is blind source separation (BSS). BSS has been widely deployed in speech and musical applications, but rarely adopted in the field of animal bioacoustics. Recently, Lin *et al.* (2017a) demonstrated that biological choruses displayed on a long-

^{a)} Author to whom correspondence should be addressed.

term spectrogram can be significantly enhanced using the periodicity-coded non-negative matrix factorization (PC-NMF). The PC-NMF is an unsupervised learning algorithm that aims to separate biological choruses from other noise sources according to the different periodicity properties of diurnal cycles. Because many fish in shallow marine environments display nighttime chorusing behavior (Rountree *et al.*, 2006; Luczkovich *et al.*, 2008), the PC-NMF is a potential solution to suppress unwanted noise components.

In the present study, we employed a rule-based fish sound detector, the PC-NMF, and GMMs, respectively, to analyze the number of fish calls, the relative strength of fish chorus, and the posterior probability of fish chorus in each 5-min recording clip. We compared our results analyzed from two nearby recording stations with different levels of mooring noise. The potential bias in acoustic monitoring results can then be identified by measuring the correlations among outputs of the three analysis approaches.

2. Methods

Five AUSOMS-mini sound recorders (AquaSound, Inc., Kobe, Japan) were deployed off-coast at Kashima-nada, Ibaraki Prefecture, Japan. These recording devices were active continuously from June 22 to July 12, 2015, and from August 4 to September 28, 2015. The five recorders spanned an area, square in shape, approximately 2 km in length. The water depth of the recording area ranges between 10 and 25 m. In this study, we only analyzed data collected at Station 3 and Station 5 during the period from August 4 to September 28, 2015. Station 3 (N35° 54' 6.31" E140° 44' 3.21") was deployed in relatively shallow water in comparison to Station 5 (N35° 54' 24.14" E140° 44' 40.81"). At Station 5, strong mooring noises were recorded during the study period. Therefore, Station 3 and Station 5 were considered to be the quiet and noisy recording sites, respectively. The AUSOMS-mini recorder can record linear pulse-code modulation (PCM) data between 70 and 160 dB re 1 μ Pa, with a flat frequency response between 0.1 and 23 kHz. Because of limitations in memory size (a 4 GB internal memory and 32 GB micro SD card), underwater sounds were sampled in 44.1 kHz and recorded in MP3 format (128 kbps) to extend the recording length. However, the highest frequency for acoustic analysis was limited to 16 kHz because of a lossy compression.

To detect fish sounds, three types of algorithm were employed in our acoustic analysis: a rule-based fish sound detector, the PC-NMF, and GMMs. The rule-based fish sound detector (hereafter rule-based detector) measures the integrated band level between 400 and 800 Hz in each section of 2^{14} samples (371 ms) for initial screening. The frequency band and screening window were selected according to the dominant frequency of recorded fish calls by manually investigating spectrograms. When the root-mean-square band level was seen to exceed a pre-defined threshold of 105 dB re 1 μ Pa, and the peak frequency was within the focal band (400–800 Hz), the waveform envelope was calculated using the Hilbert-Huang transform. Then, the time of each pulse can be detected by using the *findpeak* function of MATLAB 2015b (MathWorks, Natick, MA). Short inter-pulsepeak interval (<10 ms) were excluded to reduce false alarms, as the regular inter-pulsepeak intervals of croakers' sounds are usually longer than 10 ms (Mok *et al.*, 2011; Wang *et al.*, 2017). We only included the sequences of pulses with a mean inter-pulsepeak interval between 10 and 30 ms, and standard deviation of the inter-pulsepeak interval within 20 ms. In addition, only pulse trains with six or more pulses were recognized as croaker sounds (Mok *et al.*, 2011; Wang *et al.*, 2017). After applying this rule-based filtering procedure, we calculated the number of fish calls at 5-min intervals for later comparisons.

