ISSN 2593-7568

RESEARCH ARTICLE

Multispecies Discrimination of Seals (Pinnipeds) using Hidden Markov Models (HMMs)

Marek B Trawicki[1*]

[1]*Marquette University, Milwaukee, USA.*

Abstract

Hidden Markov Models (HMMs) were developed and implemented for the discrimination of 5 available Seals (Pinnipeds), namely the Bearded Seal (*Erignathus barbatus*), Harp Seal (*Pagophilus groenlandicus*), Leopard Seal (*Hydrurga leptonyx*), Ross Seal (*Ommatophoca rossii*), and Weddell Seal (*Leptonychotes weddellii*). The main objectives of the experiments were to study the impact of the frame size and step size and number of states for feature extraction and acoustic models on classification accuracy. Based on the experiments using Mel-Frequency Cepstral Coefficients (MFCCs) extracted from the vocalizations (15 ms frame size and 4 ms step size), HMMs containing 20 states with single underlying Gaussian Mixture Model (GMM) produced discrimination of 95.77%. From the results, the framework could be applied to analysis for other marine mammals for both classification and detection of vocalizations and species.

Key Words: *Bioacoustics; Seals (Pinnipeds); Hidden Markov Models (HMMs); Classification; Species identification*

1. Introduction

Despite the superior performance of machine learning techniques for human speech processing tasks such as speech recognition and speaker identification [1], the application of the various methods (e.g., Hidden Markov Models (HMMs) [2], Gaussian Mixture Models (GMMs) [3], Artificial Neural Networks (ANNs) [4], and Deep Neural Networks (DNNs) [5] to bioacoustics has only begun to receive attention with the establishment of different frameworks to automatically identity animal vocalizations and species [6]. HMMs been particularly successful in performing song-type recognition and speaker identification for a large variety of mammals [7] such as the African Elephant (*Loxodonta africana*) [8], Norwegian Ortolan Bunting (*Emberiza hortulana*) [9], and Tiger (*Panthera tigris*) [10].

***Corresponding Author**: Marek B Trawicki, Marquette University, Milwaukee, USA; E-mail: marek.trawicki@marquette.edu*

***Received Date**: February 23, 2024, **Accepted Date**: March 17, 2024, **Published Date**: April 04, 2024*

***Citation:** Trawicki MB. Multispecies Discrimination of Seals (Pinnipeds) using Hidden Markov Models (HMMs). Int J Auto AI Mach Learn. 2024;4(1):1-9.*

In contrast, HMMs have been primarily successful in performing classification and detection of a limited number of marine mammals, namely whale vocalizations: Blue Whale (*Balaenoptera musculus*) [11], Bryde's Whale (*Balaenoptera brydei*) [12], and Killer Whale (*Orcinus orca*) [13]. Based on the promising results across numerous animals, HMMs can be applied to discriminate vocalizations for many other species of marine mammals.

Seals (Pinnipeds) are a clade of carnivorous, fin-footed, and semi-aquatic marine mammals consisting of thirty-three species whose taxonomy is not fully understood by researchers [14]. Although the species is widespread around the world, the majority of the seals prefer the cold waters of the Northern and Southern Hemispheres [15]. Typically, the species spend most of their lives in the water but arrive ashore to mate, birth, molt, and escape from predators [16]. As some of the more common seals, the Bearded Seal (*Erignathus barbatus*), Harp Seal (*Pagophilus groenlandicus*), Leopard Seal (*Hydrurga leptonyx*), Ross Seal (*Ommatophoca rossii*), and Weddell Seal (*Leptonychotes weddellii*) produce a number of similar vocalizations for communication purposes, including growls, trills, and warbles [17]. In order to better understand the species, the seals can be classified automatically through their vocalizations by utilizing HMMs for the task of species identification [18].

The remainder of this paper is organized into the following sections: Data (Section two), Methods (Section three), Results (Section four), and Conclusion (Section five).

2. Data

The Watkins Marine Mammal Sound Database [19] contains approximately 2000 unique recordings of more than 60 species of marine mammals (e.g., dolphins, seals, and whales) with over 15,000 annotated digital sound files spanning seven decades of work at the Woods Hole Oceanographic Institution (WHOI). From the database, the "Best of Cuts" category contains 1694 sound files of high sound quality and low noise from 32 different species. Table 1 summarizes the 142 recordings from the 5 available species of seals.

