

# CNN 모델을 이용한 수중음향 자료의 돌고래 휘슬음 자동 분류 연구

윤영글<sup>1</sup> · 김한수<sup>1</sup> · 김범규<sup>1,2</sup> · 조성호<sup>1</sup> · 강돈혁<sup>1</sup> · 김선효<sup>1\*</sup>

<sup>1</sup>한국해양과학기술원 해양력강화·방위연구부, <sup>2</sup>부경대학교 인공지능융합학과

## Research on Automatic Classification of Dolphin Whistle in Underwater Acoustic Data Using Convolution Neural Network (CNN) Model

Young Geul Yoon<sup>1</sup>, Hansoo Kim<sup>1</sup>, Bym-Kyu Kim<sup>1,2</sup>, Sungho Cho<sup>1</sup>, Donhyug Kang<sup>1</sup> and Sunhyo Kim<sup>1\*</sup>

<sup>1</sup>Sea Power Reinforcement Security Research Department, Korea Institute of Ocean Science and Technology (KIOST), Busan 49111, Republic of Korea

<sup>2</sup>Department of Artificial Intelligence Convergence, Pukyong National University, Busan 48517, Republic of Korea

With the advancement of underwater acoustic observation technology and the development of high-performance equipment, passive acoustic monitoring (PAM) has become a widely used method for monitoring ecosystem cetaceans. As the amount of high-resolution acoustic data increases exponentially, researchers continuously strive to automate the classification of marine mammal sounds and improve the speed of acoustic data analysis. This study developed a method for automatically classifying the dolphin whistles of common dolphins, false killer whales, and bottlenose dolphins using a convolutional neural network model. Whistle signals were analyzed using three spectrogram images: Original, median filter, and edge detection. The F1-Scores analysis results for the three data types were 84.40%, 84.65%, and 80.29%, confirming the possibility of automating whistle classification for the three dolphin species. In Case 2 of the F1-Scores, the common dolphin scored 90.90%, the false killer whale scored 80.00%, and the bottlenose dolphin scored 85.71%, indicating that the common dolphin showed relatively high automatic detection. This study contributes to developing an automatic detection and classification technique for cetacean species in South Korea. Long-term underwater acoustic measurement data acquired through PAM can be utilized more effectively in the future.

Keywords: Deep learning, Convolution neural network (CNN), Underwater acoustics, Whistle, Automatic classification

### 서론

고래류는 높은 경제적 가치를 가지며, 생태계의 건강 상태를 유지 및 감시하는 종으로서 매우 중요한 역할을 한다(Parsons et al., 2015; Usman et al., 2020). 연구자들을 비롯한 많은 사람들은 고래류의 중요성으로 인해 보존과 보호에 관심이 많지만 아직까지 고래류의 생태계 특성에 대한 지식은 아직 충분히 확보되지 않은 상황이다(Zimmer, 2011; Putland et al., 2018; Usman et al., 2020). 고래류는 수중 환경에서 제한된 빛 투과로 인해 일상 생활의 모든 측면에서 음파를 활용하는데 특정 종류의 소리를 발생시킨다(Allen et al., 2021). 최근에는 고래류 관

찰에 친환경적인 방식으로 수중청음기를 사용하여 수중 환경에서 고래류의 소리를 장시간 측정하여 연구하는 수동 음향 모니터링(passive acoustic monitoring, PAM) 방법이 고래류 연구에 활용되고 있는 추세이다(Mellinger et al., 2007; Gibb et al., 2019, Usman et al., 2020). 수중음향을 이용한 방법은 고래류를 기상 상태에 관계없이 연속적으로 모니터링이 가능하며, 사람이 직접 관찰하기 어려운 지역을 조사할 수 있다는 장점이 있다(Gridley et al., 2020; Allen et al., 2021). PAM을 통해 종의 출현과 계절성(Clark et al., 2002; Širović et al., 2003; Munger et al., 2008), 개체 군 이동(Stafford et al., 1999; Burtenshaw et al., 2004), 소리 유형의 계절성(Winn and Winn, 1978; Oleson

