ISSN: 2687-8089

Advances in
Oceanography & Marine Biology

DOI: 10.33552/AOMB.2025.04.000589

Iris Publishers

Research article

Application of Hidden Markov Models (HMMS) for Multispecies Identification of Dolphins (*Cetaceans*)

Marek B. Trawicki*

Mathematical and Statistical Sciences, Marquette University, 1313 W. Wisconsin Avenue, Milwaukee, WI 53233, USA

***Corresponding author:** Marek B. Trawicki, Mathematical and Statistical Sciences, Marquette University, 1313 W. Wisconsin Avenue, Milwaukee, WI 53233, USA

Received Date: May 16, 2025

Published Date: May 23, 2025

Abstract

Hidden Markov Models (HMMs) were developed and applied to distinguish among the 12 available Dolphins (Cetaceans) species ranging from the Atlantic Spotted Dolphin (Stenella Frontalis) to the White-Sided Dolphin (Lagenorhynchus Acutus). The primary goals of the analyses were to investigate how varying frame duration and frame shift, dimensions of the feature vector, and number of states for feature extraction and acoustic models affect classification accuracy. In the studies utilizing 41 Mel-Frequency Cepstral Coefficients (MFCCs) extracted from the vocalizations with 5ms frame size and 2ms step size, HMMs comprising 14 states with an individual Gaussian Mixture Model (GMM) achieved classification performance spanning 63.89% (4 classes) to 100.00% (1 class) with a discrimination accuracy of 80.25% (12 classes). Based on the outcomes, the system could be extended to the study of other marine mammals for investigation of vocalizations and species.

Keywords: Bioacoustics; dolphins (cetaceans); hidden markov models (HMMs); classification; species identification

Introduction

Recently, the utilization of machine learning techniques (e.g., Hidden Markov Models (HMMs) [1] and Gaussian Mixture Models (GMMs) [2]) to bioacoustics has started to attract more interest [3]. In more precise terms, HMMs have proven effective in classifying song-types and identifying speakers across various mammalian species [4] such as the African Elephant (*Loxodonta Africana*) [5], Norwegian Ortolan Bunting (*Emberiza Hortulana*) [6], and Tiger (*Panthera Tigris*) [7]. In contrast, there has been rather limited research on classifying and detecting dolphin sounds [8] and dolphins themselves [9] using HMMs and GMMs. Given the promising findings, HMMs can be utilized to differentiate vocalizations of other mammals, particularly various species of marine mammals [10]. Dolphins (*Cetaceans*) are aquatic marine mammals consisting with forty living species varying in length and weight from 1.7 meters and 50 kilograms to 9.5 meters and 11 short ton [11].

While the species is spread around the world, the dolphins typically prefer the warmer waters of tropic zones [12]. In the water, the dolphins feed, mate, and birth along with escape from few marine enemies [13]. Through the Atlantic Spotted Dolphin (*Stenella Frontalis*) to the White-Sided Dolphin (*Lagenorhynchus Acutus*), dolphins generate a wide array of vocalizations, encompassing clicks (burst pulses) and whistles (frequency modulated), by vibrating connective tissue akin to the function of human vocal cords [14]. In order to enhance our understanding of the species, HMMs can automate the classification of dolphins from their vocalizations utilizing HMMs for species identification [15]. The rest of the paper is arranged into the subsequent sections: Data (Section 2), Methods (Section 3), Results (Section 4), and Conclusion (Section 5).

Data

The Watkins Marine Mammal Sound Database [16] comprises around 2000 distinct recordings from over 60 species of marine mammals such as dolphins, seals, and whales, including over 15,000 annotated digital sound files representing over 70 years of research conducted at the Woods Hole Oceanographic Institution (WHOI). Within the database, the "Best of Cuts" section encompasses 1694 high-quality sound files with minimal noise, which represents 32 distinct species. Figure 1 provides the time series and spectrograms of individual vocalizations, which offers insights into the complexity of the sounds. By relying solely on visual inspection, the discernment of patterns in the vocalizations is a challenge from simply the spectrograms. As a result, the vocalizations of the dolphins will be investigated through machine learning to discriminate between the different species (Table 1).

