Automated detection of broadband clicks of freshwater fish using spectro-temporal features[a)]

Navinda Kottege[b)] and Raja Jurdak
Commonwealth Scientific and Industrial Research Organization, P.O. Box 883, Kenmore, Queensland 4069, Australia

Frederieke Kroon[c)] and Dean Jones
CSIRO, P.O. Box 780, Atherton, Queensland 4883, Australia

(Received 18 June 2014; revised 22 February 2015; accepted 28 March 2015)

Large scale networks of embedded wireless sensor nodes can passively capture sound for species detection. However, the acoustic recordings result in large amounts of data requiring in-network classification for such systems to be feasible. The current state of the art in the area of in-network bioacoustics classification targets narrowband or long-duration signals, which render it unsuitable for detecting species that emit impulsive broadband signals. In this study, impulsive broadband signals were classified using a small set of spectral and temporal features to aid in their automatic detection and classification. A prototype system is presented along with an experimental evaluation of automated classification methods. The sound used was recorded from a freshwater invasive fish in Australia, the spotted tilapia (*Tilapia mariae*). Results show a high degree of accuracy after evaluating the proposed detection and classification method for *T. mariae* sounds and comparing its performance against the state of the art. Moreover, performance slightly improves when the original signal was down-sampled from 44.1 to 16 kHz. This indicates that the proposed method is well-suited for detection and classification on embedded devices, which can be deployed to implement a large scale wireless sensor network for automated species detection.

© *2015 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4919298]

[PJL] Pages: 2502–2511

I. INTRODUCTION

Bioacoustics monitoring is often used in ecology as a passive and non-invasive method to quantify the presence and abundance of a species of interest, e.g., whales,[1,2] birds,[3] bats,[4] toads,[5] and frogs.[6] The long term vision and motivation is to monitor and protect sensitive ecosystems using very large scale networks of bioacoustic sensor nodes, which highlights the importance of automated bioacoustic classification of species. Machine learning and artificial intelligence techniques have been successfully used for sound detection and classification based on the *narrowband* sound emitted by these species. While a few recent studies have characterized the *broadband* sound features of fish species in the marine[7] and freshwater[8,9] environments, these studies did not investigate how to automatically detect and classify these sounds.

Current detection and classification methods for acoustic signals are based on the assumption that sounds are either narrowband or have a relatively long duration (i.e., in the order of seconds). One such method is spectral correlation that assumes non-impulsive calls with narrow short-time frequency bandwidths consisting of a sequence of frequency up-sweeps and down-sweeps.[10] To encompass the shape of the spectrogram of the target call, a correlation kernel is constructed in a piece-wise manner. This kernel is then moved along the spectrogram of the input signal and cross-correlated in the time domain to obtain a recognition score function. Thresholding this function yields detections of the target calls. Typical call durations used with this technique are in the order of seconds, which, for example, is common in whale vocalizations.[11]

Another classification method is automated music genre classification using machine learning techniques, such as Hidden Markov Models and Support Vector Machines[12] with an extracted set of features from the input signal. While the features used in the literature do not assume narrowband signals, they rely on signals with duration in the order of seconds. This is reasonable since it generally applies to music segments used in this field.

Increasingly, short-duration broadband bioacoustic calls are being recognized to exist in nature.[13] While the literature address detection of some repetitive impulsive sounds such as those produced by Sperm whales (*Physeter macrocephalus*),[14] existing detection methods do not adequately address non-repetitive broadband bioacoustic calls.[15] This paper proposes a detection and classification approach for short-duration broadband bioacoustics based on a small set of amplitude-invariant spectro-temporal features (STFs). To develop and test this approach, sound recordings of the spotted tilapia (*Tilapia mariae*), a freshwater fish that is invasive to Australia,[16,17] were used. The presented approach is based on the retrospective selection of small STF vectors that are

[a)] Portions of this work were presented in "Classification of underwater broadband bioacoustics using spectro-temporal features," *Proceedings of the ACM International Conference on Underwater Networks and Systems* (*WUWNet'12*), November 2012.

[b)] Electronic mail: navinda.kottege@csiro.au

[c)] Current address: Australian Institute of Marine Science, PMB3, Townsville MC, Townsville Queensland 4810, Australia.

tailored for short-duration broadband sounds, followed by classification using Discriminant Analysis (DA) and Logistic Regression (LR). The computational performance of this approach for automating broadband bioacoustics detection and classification in the field was also quantified.

II. MATERIALS AND METHODS

A. Study species

Tilapia mariae (Fig. 1) is a freshwater and estuarine teleost native to West African coastal drainages in the Gulf of Guinea and naturalized in the USA, Australia, and possibly Russia[16] due to aquarium and aquaculture releases. In its native range, it can be the dominant fish species in streams, rivers, lakes, and estuaries,[18] and supports local subsistence and artisanal fisheries in some catchments.[19]

Outside its native range, *T. mariae* has a potential detrimental impact on native ichthyofauna, including competition for food[20] and breeding space,[21] thereby affecting the relative abundance of native and endemic species.[17] In Australia, *T. mariae* is a declared noxious fish under the relevant State Fisheries Acts in all states and territories, except Western Australia, and is listed on the National Noxious Fish List.[22] Notwithstanding, it has continued its range expansion in the Wet Tropics region since its first detection around 1980.[17,23] The long-term efficacy of current *T. mariae* management strategies, including the banning of its possession,[24,25] the introduction of the peacock cichlid (*Cichla ocellaris*) in Florida (USA)[26] and public awareness and education in Australia,[25] and electro-fishing are uncertain and have not been quantified.

B. Large-scale bioacoustics monitoring

The growing spread of invasive fish species, such as *T. mariae*, demands monitoring systems that can cover large spatial scales. Invasive monitoring methods such as ultrasonic fish tags are expensive and not scalable. Instead, the focus is on the passive capture of sound through a wireless network of acoustic sensor nodes, which is a more affordable and noninvasive approach. Local processing of captured sounds on each node avoids the energy and bandwidth overheads associated with sending raw audio data continuously over wireless links. In-network acoustic classification must consider the resource limitations that are inherent to battery/solar powered sensor nodes and the suitability of feature sets for the species of interest. To motivate the need for computationally efficient algorithms for automated bioacoustic classification, the audio node is described in the Appendix as the main building block for a large scale acoustic monitoring system.

FIG. 1. (Color online) Courting pair of *T. mariae* used in the case study with the larger specimen (male) on the right and the smaller specimen (female) on the left.

C. Recording setup and procedure

To study the acoustic behavior of *T. mariae*, sound and video recordings were made of five individual fish (range of total length 15 28 cm) in a freshwater aquarium (183 cm L × 60 cm H × 45 cm W) filled with rainwater (temperature: 24 °C). A Reson TC4032 hydrophone (Teledyne RESON A/S, Slangerup, Denmark) was placed in the tank, and associated behavior was recorded via an external video camera (Fig. 2). The hydrophone was sufficiently close to the fish to assume direct signal dominance as opposed to reflected sounds off the tank wall. During recordings, tank filters were switched off to minimize acoustic interference. Consequently, recordings were kept to a maximum of 60 min to minimize buildup of waste material, and filters were switched back on in between recordings.