\*Corresponding author: Tel: +82. 51. 664. 3644 Fax: +82. 51. 719. 9591

E-mail address: sunhyo@kiost.ac.kr



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

<https://doi.org/10.5657/KFAS.2024.0743>

Korean J Fish Aquat Sci 57(6), 743-751, December 2024

Received 31 October 2024; Revised 2 December 2024; Accepted 4 December 2024

저자 직위: 윤영글(선임기술원), 김한수(선임연구원), 김범규(대학원생), 조성호(책임연구원), 강돈혁(책임연구원), 김선효(선임연구원)

et al., 2007) 및 발생빈도(McDonald and Fox, 1999; Marques et al., 2013; Küsel et al., 2016) 등을 조사할 수 있다. 또한, 장기 관측을 통해 개체 수 변화와 이동 또는 서식지 환경에 대한 정보를 획득할 수 있으며, 해양 환경 변화에 대한 인간의 영향을 파악하거나 부정적인 결과를 줄이는 측면에서 중요한 역할을 할 수 있다(André et al., 2011; Zimmer, 2011; Usman et al., 2020; Allen et al., 2021).

최근 수중음향 관측 기술의 발전과 고성능 장비가 개발됨에 따라 고해상도 음향 자료의 양이 기하급수적으로 증가하고 있는 추세이다. 이로 인해 현재는 자료 수집 보다는 분석하는 데 걸리는 시간이 증가하므로 연구자들은 고래 소리 식별을 자동화하여 음향 자료 분석 속도를 향상시키기 위하여 지속적으로 노력하고 있는 추세이다(Bittle and Duncan, 2013; Allen et al., 2021). 자동 탐지 기술은 다양한 신호 처리를 통한 음향 특성 기반의 탐지 및 분류가 활용되었으나, 최근에는 기계 학습 기술을 적용하여 향상된 탐지 및 분류 방법이 개발되고 있다(Carusio et al., 2020; Allen et al., 2021; Bergler et al., 2022; White et al., 2022; Ziegenhorn et al., 2022; Frainer et al., 2023). 특히 이미지 분류나 객체 식별에서 뛰어난 성능을 보이는 합성곱 신경망(convolutional neural network, CNN) 기법을 활용한 생태학 연구가 점점 증가하고 있다(Weinstein, 2018; Allen et al., 2021; Kim et al., 2022). 국내에서 고래류에 대한 음향 연구는 큰돌고래(Shin et al., 2002), 남방큰돌고래(Choi et al., 2019), 참돌고래(Yoon et al., 2017), 흰고래(Choi et al., 2017), 상괭이(Yoon et al., 2021; Kim et al., 2024) 등 국내에 서식하는 고래 종에 대한 음향 특성 연구를 중심으로만 진행되어 왔다. 최근에는 해양 포유류의 음향 신호를 자동으로 탐지하기 위해 인공지능 기법을 적용한 연구도 수행되고 있다(Kim et al., 2022; Jeong et al., 2023). 그러나 이러한 인공지능 기법은 주로 음향 신호의 탐지에만 활용되고 있으며, 종 식별 연구에는 아직 적용되지 않아 고래 연구에서 PAM을 적용하기에는 여전히 제한적인 상황이다.

본 연구에서는 한반도 인근 해역에서 주로 서식 및 출현하는 대표적인 돌고래(Lee et al., 2022) 중 참돌고래(common dolphin *Delphinus delphis*), 흑범고래(false killer whale *Pseudorca crassidens*), 큰돌고래(bottlenose dolphin *Tursiops truncatus*)의 휘슬음을 기반으로 CNN 기법을 적용하여 자동 식별에 대한 가능성을 확인해 보았다. 복잡한 수중음향 환경에서 나타나는 고래의 휘슬음들이 인공지능 기법의 적용을 통해 효과적으로 탐지 및 분류가 가능하다면, 현재까지 국내에서 많이 활용되고 있지 않은 PAM의 활용성을 향상시킬 수 있을 것이다. 향후, 국내에서 단기/장기간의 해양 포유류 모니터링을 포함한 수중 사운드스케이프(underwater soundscape) 연구를 지속적으로 수행하기 위해서는 대용량 수중음향 자료를 분석할 수 있는 자동화된 음향 식별 기술 개발이 필수적이며, 이는 고래의 생태학적 특성을 파악하고 이해하는데 많은 정보를 제공할 것으

로 기대된다.