The PC-NMF is a BSS algorithm that can separate sound sources with different patterns of periodical occurrence. It is known that periodic structures, especially those with 24-h intervals, suggest biological events driven by a diurnal pattern (Farina and James, 2016; Sánchez-Gendriz and Padovese, 2017). This was our only assumption when extracting animal sounds using PC-NMF. Figure 1 shows a long-term spectrogram of underwater recordings collected at Station 5, and the enhanced fish chorus obtained using the PC-NMF. A long-term spectrogram is a visualization of long-duration recordings, based on the power spectrum of various short recording clips (Lin *et al.*, 2017a). In this study, we removed the first 0.1-s fragment to prevent bias of spectral analysis due to the procedure of MP3 decoding and then measured the median power spectrum for each 5-min interval. Although the fine details within each 5-min recording window were smoothed out, the presence of mooring noise, shipping noise, and fish chorus could still be present on the long-term spectrogram, as shown in the upper panel of Fig. 1. The PC-NMF enhances fish choruses according to two learning stages. In the first stage, the spectrogram is factorized into a basis matrix and a separate encoding matrix. The basis matrix represents a collection of spectral components, and the encoding matrix describes

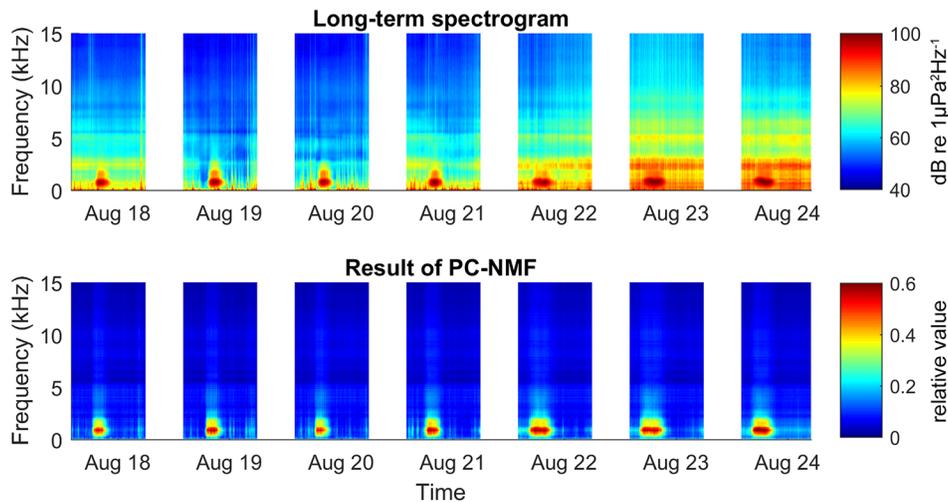


Fig. 1. (Color online) A long-term spectrogram of underwater recordings and enhanced fish chorus using the PC-NMF method. The top panel shows the long-term spectrogram of 1-week long recordings collected at Station 5, in which the blank areas indicate periods without recording. The lower panel shows the relative change in chorusing level estimated by the PC-NMF.

the time activation of each basis. In the second stage, the diurnal periodicity of each basis matrix is measured by performing a discrete Fourier transform on the components of the encoding matrix. Two groups of bases are then clustered according to the periodicity information by a secondary NMF process (Lin *et al.*, 2017a). Finally, fish choruses can be reconstructed from the associated encoding information of the spectral bases, which show strong diurnal periodicity. At the same time, noise components (mooring noise, shipping noise, etc.) can be suppressed, as the stochastic nature of these noise sources means that they do not present diurnal periodicity. From a reconstructed fish chorus, we can estimate the change in chorusing level (P_c) based on a gain function

$$P_c = P \times G, \quad (1)$$

where P represents the sound pressure level of each 5-min recording and G represents the ratio estimated by the PC-NMF. Because the BSS was performed on a logarithmically scaled spectrogram, we can only estimate the signal-noise ratio (SNR) of the fish chorus based on the median value of P_c .

Finally, we applied GMMs to identify different audio clusters based on the median power spectrums of each 5-min recording. To reduce the feature dimension and eliminate redundant information in the power spectra, we employed principle component analysis, selecting the set of components retaining at least 90% of the total variance (Lin *et al.*, 2017b). After dimensional reduction, GMMs with a full covariance matrix and a hard classification approach were used to perform the clustering. Traditionally, the choice of the number of clusters, k , can be subjectively or empirically determined. In this study, we varied the value of k over the range $3 \leq k \leq 20$ to determine the optimal number of clusters. For each value of k , we repeated the GMMs 20 times and measured the dispersion coefficient (Kim and Park, 2008), which is a method to quantify the clustering consistency. The value of k corresponding to the highest dispersion coefficient was selected as the final value. After clustering, we manually identified the cluster associated with fish choruses based on its spectral features and temporal behavior. Therefore, we were able to obtain the posterior probability of each recording clip belongs to the fish chorus cluster.