The hydrophone is a low noise, sea-state zero hydrophone with a 10 dB built-in preamplifier and a receiving sensitivity of -170 dB re 1 V/μPa. It has a linear frequency range of 15 Hz to 40 kHz with ±2 dB.[27] The hydrophone was connected to a laptop computer via a Cakewalk FA-66 device, which is a FireWire audio interface providing additional pre-amplification and analog to digital conversion. The nominal input level for this device was -50 to -10 dBu. Recordings were stored as uncompressed Pulse Code Modulated (PCM) Wave files with a sample rate of 44.1 kHz and a resolution of 24 bits. The PCM Wave files can be down-sampled to match the quality of sound recorded through the audio node.

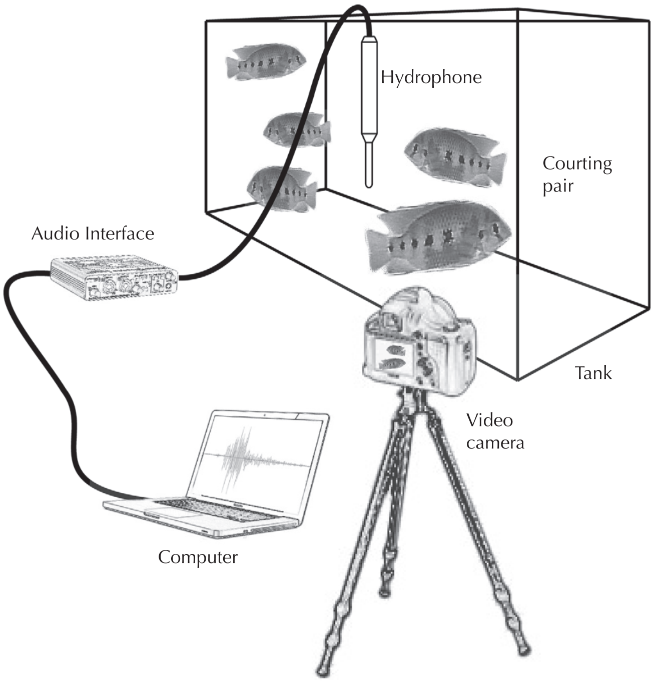


FIG. 2. The experimental setup showing the hydrophone inside the aquarium tank with *T. mariae*.

Sound and video recordings were conducted over a two day period when courting behavior was observed between a pair of fish who mainly occupied one end of the tank. This pair engaged in nest building as evidenced by pebbles on the bottom being moved to this end of the tank. The results and analysis presented in this paper are based on a 60 min recording session.

D. Detection and classification of short-duration broadband sound

The initial recordings revealed a clicking sound, which was confirmed as being produced by *T. mariae* using synchronized video footage. The audio and video tracks were synchronized in post-processing using an event captured on both recordings (a single soft tap on the aquarium glass). The *T. mariae* clicking sound was approximately 10 ms in duration with most of the spectral energy in the 3–8 kHz region with higher harmonics (Fig. 3). Various sounds were present in the recordings which included pebble movement, clicking, scraping, and chewing by the fish, most of which had broadband spectra. The recordings were then manually annotated with ground truth, i.e., the time stamps of acoustic activity (clicking sounds) by the fish were saved as a label track in Audacity.[28] The audio was also processed with a high pass filter with a cutoff frequency f_c of 1000 Hz to remove low frequency noise such as aquarium pumps and aerators of other tanks in the aquarium as well as the 50 Hz mains hum and its harmonics. As with many other bioacoustics applications, it is necessary to detect and segment "events of interest" as a first step before further signal processing is done. For this purpose, an audio detection and segmentation algorithm using an energy detector implemented as a finite state machine was used to extract the various instances of sounds and to discard the background noise.[29] This returned a series of leading-edge aligned sound segments. The manually annotated ground truth time stamps were aligned with the segmented sounds to give each detected sound a corresponding class label to be used for cross validation of the automated classification process later on. From one set of recordings (60 min duration) used for this study, there were 48 manually annotated *T. mariae* clicks, verified with video footage showing synchronous mouth movement.

The next step in the process is to extract features from the segmented sounds. As discussed previously, most standard acoustic features assume longer duration sounds and/or narrowband signals. Given that our classification target is a short duration broadband signal, first a set of nine-dimensional features derived from the music genre classification literature is considered [Timbral Texture Features (TTFs)][30,31] and then a set of six-dimensional features specifically selected for short-duration broadband sounds is considered (STFs).[32]

1. TTFs

TTFs as described in the literature are assembled by initially dividing the signal into relatively small analysis windows with a duration of 23 ms. Multiple analysis windows make up a larger texture window with a duration of 1 s. Characteristics such as spectral centroid, spectral roll-off, spectral flux, and zero-crossings as well as Mel-Frequency

