Ecological Indicators 84 (2018) 479–487

Contents lists available at ScienceDirect

Ecological Indicators

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

Research paper

A hidden Markov model approach to indicate Bryde's whale acoustics

R.L. Putland[a,*], L. Ranjard[b], R. Constantine[c], C.A. Radford[a]

[a] Leigh Marine Laboratory, Institute of Marine Science, University of Auckland, PO Box 349, Warkworth, 0941, New Zealand
[b] Research School of Biology, ANU College of Medicine, Biology and Environment, The Australian National University, Canberra, ACT0200, Australia
[c] School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland, 1142, New Zealand

A R T I C L E I N F O

Keywords:
Acoustic detection
Cetacean
Hidden markov models
Noise mitigation
Passive acoustic monitoring
Temporal variation

A B S T R A C T

Increasing sound in the ocean from human activity potentially threatens marine animals that use sound to communicate, detect prey, avoid predators and function within their ecosystem. The detection and classification of sound produced by marine animals, such as whales and fish, is an important component in noise mitigation strategies, while also providing valuable insights into their ecology. Traditionally, visual surveys are conducted to assess how these animals utilize a specific area, often underestimating the number of individuals as they don't spend much time at the surface. Long-term passive acoustic monitoring efforts have become more prevalent to monitor such animals. The large datasets collected can be impractical to manually process, necessitating the development of automated detection methods, which often produce mixed results owing to the broad frequency range and variable duration of many biological sounds. Here we describe a novel approach for automated detection of underwater biophonic sounds employing hidden Markov models (HMM). Acoustic data was collected at a single listening station in Hauraki Gulf, from October 2014 to April 2016. HMM detection models were developed for Bryde's whales (*Balaenoptera edeni*) that were used as a model organism because they are notoriously hard to study with traditional visual surveys and produce a characteristic call. Bryde's whale calls also directly overlap the sounds of anthropogenic activity, in particular the sound of vessels transiting to the busiest port in New Zealand; therefore monitoring whale calls is of utmost importance when confronting increasing sound in the ocean. Vocalizations were detected with a sensitivity of 77% and false positive rate of 23%. Bryde's whale vocalizations were detected on 11% of all recordings. Overall, there were significantly more detections during summer (n = 1716) than winter (n = 447), and significantly more during the day (n = 1991) compared to night (n = 1264). This study shows the feasibility of using HMMs on long-term acoustic datasets. The method has the potential to be used for a wide range of soniferous animals who, like the Bryde's whale, also produce unique sounds. The detection method would be particularly useful for mitigation and management strategies of species that are difficult to detect using traditional visual methods.

1. Introduction

Increasing levels of anthropogenic activity, such as shipping, has changed the acoustic soundscape of many marine and terrestrial environments (Merchant et al., 2016; Miksis-Olds and Nichols, 2016). Ecologists and managers alike are concerned that the increasing sound levels in the world's oceans may be threatening marine animals that use sound for everyday subsistence (Erbe, 2012; Hawkins and Popper, 2017). However, one of the most difficult problems faced by those charged with ecosystem management is decision making in the absence of essential information on the spatial and temporal patterns of species of concern. Passive acoustic monitoring (PAM) provides an ecological indicator for how soniferous animals use their natural habitat as well as how they may react to anthropogenic sound (Dunlop, 2016; Parks et al., 2016; Stafford et al., 2017). PAM can be used in remote areas, during day and night and in adverse weather conditions when other detection methods are not possible (Mellinger et al., 2007a,b; Russo and Voigt, 2016). In recent years, autonomous hydrophones have increasingly been deployed in regions of interest to record continuously for months to years at a time (Radford et al., 2010; Miksis-Olds et al., 2013; Nedelec et al., 2015). The large acoustic datasets collected prove almost impossible to analyze with human operators (Aide et al., 2013), therefore automated detection algorithms have become essential when monitoring sounds over large spatial and temporal scales. Using automated methods, datasets can be processed relatively quickly and consistently with human bias removed as well as sound beyond the human hearing range identified with ease (Brown and Smaragdis, 2009).