To quantify the consistency of acoustic monitoring results, we measured the correlation coefficients among the three outputs. A high correlation coefficient suggests two outputs perform consistently; on the contrary, a low correlation coefficient indicates the two outputs disagree with each other. This procedure was repeated for each recording day so that we could inspect the variability of consistency among the three outputs.

3. Results and discussion

At Station 3, the rule-based detector, the PC-NMF, and GMMs showed clear agreement with respect to the changing diurnal patterns of fish chorus. The results showed that the period between 7 and 10 pm appears to be the primary chorusing period of soniferous fish off-coast at Kashima-nada (Fig. 2). The rule-based detector was designed according to our knowledge regarding the acoustic behavior of croakers, and hence performed

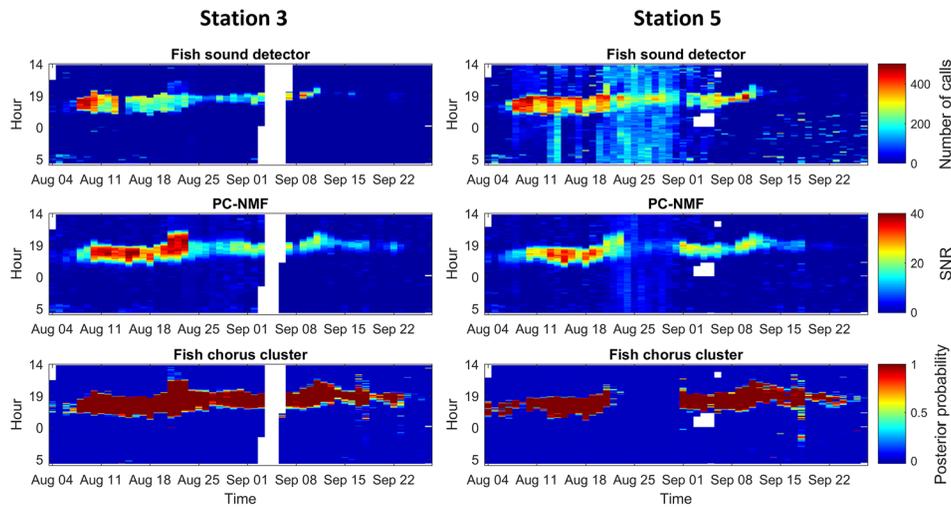


Fig. 2. (Color online) Analysis results of fish chorus recorded at Station 3 (left panels) and Station 5 (right panels) off-coast at Kashima-nada. The top panels show the number of calls detected by the rule-based fish sound detector. The center panels show the signal-noise ratio (SNR) of fish chorus estimated by the separation result of PC-NMF. The lower panels show the posterior probability of a fish chorus cluster identified by GMMs. Blank areas represent periods without recording.

reasonably well in most of the chorusing periods. In the meanwhile, the PC-NMF and GMMs (where no prior information of fish sounds is involved) also performed well; because no prior information is required, these methods can be suitably used for the initial screening of unknown naive recordings.

Although the identified chorusing periods at Station 3 were similar among the three analysis approaches, correlation coefficients indicate inconsistency among the three outputs when weak fish choruses were recorded, such as periods before Aug 9 or after September 8 (Fig. 3). Inconsistent results could happen when lots of false negatives were reported by the rule-based detector. Even for those days with strong consistency, we still noticed different changing patterns of fish chorusing behavior between the rule-based detector and the PC-NMF. For example, the number of calls at Station 3 reached a peak of 512 calls per 5-min period on August 9, and then gradually reduced to less than 300 calls per 5-min period. However, the results of analysis indicated that the relative received level of fish choruses became higher after August 8 and peaked, at 40 dB, on August 23 (Fig. 2). During prominent chorusing events, the pulsed structure of a fish call is not distinguishable when many calls are recorded simultaneously. As the rule-based detector was designed to identify the sequential structure of pulsed sounds, it may not be able to detect fish calls accurately during a noisy chorusing condition.