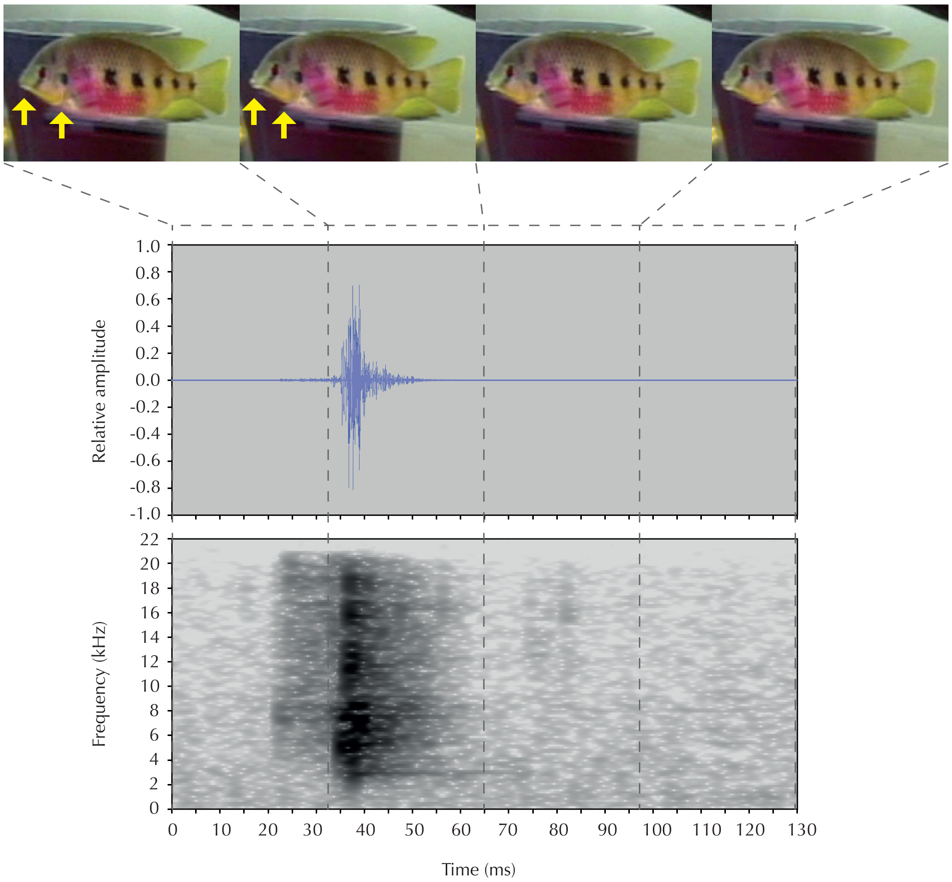


FIG. 3. (Color online) Four video frames (duration ~33 ms per frame) aligned with the corresponding audio signal recorded through the hydrophone showing the mouth and gill movement (arrows in the first two frames) of the fish. The click sound onset corresponds with the sudden opening of the mouth and the mouth remains open beyond the duration of the click.

Cepstral Coefficients (MFCCs), are then calculated for each analysis window. The final feature vector consists of the means and variances of these quantities over a full texture window. Due to the much shorter timeframe of the *T. mariae* sound, the actual values of spectral centroid, spectral roll-off, spectral flux, zero crossings, and the first five coefficients of MFCCs are considered, giving a nine-dimensional feature vector for each sound segment. This contrasts with using the means and variances of these quantities over a much larger time window as done in the literature.[30]

2. STFs

To effectively characterize the short duration broadband *T. mariae* click sound, a unique set of STFs (Ref. 32) are introduced as an alternative to existing sound classification features. Specifically, a number of attributes in the spectral and temporal domain were observed which could be used for this purpose. The spectrogram consisting of power spectral density values (assuming time and frequencies are discretized) is given by

$$P(f,t) = k|S(f,t)|^2, \text{ where } k = 2 \bigg/ \left(f_s \sum_{n=1}^{L} |\omega(n)|^2 \right), \tag{1}$$

where $S(f,t)$ is the short time Fourier transform of time domain signal $s(t)$, $\omega(n)$ is the Hamming windowing function, L is the window length, and f_s is the sampling frequency. The relative sound levels over frequency $F(f)$ and over time $T(t)$ are given by

$$T(t) = 10\log_{10}\sum_{f=0}^{f_s/2} P(f,t),$$

$$F(f) = 10\log_{10}\sum_{t=0}^{t'} P(f,t)\,, \tag{2}$$

where t' is the duration of the relevant sound segment.

Figure 4 shows a set of plots of relative sound level over frequency (F) and time (T) for a typical *T. mariae* click, averaged over (i) all the manually annotated *T. mariae* clicks and (ii) all other manually annotated sounds detected in the detection and segmentation phase. The spectral and temporal characteristics unique to the event of interest (the *T. mariae* click) were used as the basis for selecting the STFs.

As mentioned, spectrogram P is calculated for each segment as given by Eq. (1) and subsequently T and F are derived as given by Eq. (2). Figure 5 shows a typical *T. mariae* click spectrogram along with its frequency distribution and temporal evolution F and T. Maximum values and positions of these maximum values for F and T are then calculated as $c_1 = \max(T(t))$, $a_1 = \mathrm{argmax}(T(t))$, $c_2 = \max(F(f))$, $a_2 = \mathrm{argmax}(F(f))$. The lengths of T and F are b_1 and b_2, respectively. The mean values of T and F are d_1 and d_2, respectively. The set of STFs with six elements denoted by $\{v_1 \cdots v_6\}$ and defined as follows with reference to the above quantities and those shown in Fig. 5:

v_1: The ratio between the peak position of $T(t)$ and the duration of the sound segment t' (i.e., a_1/b_1);
v_2: The ratio between the peak position of $F(f)$ and the frequency bandwidth (i.e., a_2/b_2);

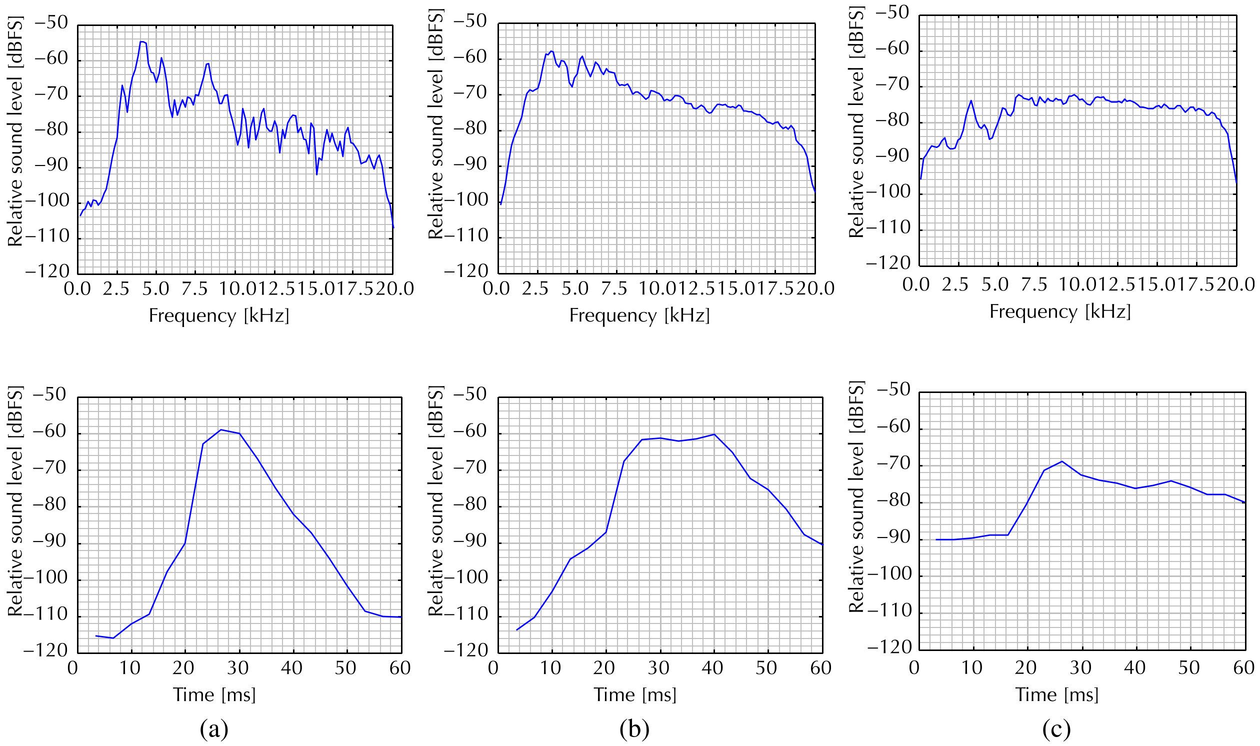


FIG. 4. (Color online) Spectral (first row) and temporal (second row) characteristics of (a) a representative example click, (b) average of 48 detected clicks, and (c) average of 773 other sounds detected in the recordings.