Table 1: *Database.*

Common Name	Binomial Name	Location	Class	Count
Bearded Seal	*Erignathus barbatus*	Alaska	BS	34
Harp Seal	*Pagophilus groenlandicus*	Gulf of St. Lawrence	HS	46
Leopard Seal	*Hydrurga leptonyx*	Antarctica	LS	10
Ross Seal	*Ommatophoca rossi*	Antarctica	RS	50
Weddell Seal	*Leptonychotes weddellii*	Antarctica	WS	2

Figure 1 provides the representative time series and spectrograms from the individuals to better understand the complexity of the vocalizations.

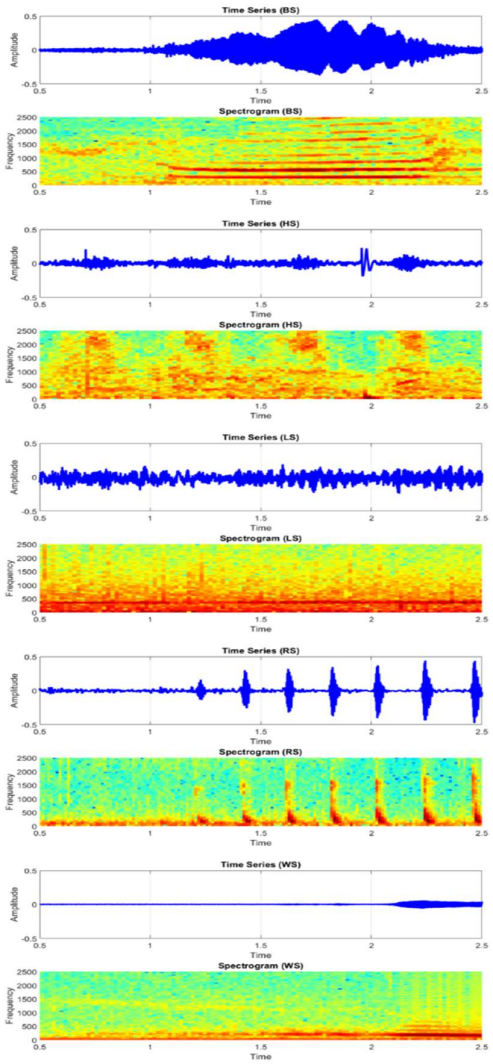

Figure 1: *Time series and spectrograms.*

Based on the spectrograms, the patterns in the vocalizations are distinctly visible for the Bearded Seal (BS), Harp Seal (HS), Leopard Seal (LS), Ross Seal (RS), and Wendell Seal (WS) across the range of frequencies 0 Hz–2500 Hz. Through the recordings, the vocalizations of the seals can be investigated to discriminate between the different species.

3. Methods

In order to perform classification through training and testing of the vocalizations, recordings must be parametrized into speech vectors that are then utilized for recognition. Given the set of training vocalizations corresponding to each particular model, the parameters of that model are determined automatically by a robust and efficient re-estimation procedure called the Baum-Welch Expectation Maximization algorithm [20,21]. Assuming training contains a sufficient number of representative vocalizations, the models are constructed to implicitly understand the many sources of variability. From the set of testing vocalizations, the likelihood of each model generating the vocalization is calculated quickly to determine the most likely model by the procedure called Viterbi algorithm [22]. Based on the classification

accuracy, the feature extraction and acoustic models can be adjusted to provide improvements in the recognition.

3.1. Feature extraction

Mel-Frequency Cepstral Coefficients (MFCCs) [23] are the classical features used in parametrization of vocalizations for numerous speech recognition applications [1]. According to the frame size and step size, the features are extracted on a frame-by-frame basis from the vocalizations. The cepstral coefficients c_i are calculated from the log filterbank amplitudes m_j using the Discrete Cosine Transform (DCT) as,

$$c_i = \sqrt{\frac{2}{N}} \sum_{j=1}^{N} m_j \cos\left(\frac{\pi i}{N}(j - 0.5)\right) \quad [1]$$

where, N is the number of filterbank channels. Since human beings perceive frequencies on a logarithmic scale [24], the filterbank channels are equally-spaced triangular filters with increasing bandwidths through increasing frequency f on the Mel scale defined as,

$$Mel(f) = 2595 \log_{10}\left(1 + \frac{f}{700}\right) \quad [2]$$

which approximates the behavior of the auditory systems and allows for the better representation of the vocalizations for recognition.
Figure **2** exhibits the extraction process of the MFCCs for each frame of the vocalizations.