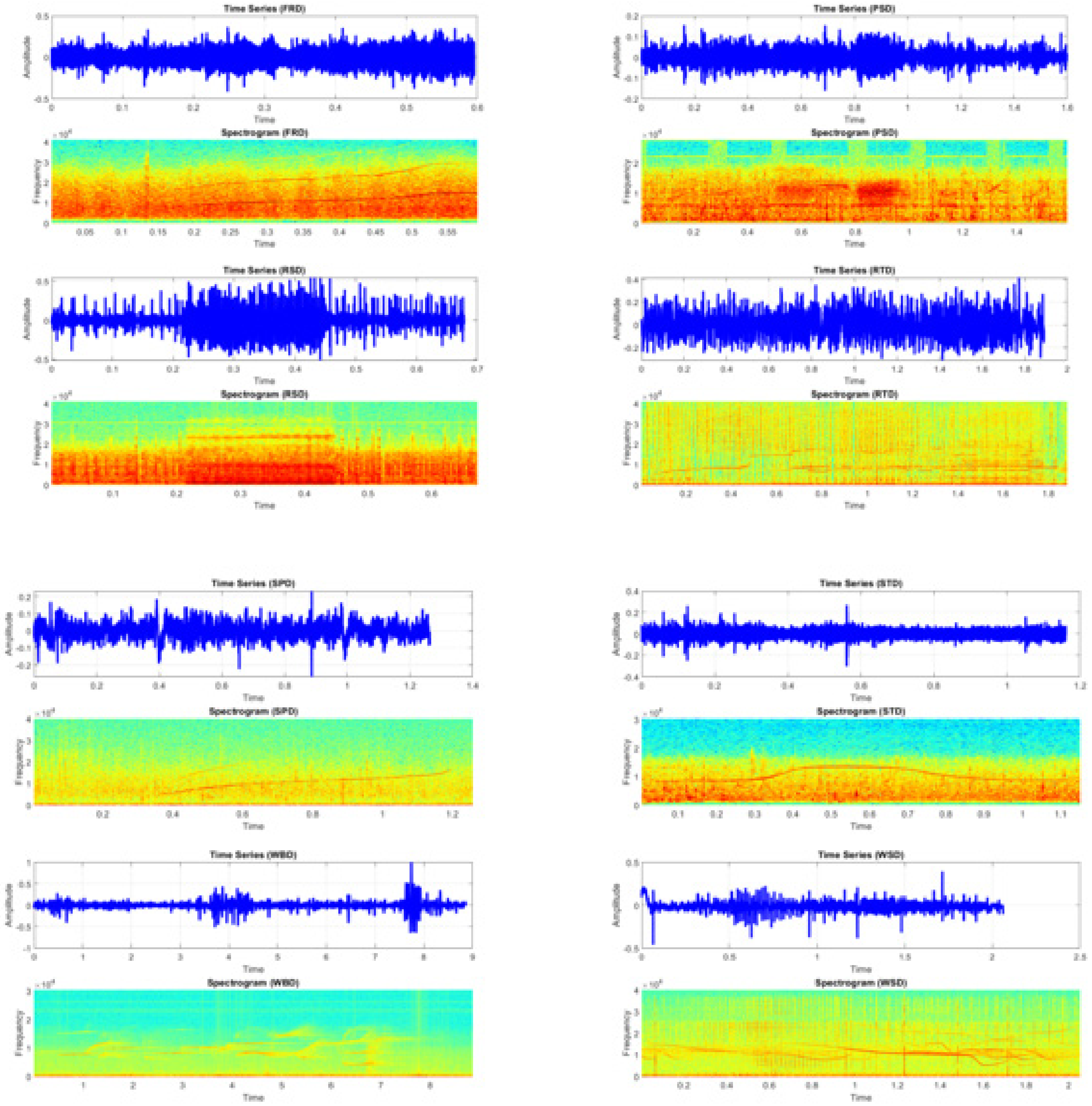

Figure 1: Time Series and Spectrograms.