## 재료 및 방법

### 돌고래 음원 자료 및 증강 방법

돌고래의 휘슬음 데이터는 “Whale FM” 웹 사이트 내에 공개되어 있는 음원 파일 중, 국내 인근 해역에서 많이 출현하는 참돌고래, 흑범고래, 큰돌고래를 선정하였다. 참돌고래 휘슬음의 주파수 범위는 3–28 kHz, 흑범고래는 1–35 kHz, 큰돌고래는 1–35 kHz로 알려져 있으며, 음향학적 특성은 지리적 위치, 사회 구조 및 환경 조건에 따라 다양하게 변동한다(McCullough et al., 2021; Antichi et al., 2023; Carlón-Beltrán et al., 2023). 휘슬음의 신호 길이는 고래 별로 상이하고, 한 개의 음원 파일 내에 다수의 휘슬음이 기록되어 있는 데이터가 존재한다. 따라서 음원 파일은 하나의 휘슬음이 포함된 1초 단위의 데이터로 변환하였으며, 각각의 음원 파일은 50개, 40개, 23개이다.

각각의 휘슬음은 학습·검증·평가으로 약 6:1:3으로 설정하였으나, 학습을 위한 데이터의 양이 충분하지 않았다. 따라서, 데이터 부족 문제를 해결하고 일반화 성능을 향상시키기 위해, 학습 및 검증에 활용하기 위한 휘슬음이 포함되어 있는 1초 단위의 원 신호 79개(전체 음원 파일의 약 70%)를 시계방향과 반시계 방향으로 0.5°로 각각 회전시켜 추가적인 데이터를 생성하였으며 전체 237개의 데이터로 증강하였다. 또한, 증강한 데이터에 가우시안 노이즈 및 랜덤 노이즈를 각각 0.3% 적용하여 추가적인 자료 증강을 통해 총 711개의 데이터를 구축하였다. 학습 및 검증에 사용된 돌고래 휘슬음 자료는 각각 참돌고래 315개, 흑범고래 252개, 큰돌고래 144개이다. 자료 증강을 수행하지 않고 모델 평가에 활용한 고래 휘슬음 원 자료는 34개(전체 음원 파일의 약 30%)로 각각 참돌고래 15개, 흑범고래 12개, 큰돌고래 7개이다.

### CNN 모델 구축 및 하이퍼파라미터 설정

최근 자연어 처리 모델의 성능이 향상됨에 따라 다양한 분야에서 높은 성능을 내기 때문에, 딥러닝 기반의 모델이 사용되고 있다. 대표적인 연구 중 하나가 컴퓨터 비전 분야의 비전 트랜스포머(vision transformer, ViT)이다(Dosovitskiy et al., 2020). ViT는 이미지 인식 분야에서 현재 최고 수준의 결과를 보여주는 SOTA (state of the art) 모델로서 입력 이미지를 패치(patch) 단위로 나누고 패치 임베딩(Patch embedding)을 거쳐 결과를 도출하는 모델구조를 가진다. 또한 ViT 구조는 일반적인 CNN에 비해 더 작은 귀납적 편향을 가진다(Dosovitskiy et al., 2020; Hatamizadeh et al., 2023; Li et al., 2022). 귀납적 편향이 적다는 것은 이미지 전체에서 정보를 찾을 수 있다는 것을 의미지만 지역해를 찾기 위한 공간도 커지기 때문에 자료만 충분하다면 기존 대비 높은 성능을 발휘하는 장점을 가질 수 있다.