Figure 2: *Feature extraction of MFCCs.*

3.2. Acoustic models

Hidden Markov Models (HMMs) [25] are statistical finite state machines used to model vocalizations and often considered as the dominant classification models in human speech processing [1] and, more recently, bioacoustics [6]. Fundamentally, HMMs consist of the following elements [1].

1. output observations $\boldsymbol{O} = \{o_1, o_2, \ldots, o_M\}$;
2. set of states $\boldsymbol{\Omega} = \{1,2, \ldots, N\}$;
3. transition probability matrices $\boldsymbol{A} = \{a_{ij}\}$;
4. output probability matrices $\boldsymbol{B} = \{b_i(k)\}$;
5. initial state distributions $\boldsymbol{\pi} = \{\pi_i\}$.

Collectively, the parameter set of HMMs are denoted as $\Phi = (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi})$.
Figure **3** supplies an example 4-state HMM with single Gaussian Mixture Models (GMMs) underlying each state.

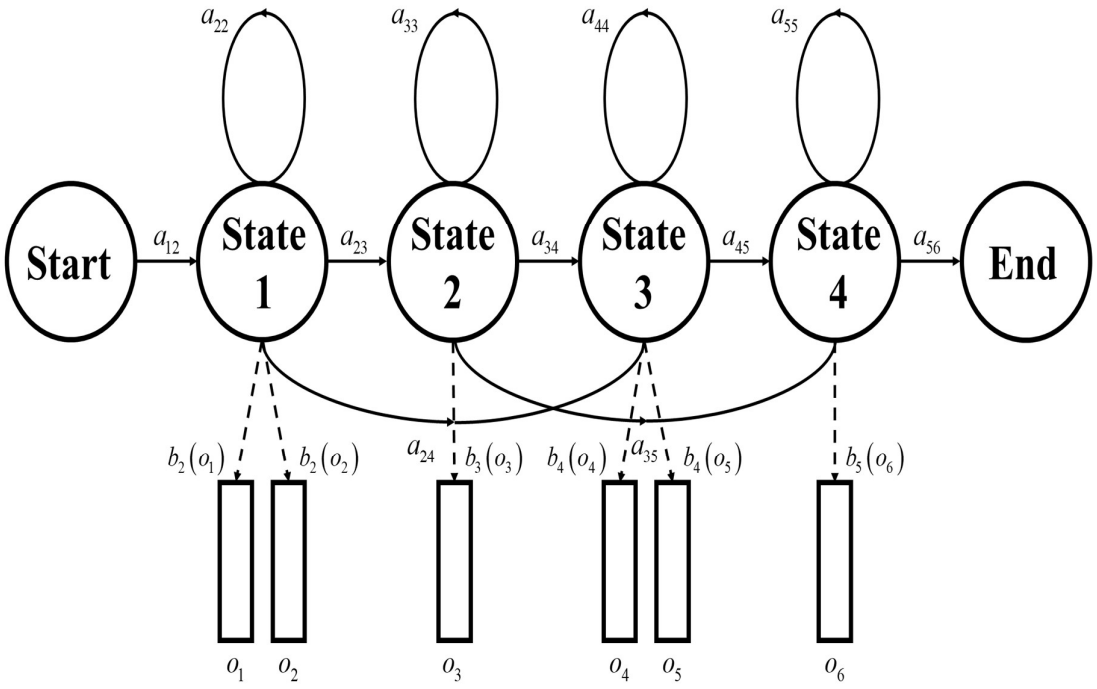

Figure 3: *Example 4-state HMM.*

4. Results

Experiments to discriminate between the 5 available species of seals were performed using leave-one-out cross-validation [26] on the 142 recordings in the database through the Hidden Markov Model Toolkit (HTK) [27]. The goals of the experiments are to determine the impact of frame size and step size and number of states for feature extraction and acoustic models on classification accuracy.

Table 2 displays the classification accuracy (number of correct classifications in parentheses) for variations in the frame size and step size (feature extraction).

Depending on the overlap between the frames of vocalizations (difference between frame size and step size (highlighted)), the classification accuracy ranges from 88.73%-95.77% (small frame sizes of 15 ms and 20 ms with relatively large overlap sizes of 9 ms–18 ms or percentages 60.00%-90.00%) to 90.15%-93.66% (large frame sizes of 20 ms and 25 ms with relatively small overlap sizes of 13 ms–22 ms or percentages 52.00%-73.33%).