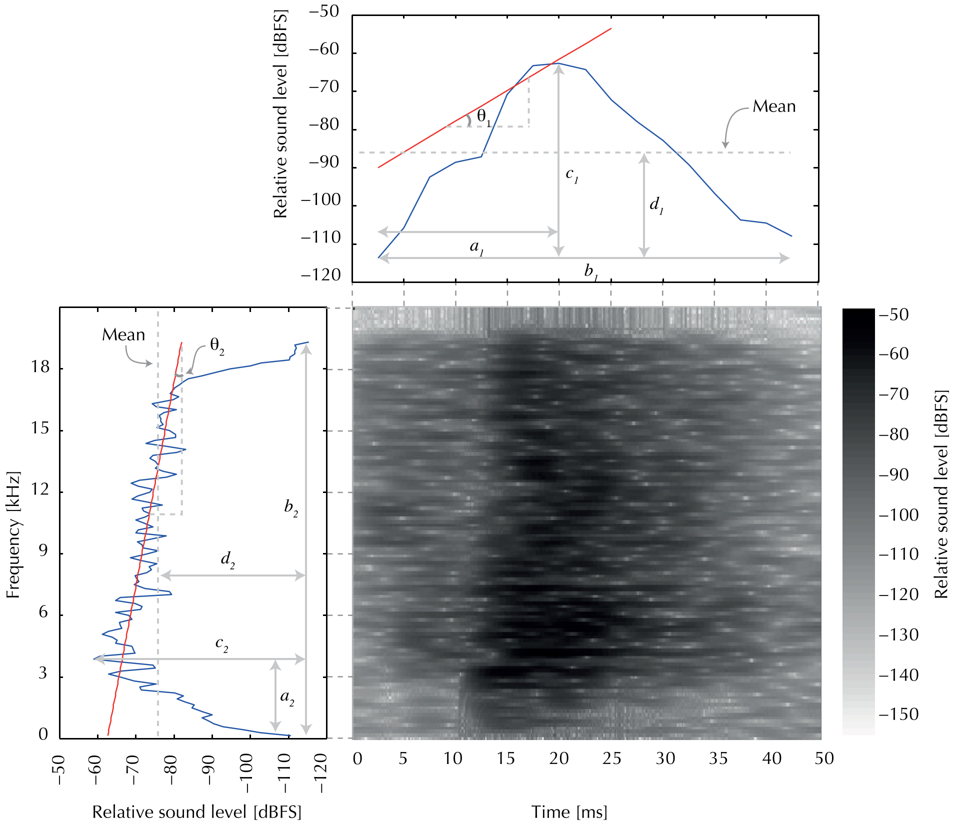


FIG. 5. (Color online) The various annotated measurements shown in the spectral and temporal characteristics of a typical *T. mariae* click are used to define the STFs in the text.

v_3: The gradient of $T(t)$ immediately before the peak [i.e., $\tan(\theta_1)$] implemented as the gradient of the least squares fit lines of the three points immediately before the peak position;
v_4: The gradient of the least squares t line of $F(f)$ [i.e., $\tan(\theta_2)$];
v_5: The ratio between the peak value of $T(t)$ and its mean value (i.e., c_1/d_1);
v_6: The ratio between the peak value of $F(f)$ and its mean value (i.e., c_2/d_2).

These features were selected to be signal strength invariant, which contributes to the robustness of the subsequent classification process. STFs were calculated for each detected sound segment and make up the feature vector associated with that sound. These were used as the input vectors for the automated classification of sound produced by *T. mariae* (Fig. 6).

Two standard classification methods commonly used in machine learning, DA and LR,[33] are used as two independent classification methods for evaluation. Performance of STF and TTF vectors are compared with the same two classification methods and the cross validated results are presented in Sec. III B. STF vectors are also compared against spectrogram correlation, which is a classification technique without explicit feature extraction.[2,10] Apart from ground truth annotation for the offline training process and performance evaluation, the complete process of detection, segmentation, feature extraction, and classification was automated. While the evaluation results presented herein were obtained using offline processing, the system can potentially be used for an online implementation.

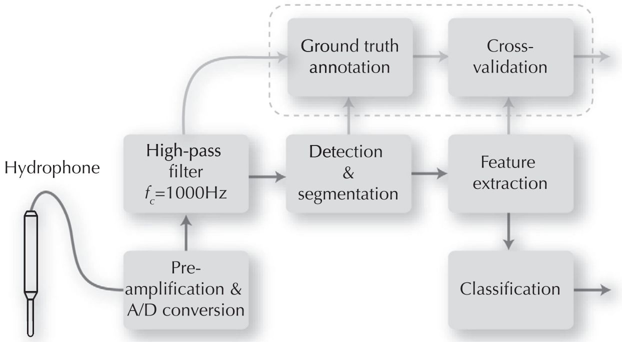


FIG. 6. Functional block diagram of the evaluation process.

3. Classification metrics

To evaluate and compare the performance of the classification methods, Accuracy, Precision, Sensitivity (Recall), and Specificity of each classification method/feature set combination are calculated using k-fold cross-validation. These four metrics are defined as follows:

$$\begin{aligned} \text{Accuracy} &= \frac{tp+tn}{tp+tn+fp+fn}, \\ \text{Precision} &= \frac{tp}{tp+fp}, \\ \text{Sensitivity} &= \frac{tp}{tp+fn}, \\ \text{Specificity} &= \frac{tn}{tn+fp}, \end{aligned} \tag{3}$$

where tp, tn, fp, and fn are the number of true positives, true negatives, false positives, and false negatives, respectively.

TABLE I. Autopsy results of breeding pair of spotted tilapia, *Tilapia mariae*.

Specimen	Total length (mm)	Standard length (mm)	Weight (g)	Gonads
Small fish	212	174	214	Eggs
Large fish	280	225	491	Sperm

To further quantify the classification performance of each feature set, an overall performance metric is defined which is the product of accuracy, precision, sensitivity, and specificity. Since it is desirable for each of these quantities to tend to 1 and since their range is 0–1, the new performance metric would reflect the overall performance of these quantities and remain in the range 0–1.

To evaluate the performance of the features and classification methods at lower sampling rates, the original 44.1 kHz recording was down-sampled to 16 kHz and then to 8 kHz. The detection, segmentation, feature extraction, and classification process was repeated for each of these down-sampled recordings.