CNN 모델은 2차원 구조의 인식 및 분류에 적합한 컨볼루

Table 1. Hyperparameters for deep learning training

	Proposed model (ViT-Base Model)
Splitting data (Raw data)	Train dataset (60%) : Validation dataset (10%) : Test dataset (30%)
Input shape	(224, 224, 3)
Epoch	30
Optimization	Adaptive moment estimation (ADAM)
Initial learning rate	0.00001
Mini-batch size	16
Validation frequency	Every 1 epoch
Number of parameters (Million)	85.82

선 신경망 기반의 모델이며, 주로 이미지, 비디오, 텍스트 등의 다양한 영역에서 특징 추출 및 분류에 적용되고 있다(Kim et al., 2020). 수중음향 분야에서는 스펙트로그램(spectrogram), 로파/데몬 그램(lofar/demon grams), bearing time recorder와 같은 이미지 분류에 많이 활용되었으며, 특히 PAM을 활용하는 해양포유류 신호 탐지 및 식별 연구에도 스펙트로그램을 입력 데이터로 하여 CNN 기법을 많이 활용하고 있다(Kim et al., 2022).

따라서, 본 연구에서는 Vision Transformer-Base 16 model에 dense layer를 추가하여 CNN 모델을 구성하였다. Dense layer는 완전 연결층으로 모델에서 출력되어 나온 vector 특징을 명확히 하여 모델의 분류 성능 향상에 기여할 수 있다(Huang et al., 2017). CNN 모델 내 하이퍼파라미터의 경우 반복적으로 다양한 실험을 통해 획득한 최적의 파라미터를 지정하였으며, 최적화 함수는 ADAM (adaptive moment estimation) 알고리즘을 사용하였고, 학습율은 0.00001로 설정하고 다중 분류 손실함수를 활용하였다. 입력 자료의 크기는 249×666×3이지만 학습을 위하여 224×224×3형태로 변환하였으며, 제안된 CNN 모델은 한번의 epoch를 자료의 분할 없이 수행 가능하도록 하였다. 총 epoch는 30이며 1회의 iteration에 활용되는 batch size는 16으로 하였다. Epoch는 전체 데이터셋을 한 번 학습하는 과정을 의미하며, 일반적으로 한 번의 epoch에서 모든 데이터를 한꺼번에 처리하지 않고, mini-batch 단위로 나눠서 처리한다. Iteration은 하나의 mini-batch를 학습에 사용하는 과정을 의미하며, epoch내에서 mini-batch 개수에 따라 iteration횟수가 결정된다. 본 연구에서 제안된 CNN 모델의 세부 파라미터는 Table 1에 정리하였다.

### CNN 모델 기반 자동 분류 기법 적용 및 모델 정확도 평가 방법

3종의 돌고래 휘슬음은 3가지 방법으로 변환한 이미지를 모델에 적용하였다. 먼저, 종별로 획득한 돌고래 휘슬음을 스펙트로그램으로 변환한 원본 이미지를 학습에 활용하였다(Fig. 1a). 두번째로, 원본 이미지에 노이즈를 제거에 사용되는 미디언 필터(median filter)를 적용하여 필터링한 이미지를 학습에 사용

하였다(Fig. 1b). 마지막으로, 미디언 필터가 적용된 데이터에 추가적으로 특징 추출 방법 중 하나인 윤곽선 검출(edge detection)을 적용해서 주파수 대역에서 우세한 신호만을 추출한 이미지를 학습에 활용하였다(Fig. 1c). 동일한 조건을 유지하기 위해 모든 이미지의 크기는 통일하여 사용하였다. 고래 휘슬음 자료의 전처리 과정 및 자료 증강 방법, CNN 모델 적용 기법의 상세 모식도는 Fig. 2에 나타내었다.

CNN 모델의 정확도 평가를 위해서는 일반적으로 AI 및 통계학적 방법이 사용된 분류 모델에서 알고리즘의 성능을 시각화하기 위해 일반적으로 혼동행렬(confusion matrix)을 활용한다(Table 2). 이를 통해 정확도(accuracy), 정밀도(precision), 재현율(recall) 및 F1-Score를 도출하여 비교 평가한다(Grandini et al., 2020).