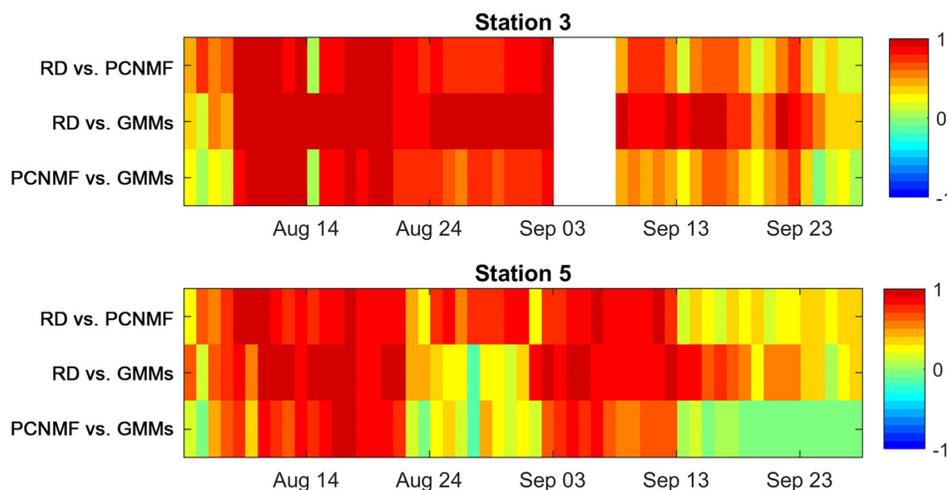


Fig. 3. (Color online) Correlation analysis among three outputs of the rule-based detector (RD), the PC-NMF, and GMMs. Correlation coefficients were measured in each recording day, and blank areas represent periods without recording.