Table 1: Summary of the 760 Recordings Spanning 12 Dolphin Species.

Common Name	Binomial Name	Location	Class	Count
Atlantic Spotted Dolphin	*Stenella Frontalis*	Unknown	ASD	57
Bottlenose Dolphin	*Tursiops Truncatus, Tursiops Gilli*	California, Florida	BND	24
Clymene Dolphin	*Stenella Clymene*	Bequia, St. Lucia	CMD	62
Common Dolphin	*Delphinus Delphis*	Canada, Canary Islands	CND	47
Fraser's Dolphin	*Lagenodelphis Hosei*	Dominica	FRD	87
Pantropical Spotted Dolphin	*Stenella Attenuata*	Costa Rica, Dominica	PSD	65
Risso's Dolphin	*Grampus Griseus*	Canary Islands, North Atlantic, Selina, Ustica	RSD	67
Rough-Toothed Dolphin	*Steno Bredanensis*	Malta, Sicily	RTD	50
Spinner Dolphin	*Stenella Longirostris*	Hawaii	SPD	111
Striped Dolphin	*Stenella Coeruleoalba*	Canada, Delaware, Italy	STD	80
White-Beaked Dolphin	*Lagenorhynchus Albirostris*	Maine, Massachusetts	WBD	57
White-Sided Dolphin	*Lagenorhynchus Acutus*	Brown's Bank, Massachusetts, Stellwagen Bank	WSD	53

Methods

To facilitate the discrimination of vocalizations through training of testing, recordings must initially be parameterized into speech vectors, which are subsequently used for recognition. By using a collection of training vocalizations corresponding to each specific model, the parameters are automatically determined through a robust and efficient re-estimation procedure called Baum-Welch Expectation Maximization [17,18]. Assuming the training data includes an adequate number of vocalizations, the models are designed to implicitly grasp the diverse sources of variability. From the collection of testing vocalizations, the likelihood of each model generating the vocalization is calculated swiftly to determine the most probable model by Viterbi [19]. Based on the classification accuracy, adjustments can be implemented to the feature extraction and acoustic models to enhance recognition.

Feature Extraction

Mel-Frequency Cepstral Coefficients (MFCCs) [20] represent the classical features utilized in vocalization parameterization for various speech recognition applications [21]. By adjusting to the frame duration and frame shift, features are extracted on a frame-by-frame basis to take advantage of their stationary nature. Upon the application of the Hamming window, the Fourier Transform is computed for each short-duration frame. Given human perception of frequencies follows a log scale [22], the feature extraction that accommodates the non-linear frequency behavior across the entire audio frequency range provides a more accurate approximation of the auditory system response and enhances vocalization representation for recognition purposes. To achieve the desired non-linear frequency resolution, filterbank channels are organized as equally-spaced triangular filters with increasing bandwidths relative to the frequency f on the Mel scale defined as

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

Figure 2 reveals the distribution of the Mel scale filterbanks spanning frequencies up to the Nyquist frequency f_N.

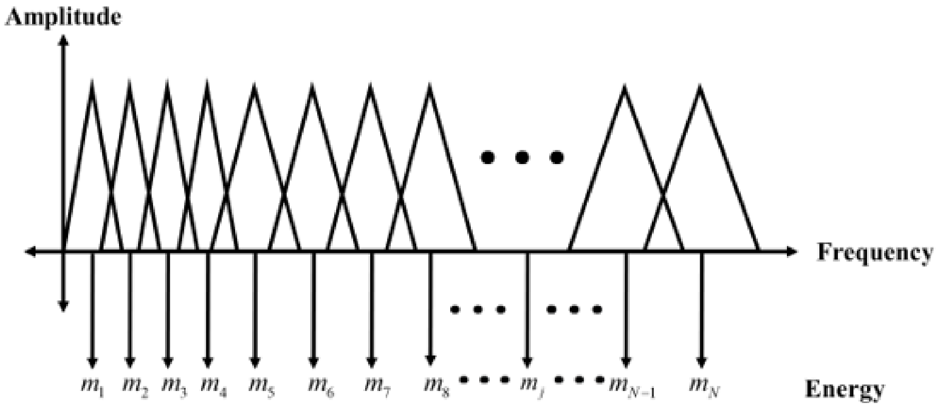

Figure 2: Mel Scale Filterbanks.