$$Accuracy = \frac{|TP|+|TN|}{|TP|+|TN|+|FP|+|FN|}$$

$$Precision = \frac{|TP|}{|TP|+|FP|}$$

$$Recall = \frac{|TP|}{|TP|+|FN|}$$

$$F1-Score = 2 * \frac{Precision \times Recall}{Precision + Recall}$$

Table 2. Confusion matrix for multi-class classification

Confusion matrix	Predictive values			
	Case 1	Case 2	Case 3	
Actual values	Case 1	TP (Case 1) FN (Case 1)	FP (Case 2) FN (Case 1)	FP (Case 3) FN (Case 1)
	Case 2	FP (Case 1) FN (Case 2)	TP (Case 2)	FP (Case 3) FN (Case 2)
	Case 3	FP (Case 1) FN (Case 3)	FP (Case 2) FN (Case 3)	TP (Case 3)

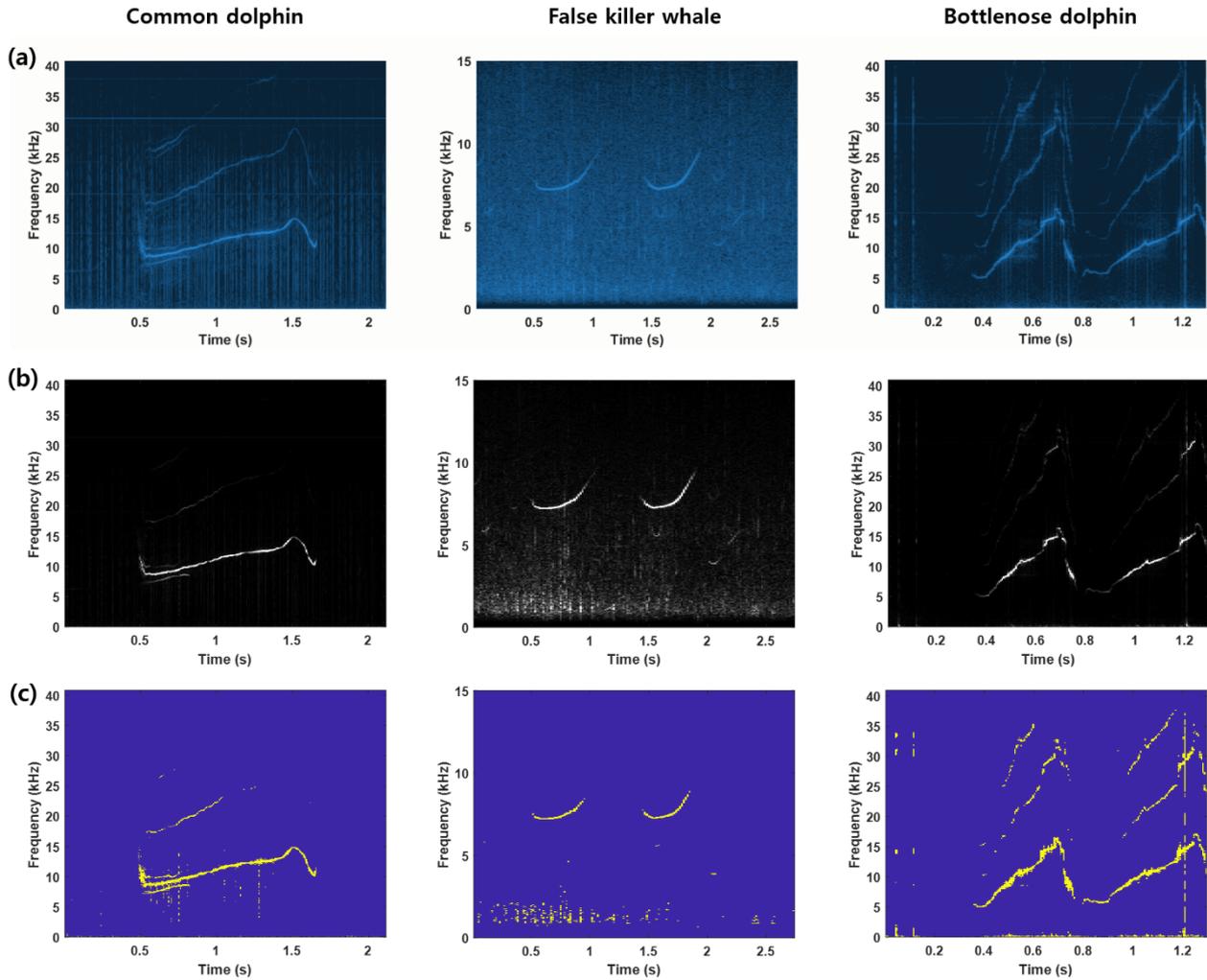


Fig. 1. (a) Images of Spectrogram, (b) images with median-filtering applied, and (c) images obtained through feature extraction method (edge detection) of common dolphins *Delphinus delphis*, false killer whales *Pseudorca crassidens*, and bottlenose dolphins *Tursiops truncatus* living in the oceans near the Korean Peninsula.

