Applied Acoustics 110 (2016) 248–255

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Identification of fish vocalizations from ocean acoustic data

Farook Sattar [a,*], Sarika Cullis-Suzuki [b], Feng Jin [c]

[a] Department of Electrical & Computer Engineering, University of Waterloo, Waterloo, Ontario, Canada
[b] Ocean Networks Canada, University of Victoria, Victoria, British Columbia, Canada
[c] Department of Electrical & Computer Engineering, Ryerson University, Toronto, Ontario, Canada

ARTICLE INFO

Article history:
Received 26 July 2015
Received in revised form 29 December 2015
Accepted 22 March 2016

Keywords:
Ocean acoustics
Fish vocalizations
Identification
High-resolution
Descriptors

ABSTRACT

A new method for identification of fish vocalizations based on auditory analysis and support vector machine (SVM) classification is presented. In this method, high resolution features have been extracted from fish vocalization data using the amplitude modulation spectrogram (AMS) of the input signals to facilitate the identification of grunts and growls made by a highly vocal wild fish, *Porichthys notatus*. The comparison results made from ocean audio recordings verify the effectiveness of the proposed method in identifying various types of fish vocalizations. The relationships between signal-to-noise ratio (SNR) and ocean temperature with the accuracy of the proposed method have also been quantified. Moreover, a context-aware prediction algorithm is introduced for estimating the continuous data.

1. Introduction

Acoustic communication is an important component of intra and inter-specific interactions among many species of fish [1]. Fish produce sound in agonistic situations [2–4], courtship and reproduction events [5–8], and unintentionally during other behaviours [9]. These sounds can range from barely audible to the human ear [10] to loud enough to disturb the sleep of nearby residents [11]. To date, over 800 species are known to make sound and many more are believed to do so [12,13].

Passive acoustics allows a non-destructive way to gain insights on spawning locations, fish abundance, and temporal aspects [12,14]. However, it also relies on the basic recognition of fish sounds, the majority of which to date, have not yet been identified [12]. What's more, once sounds have been identified, sifting through extensive audio datasets can manually become a long and tedious process [15,16]. Manual detections can be too time consuming and error-prone (e.g., due to bias or observer fatigue) to yield accurate results over long datasets [17]. Applying machine learning and automated approaches to long acoustic datasets therefore would further this field markedly.

Here we focus on the plainfin midshipman (*Porichthys notatus*), a highly vocal species of toadfish found along the northeast Pacific [18]. This fish makes four distinct vocalizations: the hum, growl, grunt, and grunt train [5]. Grunts and growls are used in antagonistic encounters with conspecifies, while the hum is produced during reproductive months by alpha males trying to attract females to mate [5,19]. Compared with other species, these fish are fairly well understood, and their call characteristics, well documented [20]. However, an automated approach to quantify and identify their sounds in natural habitats and over long time frames has never been created. Such a tool could offer ecological insights on *P. notatus* populations including abundance, habitat location and range, migratory patterns and call diversity *in situ*.

Traditionally, the identification of animal vocalizations has been done by manually analyzing large recorded datasets [16]. But machine-based algorithms offer a more efficient and potentially effective way to filter through long term acoustic data sets [21–23]. For example, in [15], an identification scheme has been presented for different Orthoptera species by using temporal information such as duration between zero-crossings, shape of the waveform and artificial neural network based multilayer perceptron classifier. Similarly, a complicated method for identification of humpback whales has been introduced in [24] by detecting frequency contour and optimizing multiple parameters. Other identification approaches have been proposed based on frequency-domain information, such as the spectrogram correlation based template matching scheme [16], Kalman filter based contour-tracking scheme [25], contour features based scheme [26], and contour signature based scheme [27]. However, in our study, as the characterizing features of *P. notatus*' grunt and growl signals both fall in the lower frequency range (≈100 Hz), and both sounds are of very short duration, a higher resolution temporal–spectral

* Corresponding author.
E-mail address: farook_sattar@yahoo.com.sg (F. Sattar).

http://dx.doi.org/10.1016/j.apacoust.2016.03.025
0003-682X/

signal representation is therefore desired for accurate identification. In many species of toadfish including *P. notatus*, call frequency is correlated with water temperature [28–31]. Therefore, knowing water temperature can help to predict dominant call frequencies and can thus become a useful parameter in automatic auditory identification schemes.

In this paper, we propose a fish sound identification scheme based on auditory analysis using amplitude modulation spectrogram (AMS). The information containing amplitude modulations of the input signal is analyzed and represented in two-dimensional AMS. The extraction of high-resolution features is performed which is motivated by the results from a neurophysiological experiment on periodicity coding in the auditory cortex [32]. A support vector machine (SVM) classifier is then trained on a large number of pre-selected AMS patterns, and classifies the input signals into grunt and growl classes. It is worth mentioning that the high-resolution features extract the subtle and detailed information and contain more distinctive information than low-resolution features.