By increasing the number of states in the HMMs from 1 state (GMM) to 20 states (HMM), the classification accuracy improves over 25% (36 additional correct classifications) to 95.77% (136/142 correct). From investigation of the frame size and step size (feature extraction) and number of states (acoustic models) utilizing 21 MFCC and 42 time derivative (21 delta and 21 delta-delta) features (feature vector size 63), the discrimination of the individuals was the highest value at 95.77% using 15 ms frame size and 4 ms step size along with 20 states containing a single GMM underlining the states of the HMMs, which are frame size and step size that could be potentially utilized as standard analysis conditions to extract features on a frame-by-frame basis from each of the vocalizations of other marine mammals.

Table 2: *Classification accuracy vs. frame size and step size.*

Accuracy		Step Size					
		2 ms	4 ms	6 ms	8 ms	10 ms	12 ms
Frame Size	15 ms	92.96% (132)	95.77% (136)	93.66% (133)	92.25% (131)	90.14% (128)	90.14% (128)
	20 ms	91.55% (130)	91.55% (130)	88.73% (126)	91.55% (130)	90.85% (129)	90.85% (129)
	25 ms	90.14% (128)	90.14% (128)	83.80% (119)	92.96% (132)	90.85% (129)	90.85% (129)
	30 ms	82.39% (117)	88.03% (125)	92.25% (131)	93.66% (133)	90.14% (128)	90.14% (128)

Table 3 shows the classification accuracy for changes in the number of states (acoustic models).

Table 3: *Classification accuracy vs. number of states.*

States	Correct	Accuracy
1	100	70.42%
5	114	80.28%
10	127	89.44%
15	135	95.07%
20	136	95.77%

Figure **4** illustrates the classification accuracy for each of the 5 available species of the seals.

Through the confusion matrix, the discrimination between the individuals was 100.00% for two of the five species of seals, namely the Harp Seal (HS) and Leopard Seal (LS). While the Ross Seal (RS) had the greatest number of misclassifications, the recordings were all classified as the Weddell Seal (WS) to produce a classification accuracy of 92.00%. Even though the Bearded Seal (BS) and Weddell Seal (WS) each had a single misclassification, the classification accuracy was drastically different for the two species of seals because of the availability of recordings: 97.06% (34/34 correct, Bearded Seal (BS)) as opposed to 50.00% (1/2 correct, Weddell Seal (WS)). Overall, the HMMs demonstrated the ability to accurately discriminate between the various 5 species of seals in the database, especially compared to chance (20.00%).

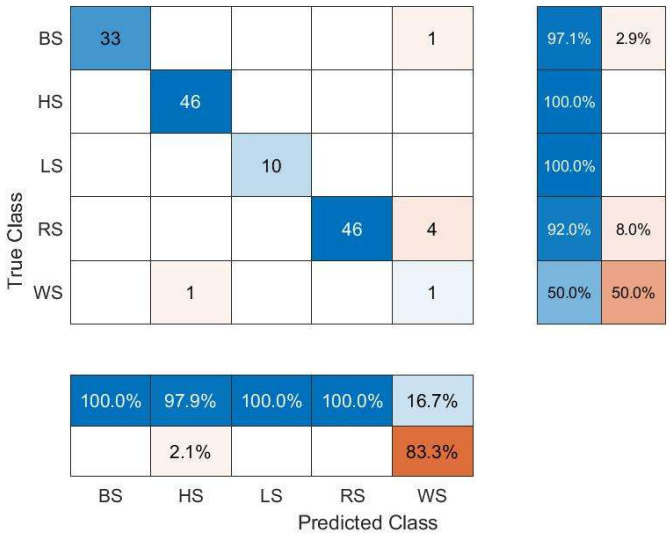

Figure 4: Confusion matrix for 5 available species of seals (Pinnipeds).

5. Conclusion

HMMs were developed and implemented for discrimination of 5 available species of seals from the WHOI database containing 142 recordings. The goals of the experiments were to determine the impact of the frame size and step size and number of states for feature extraction and acoustic models on classification accuracy. Through the study of the frame size and step size (feature extraction) and number of states (acoustic models) utilizing 21 MFCC and 42 times derivative (21 delta and 21 delta-delta) features (feature vector size 63), the discrimination of the seals achieved 95.77% using 15 ms frame size and 4 ms step size along with 20 states containing a single GMM underlining the states of the HMMs. For future work, HMMs could be applied to other marine mammals for classification and detection of vocalizations and species on a much more extensive database.