위 식에서 정확도는 분류 모델의 성능을 평가하는 가장 기본적인 지표 중 하나로, 전체 데이터 중에서 모델이 올바르게 예측한 비율을 나타낸다. 여기서 true positives는 모델이 실제로 긍정인 경우를 긍정으로 예측한 수를 의미하고, true negatives는 모델이 실제로 부정인 경우를 부정으로 예측한 수를 의미한다. 정확도는 모델이 얼마나 많은 경우에서 올바르게 예측했는지를 보여주지만, 데이터의 클래스 불균형이 있을 경우 성능을 적절히 평가하지 못하는 단점이 존재한다.

정밀도는 모델이 긍정 클래스로 예측한 결과 중에서 실제로 긍정인 경우의 비율을 나타낸다. 여기서 false positives (FP)는 실제로는 부정인 경우를 모델이 긍정으로 잘못 예측한 수를 의미합니다. 정밀도는 모델이 긍정 클래스에 대해 얼마나 정확하게 예측했는지를 평가하는 지표로, 특히 FP가 중요한 상황에서 유용하다. 재현율은 실제 긍정 클래스 중에서 모델이 이를 얼마

나 잘 찾아냈는지를 나타내는 비율이다. 여기서 false negatives (FN)는 실제로 긍정인 경우를 모델이 부정으로 잘못 예측한 수를 의미하며, 재현율은 모델이 긍정 클래스를 놓치지 않고 얼마나 잘 탐지하는지를 평가하는 지표로, FN를 최소화하는 것이 중요한 상황에서 매우 중요하다.

또한, 정확도, 정밀도, 재현율과 함께 F1-score도 중요한 평가 지표로 사용된다. F1-score는 정밀도와 재현율의 조화 평균(harmonic mean)으로, 이 두 지표의 균형을 평가하는 데 유용하다. 특히, F1-score는 데이터의 클래스 간 불균형이 있을 때 더욱 중요하게 고려되며, 높은 정밀도와 재현율을 모두 유지하기 어려운 상황에서 F1-score는 이들 간의 균형을 잘 반영하여 모델의 성능을 평가할 수 있다. F1-score가 1에 가까울수록 모델이 잘 작동한다고 볼 수 있으며, 0에 가까울수록 성능이 떨어지는 것으로 평가된다.