The cepstral coefficients c_i are obtained from the logarithm of the filterbank amplitudes m_j utilizing the Discrete Cosine Transform (DCT) as

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

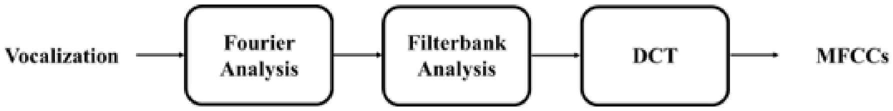

Figure 3: Feature Extraction of MFCCs.

Where N is the number of filterbank channels. Figure 3 exhibits the core process of deriving the MFCCs for each vocalization frame.

Acoustic Models

Hidden Markov Models (HMMs) [23] are statistical finite state machines used to model vocalizations, which are commonly considered one of the primary classification models in human speech processing [21] and bioacoustics [3]. Table 2 states the fundamental elements of HMMs [21]. Together, the set of parameters of HMMs is referred to as $\Phi = (A, B, \pi)$. Figure 4 supplies an example of a 4-state HMM with individual Gaussian Mixture Models (GMMs) under each state. Each state corresponds to a single, predetermined statistical model, and the choice of states in the HMMs corresponds directly to the number of distinct temporal segments in the vocalization.

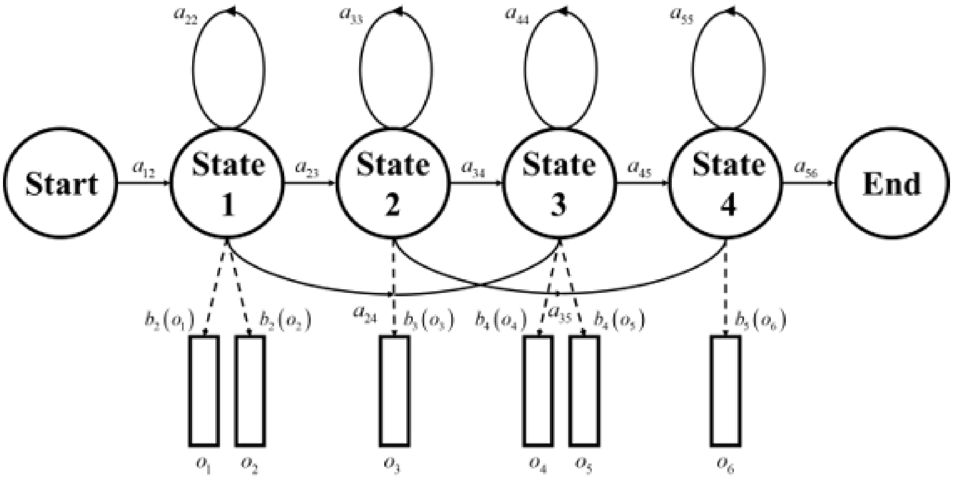

Figure 4: Example of 4-State HMM.

Results

Experiments aimed at discriminating between the 12 species of dolphins were conducted through cross-validation [24] on the 760 vocalizations in the database, partitioned into 80% for training and 20% for testing, utilizing the Hidden Markov Model Toolkit (HTK) [25]. The objectives of the experiments were to assess the impact of frame duration and frame shift, dimensions of the feature, and states on feature extraction and acoustic models on discrimination using 157 test vocalizations and 100 filterbank channels. Table 3 displays the classification accuracy with the number of correct classifications indicated in parentheses for different configurations. With any specific frame size, the classification accuracy generally rose in value with the increase in step size corresponding to more overlap between adjacent frames (1ms – 3ms): 71.34% – 73.25% (4ms frame size), 65.61% – 80.25% (5ms frame size), 69.43% – 75.80% (6ms frame size), and 63.06% – 71.97% (7ms frame size). Table 4 shows the classification accuracy corresponding to variations in the number of cepstral coefficients.