III. RESULTS

A. Description of *T. mariae* acoustic behavior

Analysis of the synchronized video revealed that the clicking sounds were emitted by the smaller fish (i.e., female) of the courting pair. Females have been found to produce sounds in 17 cichlid species, but not in *T. mariae*.[34,35] Hence, although differences in morphology exist within breeding pairs of *T. mariae*,[16] the sex of both fish were confirmed by autopsy (Table I).

The behaviors of the two fish comprising the courting pair (Fig. 1) were distinctly different. The smaller fish, i.e., female, displayed increased agonistic behavior toward the three individuals approaching the nest building end of the tank, including charging and chasing them to the other end of the tank. The larger fish, i.e., male, spent most of the time around the nest and displayed relatively less agonistic behavior. Agnostic behaviors such as Lateral display, Tail beating, Carouseling, and Mouth Fighting were also observed.[36] Hatched fry observed a month after the recordings confirmed that the observed pair had indeed been preparing for breeding.

The female produced the clicking sound mostly while the courting pair was at the nest building end of the tank or when the female was near the middle of the tank. It would position itself between the male at the nest and the other three fish at the other end of the tank. Frame by frame analysis of the video shows visible movement of its lower jaw associated with the onset of the click sound, but the jaw remains in that position for a period of up to 500 ms after the click (Fig. 3). Furthermore, there was a very faint characteristic "bubbling" sound trailing the click by about 200 ms and this lasts for up to 1 s after the click (Fig. 7).

The authors are not aware of other studies on cichlids documenting click sounds with similar duration and frequency range, except for a reported wide frequency range (1–16 kHz) for *Tilapia mossambica*.[37] The production mechanism of the observed sound is unclear, but could possibly involve the swim bladder. Even though cichlid species are shown to be sensitive up to 3 kHz (and some species up to 4 kHz),[38] sensitivity studies on *T. mariae* are not reported in the literature. However, in fishes the ability to vocalize is independent of hearing sensitivities.[39]

B. Detection and classification performance

As mentioned before, many different sounds apart from the event of interest (clicking) were present in the recordings, including the sounds of the fish moving and flicking pebbles, scraping against objects, chewing, and interacting with the water surface. The detection and segmentation algorithm found all of the 48 manually annotated clicks as well as most of the other sounds mentioned above. Once both STF and TTF vectors were extracted from the sound segments, they were classified using DA as well as LR and the overall performance was evaluated using *k*-fold cross-validation with $k = 10$ over 100 iterations. The process of detection, segmentation, feature extraction, classification, and cross-validation was initially done with the original 44.1 kHz recording and then repeated for the down-sampled 16 kHz version (64% reduction of sampling rate) and the 8 kHz version (81% of reduction of sampling rate). The total number of sounds detected by the detection and segmentation algorithm decreased with the sampling rate, with 821 sounds being detected at 44.1 kHz, 430 at 16 kHz, and to 236

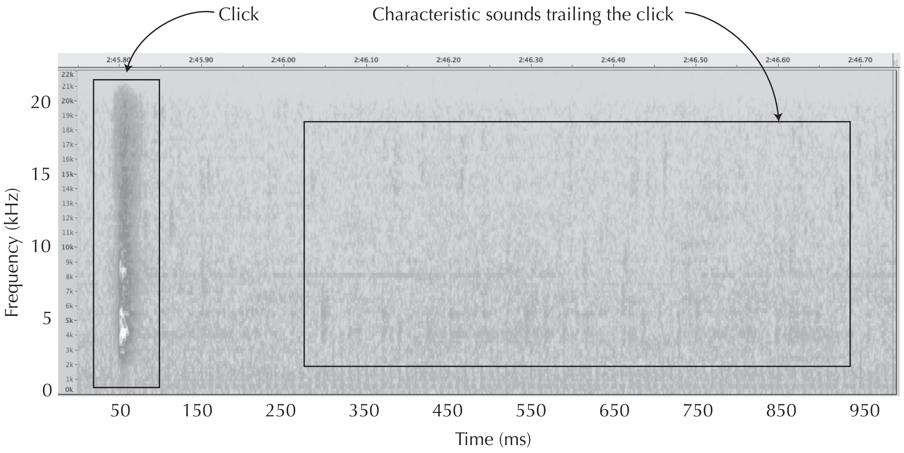


FIG. 7. Spectral and temporal characteristics of a typical *T. mariae* click.

TABLE II. Performance of DA and LR with the proposed STFs and TTFs along with spectrogram correlation at sampling rates of 44.1, 16, and 8 kHz.

f_s (kHz)	Features	Method	Accuracy	Precision	Sensitivity	Specificity
44.1	STF	DA	0.97	0.67	0.94	0.97
		LR	0.98	0.84	0.75	0.99
	TTF	DA	0.93	0.45	0.90	0.93
		LR	0.95	0.63	0.24	0.99
	Spect. corr.		0.85	0.00	0.00	0.90
16	STF	DA	0.95	0.69	0.95	0.95
		LR	0.96	0.85	0.78	0.98
	TTF	DA	0.89	0.52	0.27	0.97
		LR	0.90	0.79	0.13	0.99
	Spect. corr.		0.85	0.25	0.19	0.93
8	STF	DA	0.85	0.58	0.95	0.83
		LR	0.93	0.83	0.82	0.96
	TTF	DA	0.86	0.76	0.43	0.96
		LR	0.87	0.76	0.56	0.95

at 8 kHz. Despite this reduction, the 48 manually annotated clicks always appeared in the detected segments.

For the higher sampling rate of 44.1 kHz, STF outperformed TTF for accuracy, precision, specificity by 4%, and sensitivity by 48% when using DA. Going from 44.1 to 16 kHz, the average accuracies for STFs dropped from 0.98 to 0.96 and for TTFs from 0.94 to 0.90. With down-sampling to 16 kHz, precision of STFs increased by 1.5% and accuracy decreased by 2%. For TTFs, precision improved by 15% and accuracy decreased by 5%. However, sensitivity decreased drastically by 70% for DA and 46% for LR. At 8 kHz, the average accuracies of STFs and TTFs dropped to 0.89 and 0.86, respectively, which is approximately a 9% drop compared to the results at 44.1 kHz (Table II).

Despite an accuracy of 0.85, spectrogram correlation has very low utility due to its low precision and sensitivity. This stems from the fact that the correlation kernel used in spectrogram correlation relies on a relatively long duration, slow varying, narrowband signal such as the frequency sweeps common in whale calls.[1] The short-duration impulsive broadband click of *T. mariae* does not conform to these characteristics yielding poor performance when using spectrogram correlation. Due to this, spectrogram correlation was not evaluated for the 8 kHz down-sampled recording.