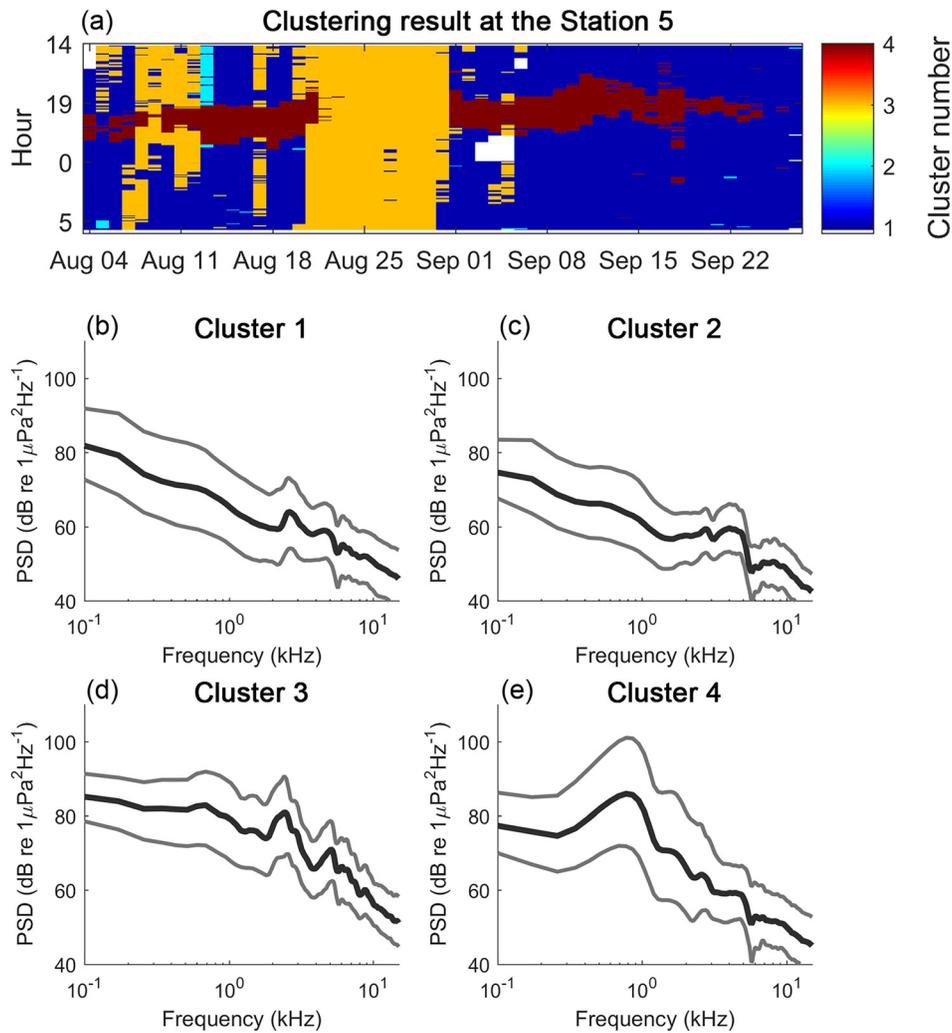


Fig. 4. (Color online) Temporal occurrence and spectral characteristics of the four clusters at Station 5. (a) The clustering result was analyzed using GMMs. Blank areas represent periods without recording. (b)–(e) Summary of the spectral characteristics of each cluster. Each panel shows the 25th, 50th, and 75th percentiles of the power spectral densities in different frequency ranges.

Relative to Station 3, Station 5 represented a noisier recording site. The results from Station 5 showed that the three analysis methods performed inconsistently from August 22 to September 2 (Fig. 3). Large numbers of fish sounds were detected before 7 pm and after 10 pm from the same period by using the rule-based detector. However, the PC-NMF was only able to detect weak choruses within the same period. In addition, the fish chorus cluster identified by GMMs also displayed a very low posterior probability (Fig. 2). Although at present the three methods cannot perform well in noisy conditions, we can still identify false positives for the rule-based detector. We can also identify false negatives for the two unsupervised approaches according to the results of rule-based detector. Therefore, the correlations among the three outputs can be employed as essential indicators for the uncertainty of passive acoustic monitoring.

To investigate the potential cause of the inconsistent results between different analysis methods, we investigated the spectral features of the four clusters identified by GMMs at Station 5 (Fig. 4). The third cluster appears to be highly influenced by mooring noise, which displayed a broadband structure for the frequency range below 2.5 kHz. The time at which the third cluster occurred at Station 5 represents heavy contamination of mooring and wave noises caused by stormy weather. This well explained the inconsistent results that can be noted in Fig. 3.

Lots of factors can affect the precision of automatic detection and classification. Most bioacoustics studies manually annotate data and measure the detection and classification accuracy to evaluate the uncertainty of passive acoustic monitoring (Acevedo *et al.*, 2009; Dugan *et al.*, 2010). However, it is impossible to measure the detection and classification accuracy in all recording conditions. A rule-based detector may perform poorly when an unfamiliar noise was encountered. Moreover, misdetections could also

happen when multiple sound sources overlap with each other. Although the precision remains uncertain, supervised approaches can still be an important analysis method in passive acoustic monitoring of soniferous fish. In the future, supervised approaches can be first used to investigate the occurrence of soniferous animals and then unsupervised approaches can be employed to identify potential defects. In this way, the potential biases of passive acoustic monitoring can be reduced on the basis of the multi-method approach proposed in this study.

4. Conclusions

The three methods employed in this study, including a rule-based detector and two unsupervised learning algorithms, performed well when identifying fish chorusing events provided that the ambient noise level was sufficiently low. However, the three approaches show defects in a different recording condition. By quantifying the consistency among the three outputs, false positives and false negatives of acoustic monitoring results can be identified. In the future, more acoustic recorders will be deployed in different marine habitats. It is unlikely that we can manually identify a large number of ground truths and measure the accuracy of our detectors in all recording conditions. Nevertheless, we can employ unsupervised learning algorithms to investigate the variability of long-term recordings, and subsequently optimize our rule-based detectors by inspecting those inconsistent acoustic monitoring results. Based on this multi-method approach, the uncertainty of passive acoustic monitoring can be minimized, and the screening effort required for bioacoustics studies can also be reduced. We believe that by using the proposed multi-method approach, the decisions involved in conservation management can be better-informed by passive acoustic monitoring.

Acknowledgments

This work was supported by the JST CREST Grant No. JPMJCR11A1 (Japan) and the Ministry of Science and Technology, Taiwan (Republic of China) under the project entitled “Investigation on the interactions between ecological environment, wildlife animals, and human activities using soundscape information” (Grant No. MOST105-2321-B-001-069-MY3).