By augmenting the number of cepstral coefficients up to 41 MFCC features along with the time derivative (41 delta and 41 delta-delta) features resulting in a feature vector size 123, the classification accuracy achieved its peak value at 80.25% (126/157 correct classifications), which represents an enhancement of 18.47% over the smallest number of cepstral coefficients. Table 5 indicates the classification accuracy for variations in the states. With the increase in the number of states within the HMMs (transitioning from 1 state (GMM) to 14 states (HMM)), the classification accuracy improved significantly by 33.75% (53 additional correct classifications) to 80.25% (126/157 correct classifications). Figures 5&6 illustrate the classification accuracies relative to the number of whale species and all 12 available dolphin species. Upon exploration of frame duration and frame size and dimension of feature vector (feature extraction) and number of states (acoustic models) with 41 MFCC and 82 times derivative (41 delta and 41 delta-delta) features (feature vector size 123), discrimination ranged from 63.89% (4 classes) to 100.00% (1 class) with 80.25% (12 classes) using 5ms frame size and 2ms step size along with 14 states containing a single GMM underlining the states of the HMMs.

Table 2: Elements of HMMs.

Quantity	Notation
Output Observations	$\mathbf{O} = \left\{o_1, o_2, \ldots, o_M\right\}$
Set of States	$\mathbf{\Omega} = \left\{1, 2, \ldots, N\right\}$
Transition Probability Matrices	$\mathbf{A} = \left\{a_{ij}\right\}$
Output Probability Matrices	$\mathbf{B} = \left\{b_i\left(k\right)\right\}$
Initial State Distributions	ð $= \left\{\pi_i\right\}$

Table 3: Classification Accuracy vs. Frame Size and Step Size.

Accuracy			Step Size	
1 ms		2 ms	3 ms	
Frame Size	4 ms	71.34% (112)	71.97% (113)	73.25% (115)
	5 ms	65.61% (103)	80.25% (126)	73.89% (116)
	6 ms	69.43% (109)	75.80% (119)	73.25% (115)
	7 ms	63.42% (98)	63.06% (99)	71.97% (113)

Table 4: Classification Accuracy vs. Number of Cepstral Coefficients.

Cepstral Coefficients	Feature Vector Size	Correct	Accuracy
5	15	97	61.78%
9	27	111	70.70%
13	39	105	66.88%
17	51	106	67.52%
21	63	105	66.88%
25	75	112	71.34%
29	87	107	68.15%
33	99	122	77.71%
37	111	120	76.43%
41	123	126	80.25%

Table 5: Classification Accuracy vs. Number of States.

States	Correct	Accuracy
1	73	46.50%
2	86	54.78%
4	103	65.61%

6	105	66.88%
8	113	71.98%
10	120	76.43%
12	118	75.16%
14	126	80.25%

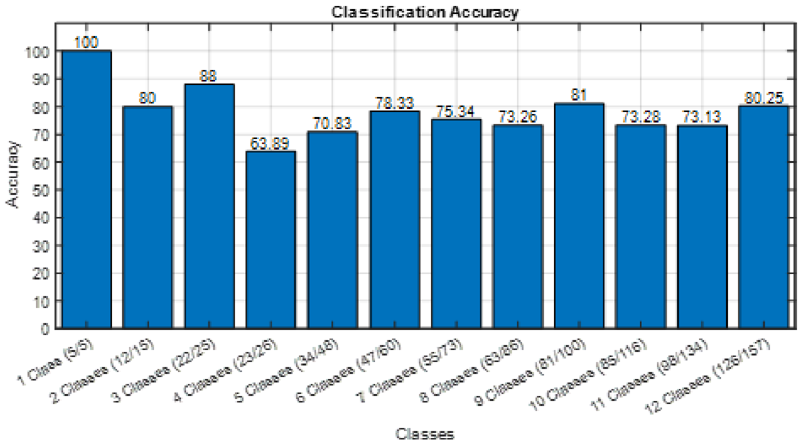

Figure 5: Classification Accuracy vs. Number of Classes.