Receiver Operator Characteristic (ROC) curves for 44.1 and 16 kHz demonstrate the robustness of STFs compared to TTFs as the sampling rate is reduced to 16 kHz (Fig. 8). These curves also depict the poor performance of spectrogram correlation both for 44.1 and 16 kHz sampling. Overall, for both classification methods STFs show better performance compared to TTFs.

1. Effects of reduced sampling rate

To observe the effect of reduced sampling rate on each of the STF components, the density distributions of features associated with (i) detected *T. mariae* clicks and (ii) all detected sounds (full data set) for both 44.1 and 16 kHz sampling rates are plotted (Fig. 9). *T. mariae* clicks have their peak energy in frequencies below 8 kHz whereas the other sounds present in the recordings have their energy spread into the higher frequencies. At 16 kHz sampling, the peak

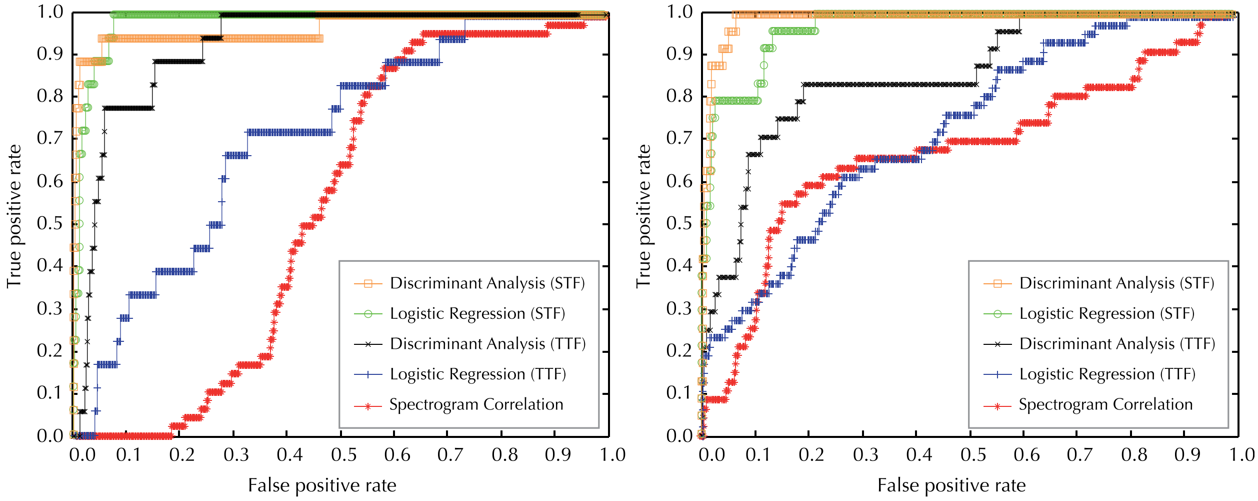
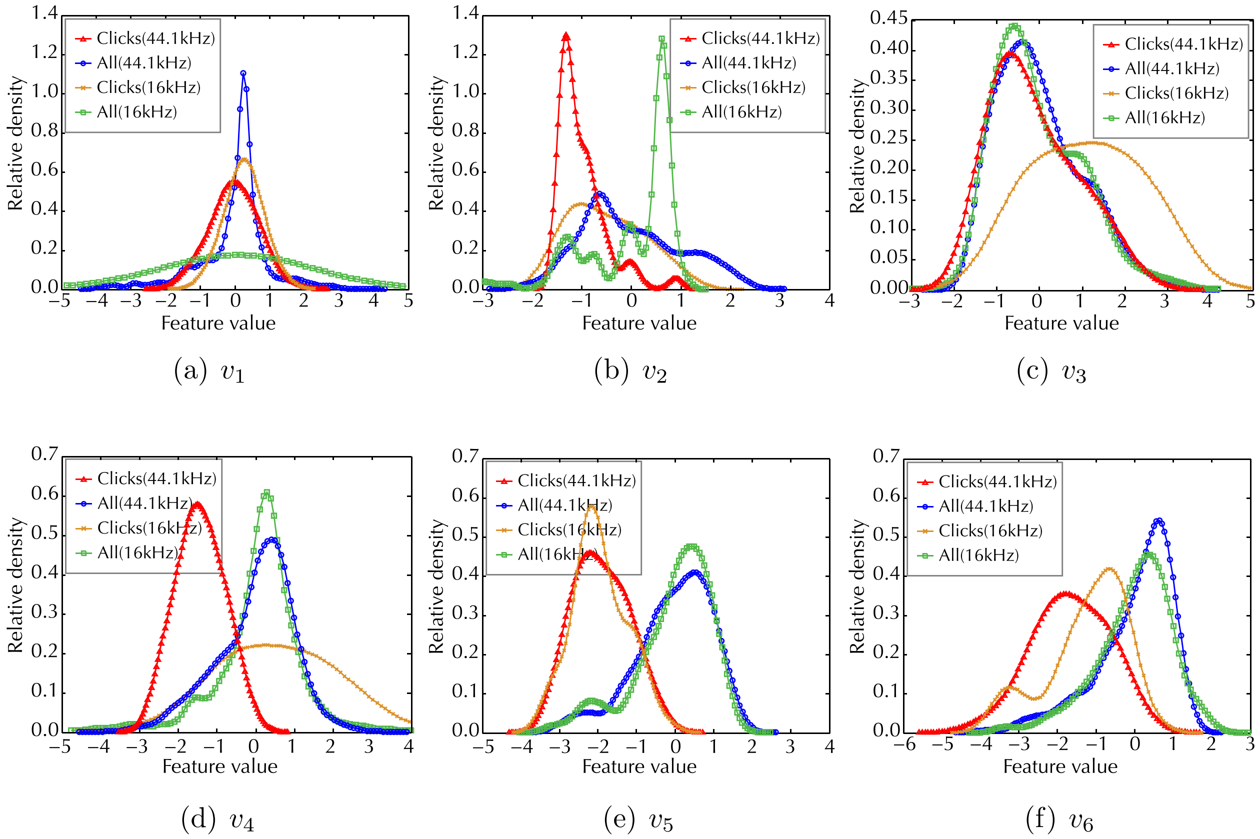


FIG. 8. (Color online) ROC curves at $f_s = 44.1$ kHz (left) and $f_s = 16$ kHz (right), for DA and LR with the proposed STFs and TTFs along with spectrogram correlation.



(a) v_1

(b) v_2

(c) v_3

(d) v_4

(e) v_5

(f) v_6

FIG. 9. (Color online) Density distribution of the six features $\{v_1 \cdots v_6\}$, comparing the positive examples (triangles and crosses) with the full data set (circles and squares) for the original 44.1 kHz recording and the down-sampled 16 kHz version.

energy portion of the *T. mariae* click is still preserved while the other sounds lose the more energetic part of their spectra. This explains why feature v_4 loses its separation from the full data set [Fig. 9(d)] when down-sampled to 16 kHz as this feature is based on the shape of the frequency distribution over the full bandwidth. However, features v_2 and v_3 have increased separation between clicks and the full dataset at the lower sampling rate.