References and links

- Acevedo, M. A., Corrada-Bravo, C. J., Corrada-Bravo, H., Villanueva-Rivera, L. J., and Aide, T. M. (2009). “Automated classification of bird and amphibian calls using machine learning: A comparison of methods,” *Ecol. Inf.* **4**, 206–214.
- Dugan, P. J., Rice, A. N., Urazghildiev, I. R., and Clark, C. W. (2010). “North Atlantic right whale acoustic signal processing: Part I. comparison of machine learning recognition algorithms,” in *2010 Long Island Systems, Applications and Technology Conference (IEEE)*, pp. 1–6.
- Farina, A., and James, P. (2016). “The acoustic communities: Definition, description and ecological role,” *Biosystems* **147**, 11–20.
- Kim, J., and Park, H. (2008). “Sparse nonnegative matrix factorization for clustering,” CSE technical reports, Georgia Institute of Technology, Atlanta, 2008.
- Lin, T.-H., Fang, S.-H., and Tsao, Y. (2017a). “Improving biodiversity assessment via unsupervised separation of biological sounds from long-duration recordings,” *Sci. Rep.* **7**, 4547.
- Lin, T.-H., Tsao, Y., Wang, Y.-H., Yen, H.-W., and Lu, S.-S. (2017b). “Computing biodiversity change via a soundscape monitoring network,” in *PNC 2017 Annual Conference and Joint Meetings*, pp. 128–133.
- Lin, T.-H., Yu, H.-Y., Chen, C.-F., and Chou, L.-S. (2015). “Passive acoustic monitoring of the temporal variability of odontocete tonal sounds from a long-term marine observatory,” *PLoS One* **10**, e0123943.
- 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.
- Mann, D. A., Hawkins, A. D., and Jech, J. M. (2008). “Active and passive acoustics to locate and study fish,” in *Fish Bioacoustics*, edited by J. F. Webb, R. R. Fay, and A. N. Popper (Springer, Berlin), pp. 279–309.
- Miyajima-Taga, Y., Akamatsu, T., Matsuo, I., Takahashi, M., Matsuzaki, H., and Nijima, K. (2016). “The study of automatic detection of croaker sounds,” *J. Marine Acoust. Soc. Jpn.* **43**, 116–125.
- Mok, H.-K., Lin, S.-Y., and Tsai, K.-E. (2011). “Underwater ambient biological noise in the waters on the west coast of Taiwan,” *Kuroshio Sci.* **5**, 51–57.
- Rountree, R. A., Gilmore, R. G., Goudey, C. A., Hawkins, A. D., Luczkovich, J. J., and Mann, D. A. (2006). “Listening to fish: Applications of passive acoustics to fisheries science,” *Fisheries* **31**, 433–446.
- Russ, G. R., and Alcalá, A. C. (1996). “Marine reserves: Rates and patterns of recovery and decline of large predatory fish,” *Ecol. Appl.* **6**, 947–961.
- Sánchez-Gendríz, I., and Padovese, L. R. (2017). “A methodology for analyzing biological choruses from long-term passive acoustic monitoring in natural areas,” *Ecol. Inf.* **41**, 1–10.

- Stolkin, R., Radhakrishnan, S., Sutin, A., and Rountree, R. (2007). "Passive acoustic detection of modulated underwater sounds from biological and anthropogenic sources," in *OCEANS 2007*, pp. 1–8.
- Terhune, J. M., Burton, H., and Green, K. (1993). "Classification of diverse call types using cluster analysis techniques," *Bioacoustics* **4**, 245–258.
- Wall, C. C., Simard, P., Lembke, C., and Mann, D. A. (2013). "Large-scale passive acoustic monitoring of fish sound production on the West Florida Shelf," *Marine Ecol. Prog. Ser.* **484**, 173–188.
- Wang, Z.-T., Nowacek, D. P., Akamatsu, T., Wang, K.-X., Liu, J.-C., Duan, G.-Q., Cao, H.-J., and Wang, D. (2017). "Diversity of fish sound types in the Pearl River Estuary, China," *PeerJ* **5**, e3924.