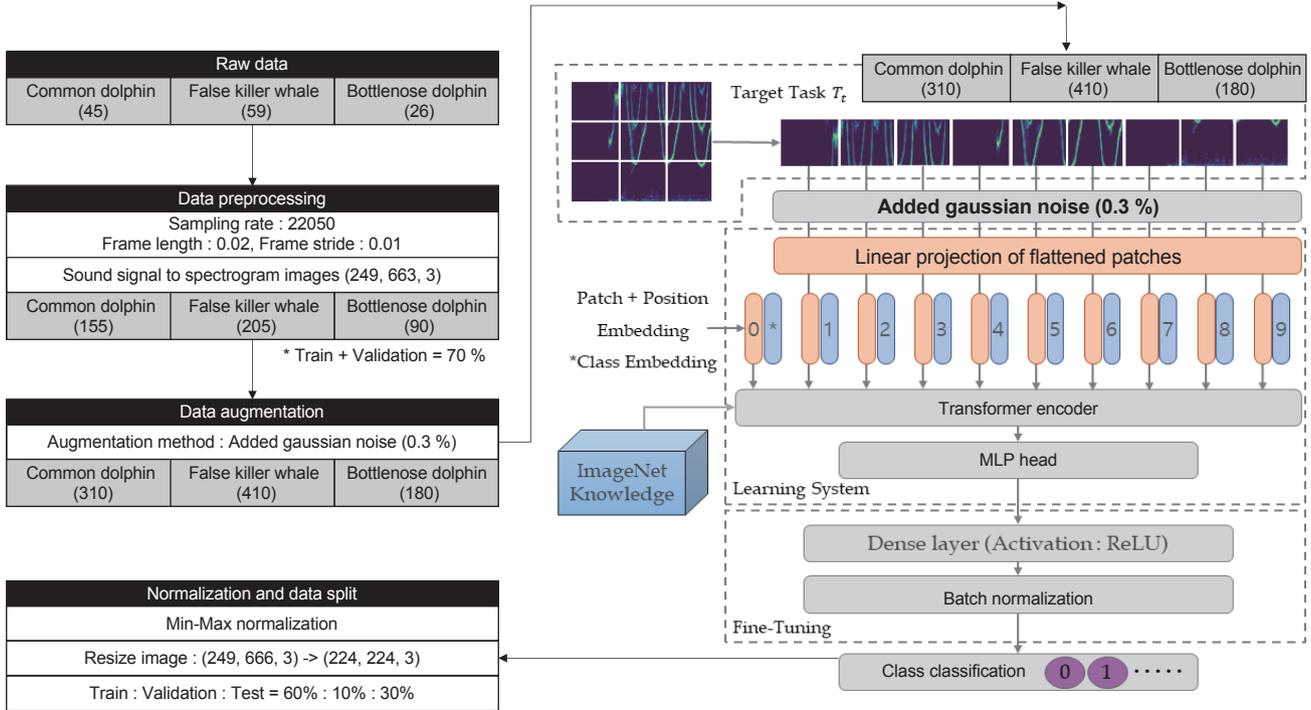


Fig. 2. A schematic diagram of the automatic classification method for three whale species using a convolution neural network (CNN) model.

**결 과**

본 연구에서는 한반도 인근 해역에 서식하는 참돌고래, 흑범고래, 큰돌고래의 휘슬음 신호를 스펙트로그램 이미지로 변환하여, CNN 기반의 제안 모델을 이용하여 돌고래 휘슬음 자동 분류를 수행하였다. 또한 3종의 돌고래 휘슬음 이미지는 스펙트로그램으로 변환한 원본 이미지(Case 1), 미디언 필터를 적용한 이미지(Case 2), 윤곽선 검출 기법을 적용한 이미지(Case 3)로

구성하여 입력 데이터에 따라 자동 분류를 수행하였다. 고래 종별 평가 데이터셋의 혼동행렬을 확인한 결과(Fig. 3), Case 3을 제외하고 전반적으로 정확하게 자동 식별되었다. 참돌고래와 큰돌고래는 Case 1, 2번에서 성능이 높게 나타났으며, 흑범고래만 Case 3에서 성능이 더 높게 나타났다. 또한 모든 케이스에서 흑범고래와 큰돌고래는 참돌고래로 잘못 식별되는 경우가 발생하였다. 각 Case별 각각의 정확도, 정밀도, 재현율, F1-score를 산출하였다(Table 3). 스펙트로그램 이미지 원본으로 학습한