2. Method

The proposed identification scheme for fish vocalizations is based on auditory analysis for feature extraction followed by a machine learning algorithm for classification. The overall flowchart of our method is shown in Fig. 1. The hydrophone recordings of fish data are partitioned first into blocks of particular segments. Each 1D data block is then converted into a 2D feature map. A high-resolution feature set (descriptors) is then constructed from the feature maps and used as input to the SVM classifier.

2.1. Data preprocessing

For each of the five days that were analyzed here, five minutes of each hour of a 24-h cycle were processed manually by identifying grunts and growls, thus forming 24 five-minute clips per day. Each five minute spectrogram was then examined manually (visually and audibly) using Audacity 2.0.6 [33] by an expert, who recorded all start and end time stamps (in seconds) for each vocalization. Based on the time stamps of the annotated data, all grunt and growl segments were then extracted, resampled (from 44,100 Hz to 16,000 Hz) and resized into N-sample data blocks ($N = 8192$ here, referring to ≈ 0.5 s) followed by time windowing using N-sample Hamming windows [34]. It should be noted that resampling is usually done to reduce the computational complexity of the method, by running it on a signal sampled at a lower rate. Resizing, meanwhile, is done to save memory by compressing the signal without changing its spectral content [35].

2.2. Feature extraction

We have proposed a high-resolution descriptor (i.e., feature set) for the identification of fish species from their vocalizations. Each input fish data block is first bandpass filtered into 25 subbands by a mel-frequency bank [36]. The envelope of each subband is then obtained by using full-wave rectification followed by decimation with a factor of 3. The decimated envelope signals are subsequently partitioned into segments of 128 samples (0.572 ms) using 50% overlapping, Hamming window. The 256-point fast Fourier transform (FFT) of the zero-padded segments is then calculated. The FFT computes the modulation spectrum in each subband with a frequency resolution of 15.6 Hz. For each subband, the FFT magnitudes are multiplied by 15 uniformly spaced triangular-shaped windows across the 15.6–400 Hz range and summed up to generate 15 modulation spectrum amplitudes representing AMS feature

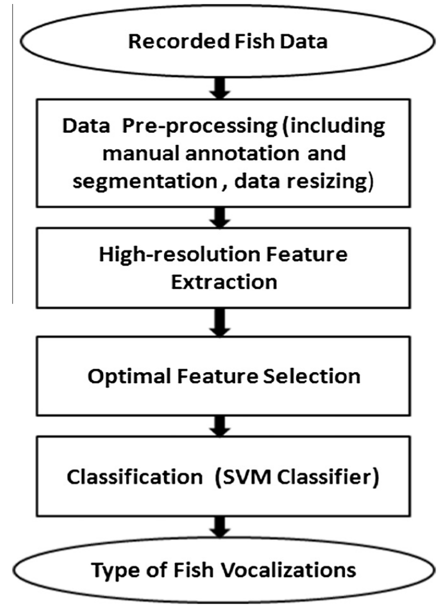

Fig. 1. The overall flowchart of the proposed scheme.

matrix $S_l(n,m)$, where $n,\ m$ and l indicate the time index, modulation index, and subband/channel index, respectively, with $1 \leqslant \{n,m,l\} \leqslant \{N,M,L\}$. Then, as shown in Fig. 2, the proposed high-resolution descriptor, $\mathbf{d}$ of size $(1 \times ML)$ is constructed as follows:

$$\mathbf{d} = [\Lambda_1(m), \Lambda_2(m), \ldots, \Lambda_L(m)]; \tag{1}$$

where

$$\Lambda_l(m) = \frac{1}{N}\sum_{n=1}^{N} S_l(n,m) \tag{2}$$

Here, we set $M = 25$ and $L = 15$.

Illustrative plots of our high-resolution descriptors for grunt and growl vocalizations are shown in Fig. 3.

2.3. Feature selection

Feature selection is adopted here to improve classification by removing redundant information in high-dimensionality spaces. The sequential floating forward selection (SFFS) algorithm [37,38], finds an optimum subset of features by appending features to and discarding features from subsets of selected features and has been adopted to guide the search, as the SFFS algorithm shows below. A separation index based on distance and separability measures is considered in the SFFS algorithm as an objective function, which evaluates the candidate set by returning a measure of their 'goodness'. This SFFS scheme automatically selects the best feature subset of high-resolution features related to fish vocalizations. The size of the feature space is 375, which corresponds to the length of the high-resolution features.

The SFFS algorithm adopted for feature selection:

1. Start with initialization:
   $i \leftarrow 0$;
   $\mathbf{D}_0 \leftarrow \{\emptyset\}$;
   $J(0) \leftarrow 0$
2. Inclusion – select the most significant feature with respect to $\mathbf{D}_k$:
   $\mathbf{d}' = arg\ \max_{\mathbf{d} \notin \mathbf{D}_k} J(\mathbf{D}_k + \mathbf{d})$;
   $\mathbf{D}_{k+1} = \mathbf{D}_k + \mathbf{d}';\quad k = k+1$