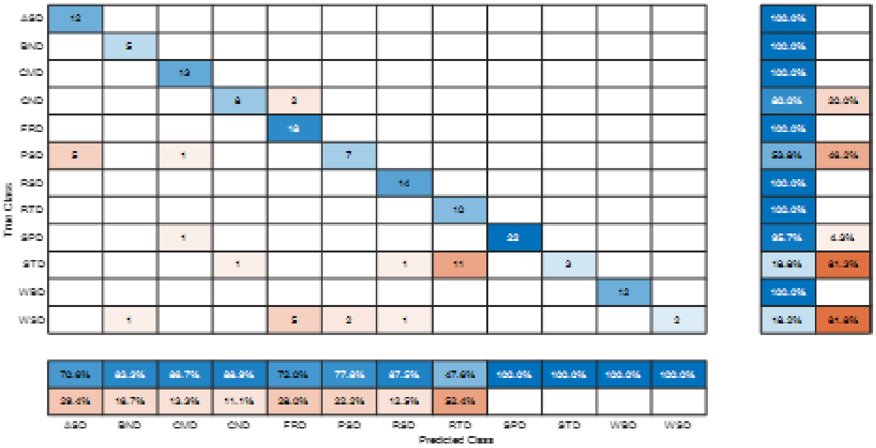

Figure 6: Confusion Matrix for 12 Available Species of Dolphins.

For three species of dolphins, the classification accuracies were only 18.18% (2/11 correct classification, White-Sided Dolphin (WSD)), 18.75% (3/16 correct classification, Striped Dolphin (STD)), and 53.85% (7/13 correct classification, Pantropical Spotted Dolphin (PSD)), which considerably reduced the overall classification accuracy of the 12 classes of dolphins. In the end, the HMMs demonstrated capability to effectively differentiate among the diverse species of dolphins within the database.

Conclusion

Hidden Markov Models (HMMs) were developed and applied to discern among the 12 species of dolphins sourced from the WHOI database, which contains 760 vocalizations. The primary objectives of the study were to assess the influence of frame duration and frame shift, dimension of the feature vector, and number of states for feature extraction and acoustic models on discrimination. From the analysis of the frame duration and frame shift (feature extraction), dimension of the feature vector (feature extraction), and number of states (acoustic models), the HMMs revealed robust classification accuracies between the individuals, including by increasing the number of dolphin species. Through 41 MFCC and 82 times derivative (41 delta and 41 delta-delta) features (feature vector size 123), discrimination of the dolphins ranged from 63.89% (4 classes) to 100.00% (1 class) with 80.25% (12 classes) using 5ms frame size and 2ms step size along with 14 states containing a single GMM under the states of the HMMs. In future research, HMMs offer potential applications in classifying and detecting vocalizations and species of other marine mammals.

Statements and Declarations

Funding: N/A.