References

1. Huang X, Acero A, Hon HW, et al. Spoken Language Processing: A Guide to Theory, Algorithm, and System Development. Prentice Hall PTR, Upper Saddle River, New Jersey. 2001.

2. Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE. 1989;77:257-86.

3. Juang BH, Levinson S, Sondhi M. Maximum likelihood estimation for multivariate mixture observations of Markov chains (corresp). IEEE Trans Inf Theory. 1986;32:307-9.

4. Lippmann RP. Review of neural networks for speech recognition. Neural Comput. 1989;1:1-38.

5. Hinton G, Deng L, Yu D, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Signal Process Mag. 2012;29:82-97.

6. Clemins PJ. Automatic Classification of Animal Vocalizations. Marquette University, Wisconsin, United States. 2005.

7. Ren Y, Johnson MT, Clemins PJ, et al. A framework for bioacoustic vocalization analysis using hidden Markov models. Algorithms. 2009;2:1410-28.

8. Clemins PJ, Johnson MT, Leong KM, et al. Automatic classification and speaker identification of African elephant (*Loxodonta africana*) vocalizations. J Acoust Soc Am. 2005;117:956-63.

9. Trawicki MB, Johnson MT, Osiejuk TS. Automatic song-type classification and speaker identification of norwegian ortolan bunting (*Emberiza hortulana*) vocalizations. 2005 IEEE Workshop on Machine Learning for Signal Processing, Mystic, CT, USA. 2005.

10. Ji A, Johnson MT, Walsh EJ, et al. Discrimination of individual tigers (*Panthera tigris*) from long distance roars. J Acoust Soc Am. 2013;133:1762-9.

11. Buchan SJ, Mahu R, Wuth J, et al. An unsupervised hidden Markov model-based system for the detection and classification of blue whale vocalizations off Chile. Bioacoustics. 2020;29:140-67.

12. Putland RL, Ranjard L, Constantine R, et al. A hidden Markov model approach to indicate Bryde's whale acoustics. Ecol Indic. 2018;84:479-87.

13. Brown JC, Smaragdis P. Hidden Markov and gaussian mixture models for automatic call classification. J Acoust Soc Am. 2009;125:EL221-4.

14. Berta A, Churchill M. Pinniped taxonomy: review of currently recognized species and subspecies, and evidence used for their description. Mamm Rev. 2012;42:207.

15. Berta A. Return to the Sea: The Life and Evolutionary Times of Marine Mammals. University of California Press, Oakland, California. 2020.

16. Riedman M. The Pinnipeds: Seals, Sea Lions, and Walruses. University of California Press, Oakland, California. 1990.

17. Perrin WF, Wursig B, Thewissen JG. Encyclopedia of marine mammals. (2ndedn), Academic Press, London, UK. 2009.

18. Kvsn RR, Montgomery J, Garg S, et al. Bioacoustics data analysis–A taxonomy, survey and open challenges. IEEE Access. 2020;8:57684-708.

19. Watkins WA, Fristrup K, Daher MA, et al. SOUND database of marine animal vocalizations: structures and operations. Technical Report, Woods Hole Oceanographic Institution, Woods Hole, MA, USA. 1992.

20. Baum LE, Petrie T, Soules G, et al. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. Ann Math Stat. 1970;41:164-71.

21. Moon TK. The expectation-maximization algorithm. IEEE Signal Process Mag. 1996;13:47-60.

22. Forney GD. The viterbi algorithm. Proc IEEE. 1973;6:268-78.

23. Davis S, Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Trans Acoust. 1980;28:357-66.

24. Von Bekesy G. Experiments in Hearing. McGraw-Hill, New York, USA. 1989.

25. Rabiner L, Juang B. An introduction to hidden Markov models. IEEE Assp Mag. 1986;3:4-16.

26. Stone M. Cross-validatory choice and assessment of statistical predictions. J R Stat Soc Series B Stat Methodol. 1974;36:111-33.

27. Young S, Evermann G, Gales M, et al. The HTK Book (HTK version 3.4. 1). Cambridge University, Cambridge, England. 2009.