There are many existing methods for automated detection of

* Corresponding author.
E-mail address: rput037@aucklanduni.ac.nz (R.L. Putland).

http://dx.doi.org/10.1016/j.ecolind.2017.09.025
Received 26 June 2017; Received in revised form 8 September 2017; Accepted 13 September 2017
Available online 19 September 2017
1470-160X/

bioacoustic signals, including energy summation, matched filtering, spectrogram correlation and dynamic time warping (Mellinger, 2004; Munger et al., 2005). The performance of each often depends on the characteristics of a particular species acoustic repertoire and behavior, and the physical environment in which the sounds were recorded (Munger et al., 2005; Russo and Voigt, 2016). Spectrogram correlation is the most commonly used method of automatic detection (Mellinger et al., 2007a,b) where a spectrogram is directly correlated with template vocalizations. Spectrogram cross correlation was very effective in identifying North Pacific right whale (*Eubalena japonica*) calls within two years of acoustic monitoring in the Gulf of Alaska (Munger et al., 2005). However, biological sounds often have a complex nature with variable frequency range and duration, and the cross correlation approach is unable to adapt to changes in duration and alignment.

Hidden Markov models (HMMs) are widely adopted in speech recognition tasks because their flexibility allows sounds to be classified from a series of given observations (Ren et al., 2009). A HMM is defined as a statistical-state machine where each state represents a stationary spectral configuration, and transitions between states represent spectral changes over time (Ren et al., 2009). HMM approaches have been successfully used to automatically detect birds (Somervuo et al., 2006; Ranjard et al., 2016), fish (Vieira et al., 2015) and mammals (Scheifele et al., 2015), often outperforming other detection methods (Weisburn et al., 1993; Kogan and Margoliash, 1998; Brown and Smaragdis, 2009), because they account for the change in spectral characteristics over time, unlike more common and straightforward classifiers (Ren et al., 2009). HMMs can also cope with some sounds being extremely common whilst others rare (Ren et al., 2009). For example, the HMM approach detected 97% of bowhead whale (*Balaena mysticetus*) calls compared to 84% for matched filter methods (Weisburn et al., 1993). Matched filtering is a time series correlation method that uses synthetic waveforms instead of recording examples (Weisburn et al., 1993). Furthermore, HMMs were compared to dynamic time warping techniques for zebra finch calls. Dynamic time warping measures the similarity between time temporal sequences which may vary in speed, it is therefore often used in speech recognition to cope with different speaking speeds (Kogan and Margoliash, 1998). HMMs required more training but outperformed dynamic time warping when tested with noisy recordings that had confusing calls and notes (Kogan and Margoliash, 1998).

The detection of specific sounds over a large spatial or temporal scale is critical for supporting spatio-temporal management of ecologically important populations or species (Kalan et al., 2015; McDonald et al., 2017; Williamson et al., 2017). Here, we investigate the potential for using HMMs to detect cryptic soniferous species, such as the Bryde's whale (*Balaenoptera edeni*). Bryde's whales produce a variety of vocalizations, including low frequency pulses, tonals, moans and downsweeps from 60 to 950 Hz (Oleson et al., 2003; Heimlich et al., 2005; Rice et al., 2014; Roch et al., 2016). The relationship between geographic differences in vocalizations is unclear, but it has been suggested different vocalizations may delineate different stocks (Širović et al., 2014). In New Zealand, only long moans or down-sweeps have been attributed to Bryde's whales (McDonald, 2006; Constantine et al., 2015). The down-sweep vocalization matches the description of Be3 calls from the eastern tropical Pacific (Oleson et al., 2003; Heimlich et al., 2005) with little frequency modulation, frequency range 15–150 Hz and duration of 2.6 ± 0.8 s (Putland et al., 2017). There is now sufficient description of the sound to support efforts for automated acoustic detection. The stereotyped nature of the vocalization is ideal because a HMM recognizer can be trained to detect these specific characteristics.