Due to other sounds losing most of their energetic portion of the spectrum when down-sampled, the total number of sounds picked up by the detection and segmentation algorithm decreases with down-sampling from 821 at 44.1 kHz to 430 at 16 kHz and 236 at 8 kHz (Fig. 10). For example, at 16 kHz, which is the sampling rate used on the audio node, the number of samples to be processed is 70% lower than at 44.1 kHz. Despite this substantial reduction, the 48 manually annotated clicks always appear in the detected segments. Moreover, the classification performance for STFs comes with an actual increase in precision and sensitivity and only a small decrease in accuracy (2%). This is reflected as an overall increase in the performance metric (0.58 to 0.60). In contrast, the performance for TTFs is nearly halved, highlighting the robustness of STFs to lower sampling rates.

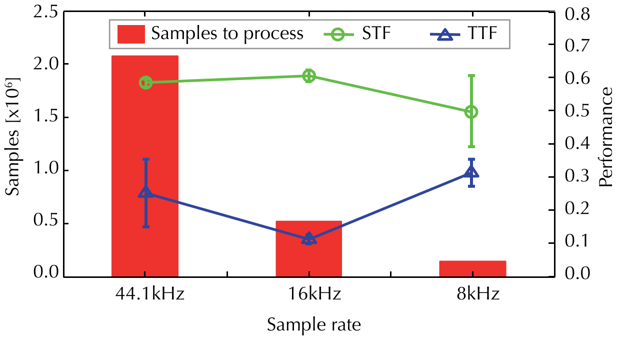


FIG. 10. (Color online) Reduction in the number of samples to be processed during sound classification with the decreasing sampling rate. The error bars represent the maxima and minima for the two different classification algorithms DA and LR.

Therefore, it can be concluded that a sampling rate of 16 kHz strikes a much better balance between the number of samples to be processed and classification performance of STFs.

IV. DISCUSSION

This paper presented a novel set of signal strength invariant STFs which effectively characterize short duration broadband bio-acoustic calls. These features can then be used with standard machine learning techniques to accurately and efficiently detect and classify species such as *T. mariae*. When combined with the audio detection and segmentation algorithm used in this work, it was shown that an accuracy >0.95 can be achieved while the number of samples processed can be reduced by 70% when the signal is down-sampled. It was demonstrated that the proposed features remain robust even at the lower sampling rate of 16 kHz. This along with the relatively small feature set size makes it ideally suited for implementation on the wireless audio node platform as described in supplemental material to this paper.[40] These audio nodes could potentially serve as the building blocks of a large scale freshwater monitoring system deployed along inland waterways and lakes enabling real-time in-network detection and classification of freshwater species.

It was further demonstrated that the automatic classification of the distinct clicking sound through DA and LR performs with high accuracy and specificity, but with lower precision and sensitivity. DA emerges as the better performing method for automatic classification of vocalizations by this particular species.

This study is the first to confirm vocalization by a female cichlid in a breeding pair of *T. mariae*. In addition, to the best of our knowledge, these types of click sounds have not been reported in the literature for fresh water fish. The observation of female vocalization expand on previous studies on other cichlids, of which females have been found to produce sound in 17 species.[34,35] In *T. mariae*, males have been documented to vocalize during aggressive male-male interaction.[34,35] The exact role and general occurrence of female vocalization in this and other cichlid species, as well as the mechanism producing the sound, are unclear and require further research. Furthermore, the occurrence of similar clicking sounds in other species is worthy of investigation.

The ability to detect and to automatically classify *T. mariae* through the passive capture of sound enables not only the detection of presence of a species, but also the possibility of population control. Through integration with wireless sensor network technology, detection of this invasive species can be automated over large spatial scales to cover freshwater bodies such as rivers and lakes. This in itself could enable water management specialists to implement control measures in the specific regions where the species has been detected. Population control can be enhanced through the automation of active measures as well. For instance, the click sound can be played back synthetically in areas suspected to be on the invasive front in order to lure *T. mariae* to a spot where they can be easily captured. Alternatively, synthetic sound can be used to disrupt mating calls.

While the current study has demonstrated effective sound classification for species detection in a controlled environment, the authors acknowledge that species detection in an uncontrolled environment could be more difficult. The small size of the test tank introduced reverberations which in turn manifest themselves as interfering signals. A natural environment (e.g., Riparian environment) would not typically have these strong reverberations but would present other interfering noise sources such as water flow induced hydrodynamic sounds. However, the short duration broadband nature of the detected clicking sound makes it robust to most interference, noise, and frequency selective fading common in the underwater medium.[41] It can be anticipated that the aforementioned flow noise and dispersion due to suspended solids, may affect acoustic propagation and therefore limit the detection range and classification quality in natural freshwater environments. These are issues that need to be examined in experimental field work in order to design an effective long-term bioacoustics monitoring program in such environments.

ACKNOWLEDGMENTS

Funding for this project was provided by the CSIRO Water for a Healthy Country Flagship Program.

[1] D. K. Mellinger, "A comparison of methods for detecting right whale calls," Can. Acoust. **55**, 55–65 (2004).

[2] M. F. Baumgartner and S. E. Mussoline, "A generalized baleen whale call detection and classification system," J. Acoust. Soc. Am. **129**, 2889–2902 (2011).

[3] T. S. Brandes, "Automated sound recording and analysis techniques for bird surveys and conservation," Bird Conserv. Int. **18**, S163–S173 (2008).

[4] S. Parsons and G. Jones, "Acoustic identification of twelve species of echolocating bat by discriminant function analysis and artificial neural networks," J. Exp. Biol. **203**, 2641–2656 (2000).

[5] W. Hu, N. Bulusu, C. T. Chou, S. Jha, A. Taylor, and V. N. Tran, "Design and evaluation of a hybrid sensor network for cane toad monitoring," ACM Trans. Sensor Networks **5**, 4:1–4:28 (2009).

[6] A. Taylor, G. Watson, G. Grigg, and H. McCallum, "Monitoring frog communities: An application of machine learning," in *Proceedings of the Eighth Annual Conference on Innovative Applications of Artificial Intelligence* (AAAI Press, 1996), pp. 1564–1569.

[7] K. S. Boyle and T. C. Tricas, "Pulse sound generation, anterior swim bladder buckling and associated muscle activity in the pyramid butterflyfish, *Hemitaurichthys polylepis*," J. Exp. Biol. **213**, 3881–3893 (2010).

[8] N. Longrie, S. Van Wassenbergh, P. Vandewalle, Q. Mauguit, and E. Parmentier, "Potential mechanism of sound production in *Oreochromis niloticus* (Cichlidae)," J. Exp. Biol. **212**, 3395–3402 (2009).