References

1. L Rabiner (1989) A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proceedings of the IEEE 77: 257-286.
2. B Juang, SE Levinson, M Sondhi (1986) Maximum Likelihood Estimation for Multivariate Mixture Observations of Markov Chains. IEEE Transactions on Information Theory 32(2): 307-309.
3. PJ Clemins (2005) Automatic Classification of Animal Vocalizations. Milwaukee: Marquette University.
4. YJ Ren, M Johnson, PJ Clemins, M Darre, S S Glaeser, et al. (2009) A Framework for Bioacoustic Vocalization Analysis using Hidden Markov Models. Algorithms 2(4): 1410-1428.
5. PJ Clemins, MT Johnson, KM Leong, A Savage (2005) Automatic Classification and Speaker Identification of African Elephant (Loxodonta Africana) Vocalizations. The Journal of the Acoustical Society of America 117(2): 956-963.
6. MB Trawicki, MT Johnson, T Osiejuk (2005) Automatic Song-Type Classification and Speaker Identification of Norwegian Ortolan Bunting (Emberiza Hortulana) Vocalizations. in 2005 IEEE Workshop on Machine Learning for Signal Processing, Mystic.
7. An Ji, MT Johnson, EJ Walsh, J McGee, DL Armstrong (2013) Discrimination of Individual Tigers (Panthera Tigris) from Long Distance Roars. The Journal of the Acoustical Society of America 133(3): 1762-1769.
8. P P Parada, A Cardenal-Lopez (2014) Using Gaussian Mixture Models to Detect and Classify Dolphin Whistles and Pulses. The Journal of the Acoustical Society of America 135(6): 3371-3380.
9. MA Roch, MS Soldevilla, JC Burtenshaw, EE Henderson, J A Hildebrand (2007) Gaussian Mixture Model Classification of Odontocetes in the Southern California Bight and the Gulf of California. The Journal of the Acoustical Society of America 121(3): 1737-1748.
10. MJ Bianco, P Gerstoft, J Traer, E Ozanich, MA Roch, et al. (2019) Machine Learning in Acoustics: Theory and Applications. The Journal of the Acoustical Society of America 146(5): 3590-3628.
11. D Ranneft, H Eaker, R Davis (2001) A Guide to the Pronunciation and Meaning of Cetacean Taxonomic Names. Aquatic Mammals 27(2): 183-195.
12. A Berta (2012) Return to the Sea: The Life and Evolutionary Times of Marine Mammals. Berkeley: University of California Press.
13. W F Perrin, B Wursig, J G M Thewissen (2009) Encyclopedia of Marine Mammals. Academic Press: Cambridge.
14. W A Whitlow (2012) The Sonar of Dolphins. Berlin: Springer Science & Business Media.
15. RR KVSN, J Montgomery, S Garg, M Charleston (2020) Bioacoustics Data Analysis - A Taxonomy, Survey and Open Challenges. IEEE Access 8: 57684-57708.
16. W A Watkins, K Fristrup, M A Daher, T Howald (1992) SOUND Database of Marine Animal Vocalizations Structure and Operations. Woods Hole Oceanographic Institution, Woods Hole.
17. L E Baum, T Petrie, G Soules, N Weiss (1970) A Maximization Technique Occurring in the Statistical Analysis of Probability Functions of Markov Chains. The Annals of Mathematical Statistics 41(1): 164-171.
18. T K Moon (1996) The Expectation-Maximization Algorithm. IEEE Signal Processing Magazine 13(6): 47-60.
19. G Forney (1973) The Viterbi Algorithm. Proceedings of IEEE 61(3): 268-278.
20. S B Davis, P Mermelstein (1980) Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. IEEE Transactions on Acoustics, Speech, and Signal Processing 28(4): 357-366.
21. X Huang, A Acero, H W Hon (2001) Spoken Language Processing. Upper Saddle River: Prentice-Hall, Inc.
22. G Von Bekesy (1989) Experiments in Hearing. New York City: McGraw-Hill Book Company.
23. L Rabiner, B Juang (1986) An Introduction to Hidden Markov Models. IEEE ASSP Magazine 3(1): 4-16.
24. M Stone (1974) Cross-Validatory Choice and Assessment of Statistical Predictions. Journal of the Royal Statistical Society: Series B Methodological 36(2): 111-147.
25. S Young, G Evermann, M Gales, T Hain, D Kershaw, et al. (2009) Hidden Markov Model Toolkit (HTK) (Version 3.4). Cambridge: Cambridge University Engineering Department.

Citation: Marek B. Trawicki*. Application of Hidden Markov Models (HMMS) for Multispecies Identification of Dolphins (*Cetaceans*). Ad Oceanogr & Marine Biol. 4(3): 2025. AOMB.MS.ID.000589. **DOI:** 10.33552/AOMB.2025.04.000589.