From an ecological perspective, there has been limited research into the temporal variation of Bryde's whale vocalizations. In the eastern tropical Pacific, none of the five phrase types recorded showed a consistent seasonal pattern over the two years of monitoring (Heimlich et al., 2005). Furthermore, in the Gulf of Mexico Bryde's whales were heard sporadically throughout the 10 months of recording although the vocalization rate was significantly higher during dusk and night then dawn and day (Širović et al., 2014). In New Zealand the resident endangered population remains in close proximity to the Hauraki Gulf year round (Wiseman et al., 2011; Constantine et al., 2015). Our objective was to use HMMs to investigate diel and seasonal trends of Bryde's whale vocalizations in a long-term passive acoustic dataset. We then explored the potential implications of these findings for conservation and management efforts.

2. Materials and methods

2.1. Acoustic data

Passive acoustic monitoring was conducted in the Hauraki Gulf, New Zealand; a large, island-studded embayment recognized for its high biodiversity value as well as regular use by both recreational and commercial boating activities (Kelly et al., 2014). Acoustic data was collected using an omnidirectional hydrophone (ST202 Ocean Instruments, NZ; www.oceaninstruments.co.nz) that was calibrated prior and post deployment using a piston phone. The hydrophone was placed in Jellicoe Channel on the outer border of the Hauraki Gulf Marine Park (Fig. 1), from October 2014 to April 2016. This location was chosen because using normal mode propagation theory previous research in the Hauraki Gulf has shown the cut-off mode in 50 m of water was 22 Hz (Tindle et al., 1978; Tindle, 1982). Therefore, at Jellicoe Channel the low frequency Bryde's whale vocalization (peak frequency 35 Hz) would not be affected, whereas in shallower regions of the Gulf recordings may distort the sound signal. Furthermore, Bryde's whales have been seen in the area during marine mammal surveys (Kozmian-Ledward, 2014; Constantine et al., 2015). Bryde's whale vocalizations directly overlap anthropogenic sound from boat activity making monitoring their calls of utmost importance for mitigation strategy (Putland et al., 2017). The hydrophone was suspended 2 m off the seafloor and retrieved using an acoustic release (Desert Star Systems; www.desertstar.com). The hydrophone was pre-programed to sample at 144 kHz, 24 bits, for 2 mins every 20 min for the duration of each deployment (Table 1).

2.2. Training and testing

All recordings during October 2014 (1184 recordings) were reviewed aurally and visually using scrolling spectrograms in Audacity® (version 2.1.2) (Hanning window, Fast Fourier Transformation (FFT) length = 1024 and then FFT length = 16384) and annotated using the three categories: Bryde's whale (w), vessel passage (v) or ambient sound (a) (Fig. 2). Training files for building the automated recognizer included all recordings with Bryde's whale vocalizations during October 2014 (n = 94) as well as two full days of recordings (containing vessel passages and ambient sound) to represent the variety of background sound recorded at the station (n = 144). The MatlabHTK package (Ranjard et al., 2016) was used to load annotated recordings and generate a series of hidden Markov model toolkit (HTK) recognizers. According to the HMM principle (Young et al., 2006), each defined window within an input sound file has a certain probability of matching a specific state defined during the training stage of the recognizer. Five different recognizers were built using a variety of different frequency ranges: 10–500 Hz, 10–1000 Hz, 10–2000 Hz, 50–1000 Hz and 50–2000 Hz and the default MatlabHTK parametrization of the sound signal, window size 30 ms and number of cepstral parameters = 24 (Ranjard et al., 2016). The frequency ranges extended beyond the high frequency of Bryde's whale down-sweeps, because vessel sound is prevalent across all frequencies whereas Bryde's whale vocalizations are not (Putland et al., 2017).

Every two minute recording from the new moon (7th) and full moon (23rd) in November 2014 (144 recordings) was manually annotated