[9] K. P. Maruska, U. S. Ung, and R. D. Fernald, "The African cichlid fish *Astatotilapia burtoni* uses acoustic communication for reproduction: Sound production, hearing, and behavioral significance," PLoS One **7**(5), e37612 (2012).

[10] D. K. Mellinger and C. W. Clark, "Recognizing transient low-frequency whale sounds by spectrogram correlation," J. Acoust. Soc. Am. **107**, 3518–3529 (2000).

[11] M. F. Baumgartner, S. M. V. Parijs, F. W. Wenzel, C. J. Tremblay, H. C. Esch, and A. M. Warde, "Low frequency vocalizations attributed to Sei whales (*balaenoptera borealis*)," J. Acoust. Soc. Am. **124**, 1339–1349 (2008).

[12] T. Pohle, E. Pampalk, and G. Widmer, "Evaluation of frequently used audio features for classification of music into perceptual categories," in *Proceedings of the Fourth International Workshop on Content-Based Multimedia Indexing* (*CBMI'05*) (2005).

[13] I. A. Martins and J. Jim, "Bioacoustic analysis of advertisement call in *Hyla nana* and *Hyla sanborni* (Anura, Hylidae) in Botucatu, São Paulo, Brazil," Braz. J. Biol. **63**, 507–516 (2003).

[14] V. Kandia and Y. Stylianou, "Detection of sperm whale clicks based on the Teager? Kaiser energy operator," Appl. Acoust. **67**, 1144–1163 (2006).

[15] T. S. Brandes, "Feature vector selection and use with hidden Markov models to identify frequency modulated bioacoustic signals amidst noise," IEEE Trans. Audio, Speech, Lang. Proc. **16**, 1173–1180 (2008).

[16] M. Bradford, F. J. Kroon, and D. J. Russell, "The biology and management of *Tilapia mariae* (Pisces: Cichlidae) as a native and invasive species: A review," Mar. Freshwater Res. **62**, 902–917 (2011).

[17] F. J. Kroon, D. J. Russell, P. A. Thuesen, T. Lawson, and A. Hogan, "Using environmental variables to predict distribution and abundance of invasive fish in the wet tropics," Technical Report, CSIRO (2011).

[18] R. C. Akpaniteaku and J. N. Aguigwo, "Seasonal variation in catch of tilapiines (Osteichthyes: Chichlidae) in Agulu and Nawfia Lakes, Anambra State, Nigeria," J. Sustain. Trop. Agr. Res. **6**, 19–22 (2003).

[19] C. S. Nwadiaro, "Fecundity of cichlid fishes of the Sombreiro River in the lower Niger delta," Rev. Zool. Africaine **101**, 433–437 (1987).

[20] J. Courtenay, R. Walter, and J. E. Deacon, "Fish introductions in the American Southwest: A case history of Rogers Spring, Nevada," Southwest. Natur. **28**, 221–224 (1983).

[21] W. R. Brooks and R. C. Jordan, "Enhanced interspecific territoriality and the invasion success of the spotted tilapia (*Tilapia mariae*) in South Florida," Biol. Invasions **12**, 865–874 (2010).

[22] Bureau of Rural Sciences, "A strategic approach to the management of ornamental fish in Australia," Technical Report, Bureau of Rural Sciences (2007).

[23] A. C. Webb, "Status of non-native freshwater fishes in tropical northern Queensland, including establishment success, rates of spread, range and introduction pathways," J. Proc. R. Soc. N. S. W. **140**, 63–78 (2007).

[24] M. R. Clark, "Probable establishment and range extension of the spotted tilapia, *tilapia mariae boulenger* (Pisces: Cichlidae) in East Central Florida," Florida Sci. **44**, 168–171 (1981).

[25] DEEDI, "Control of exotic pest fishes: An operational strategy for Queensland freshwaters 2011–2016," Technical Report, Queensland Department of Employment, Economic Development and Innovation (2011).

[26] P. L. Shafland, "Introduction and establishment of a successful butterfly peacock fishery in southeast Florida canals," in *Proceedings of the American Fisheries Society Symposium* (1995), pp. 443–451.

[27] Reson, "TC4032 Datasheet" (2005), http://www.reson.com/products/hydrophones/tc-4032/ (Last viewed June 1, 2014).

[28] Audacity Development Team, "Audacity [Computer program]" (2010), http://audacity.sourceforge.net/ (Last viewed June 1, 2014).

[29] B. Croker and N. Kottege, "Using feature vectors to detect frog calls in wireless sensor networks," J. Acoust. Soc. Am. **131**, EL400–EL405 (2012).

[30] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," IEEE Trans. Speech Audio Proc. **10**, 293–302 (2002).

[31] T. Giannakopoulos, D. Kosmopoulos, A. Aristidou, and S. Theodoridis, "Violence content classification using audio features," in *Advances in Artificial Intelligence*, *Vol. 3955 of Lecture Notes in Computer Science*, edited by G. Antoniou, G. Potamias, C. Spyropoulos, and D. Plexousakis (Springer, Berlin Heidelberg, 2006), pp. 502–507.

[32] N. Kottege, F. Kroon, R. Jurdak, and D. Jones, "Classification of underwater broadband bio-acoustics using spectro-temporal features," in *Proceedings of the Seventh ACM International Conference on Underwater Networks and Systems* (*ACM*) (2012).

[33] S. J. Press and S. Wilson, "Choosing between logistic regression and discriminant analysis," J. Am. Stat. Assoc. **73**, 699–705 (1978).

[34] P. Lobel, "Possible species specific courtship sounds by two sympatric cichlid fishes in Lake Malawi, Africa," Env. Biol. Fishes **52**, 443–452 (1998).

[35] N. Longrie, P. Poncin, M. Denoel, V. Gennotte, J. Delcourt, and E. Parmentier, "Behaviors associated with acoustic communication in Nile Tilapia (*Oreochromis niloticus*)," PLoS One **8**, e61467 (2013).

[36] R. F. Oliveira and V. C. Almada, "Mating tactics and male-male courtship in the lek-breeding cichlid *Oreochromis mossambicus*," J. Fish Biol. **52**, 1115–1129 (1998).

[37] W. J. R. Lanzing, "Sound production in the cichlid *Tilapia mossambica* Peters," J. Fish Biol. **6**, 341–347 (1974).

[38] T. Schulz-Mirbach, B. Metscher, and F. Ladich, "Relationship between swim bladder morphology and hearing abilities—A case study on Asian and African Cichlids," PLoS One **7**(8), e42292 (2012).

[39] F. Ladich and T. Schulz-Mirbach, "Hearing in cichlid fishes under noise conditions," PLoS One **8**(2), e57588 (2013).

[40] See supplementary material at http://dx.doi.org/10.1121/1.4919298 for CSIRO Audio Node.

[41] N. Kottege and U. R. Zimmer, "Underwater acoustic localization for small submersibles," J. Field Robotics **28**, 40–69 (2011).