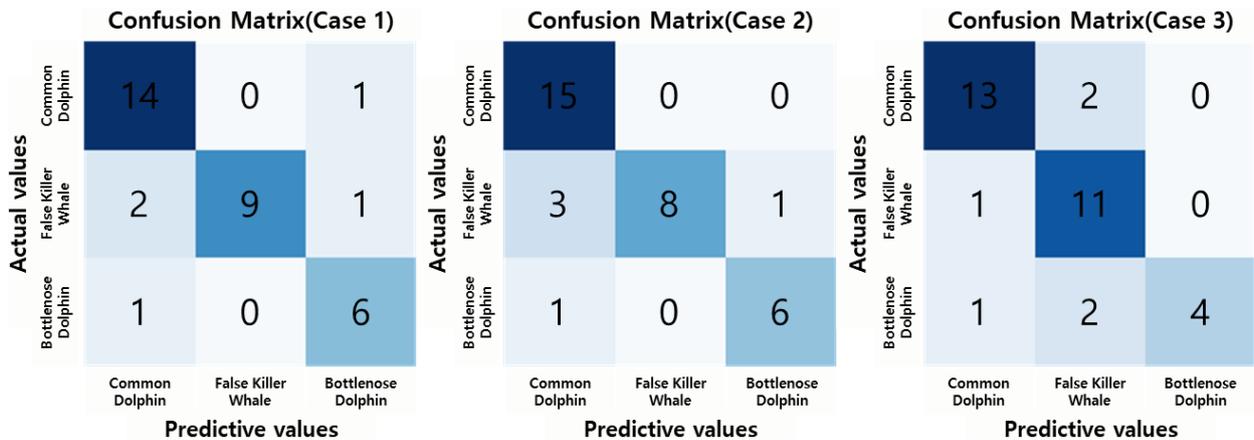


Fig. 3. Confusion matrix of predictive values with raw data (Case 1), median filtering data (Case 2), and edge detection data (Case 3).

Table 3. The accuracy, precision, recall, and F1-score of the spectrogram image, median filtered image, and edge detection image (Case 1–3)

	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Case 1	85.29	85.78	84.68	84.40
Case 2	85.29	88.22	84.13	84.65
Case 3	82.36	86.67	78.49	80.29

Table 4. The accuracy, precision, recall, and F1-score of the common dolphin *Delphinus delphis*, false killer whale *Pseudorca crassidens*, and bottlenose dolphin *Tursiops truncatus* (only Case 2)

Case 2	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Common dolphin	100.00	83.33	100.00	90.90
False killer whale	66.67	100.00	66.67	80.00
Bottlenose dolphin	85.71	85.71	85.71	85.71

모델에 대한 평가 성능 지표는 정확도 85.29%, 정밀도 85.78%, 재현율 84.68%, F1-score 84.40%로 나타났다. 미디언 필터를 적용한 이미지로 학습한 모델에 대한 평가 성능 지표는 정확도 85.29%, 정밀도 88.22%, 재현율 84.13%, F1-score 84.65%로 나타나, 스펙트로그램 이미지 원본으로 학습한 모델에 비해 전반적으로 소폭 상승하였다. 반면에 윤곽선 검출 기법을 적용한 이미지의 성능 지표는 정확도 82.36%, 정밀도 86.67%, 재현율 78.49%, F1-score 80.29%로 3개의 이미지 중 가장 성능이 낮았다.

우수한 성능을 보인 미디언 필터를 적용한 모델(Case 2)에서 고래 종에 따른 정확도, 정밀도, 재현율, F1 점수를 산출하였다(Table 4). 참돌고래의 성능 지표는 정확도 및 F1-score가 각각 100.00%, 90.90%로 나타났으며, 정확도랑 재현율이 높게 나타나 실제 참돌고래 신호를 정확하게 식별하였다. 반면에 흑범고래의 성능 지표는 정확도 및 재현율이 66.67%로 다른 두 고래에 비해 낮게 나타났으며, 정밀도는 더 높아 다른 고래의 신호를 흑범고래로 오 식별하는 경우가 없었다. 큰돌고래의 성능 지표는 정확도, 정밀도, 재현율, F1-score가 85.71%로 동일하게 나타났다. 종합적으로 정확도와 F1-score는 각 고래 종류마다 데이터의 불균형 정도에 따라 차이를 보였으나, 돌고래 휘슬음 신호를 탐지하고 자동 분류함이 가능함을 확인하였다.

## 고 찰

최근에는 해양 생태계를 파악하기 위한 방법으로 PAM 기법을 통해 수집된 대용량 수중음향 자료에서 해양 포유류의 소리를 탐지하고 식별하기 위하여 CNN 기법을 활용한 연구가 많이

활용되고 있다(Lindseth and Lobel, 2018). 본 연구에서는 국내에서 관찰되는 돌고래 3종(참돌고래, 흑범고래, 큰돌고래)에 대한 휘슬음 데이터를 CNN 모델을 이용하여 자동 신호 분류를 수행하였다. 또한 스펙트로그램 기반의 3가지 이미지를 학습시켜 모델 간의 성능을 비교하였으며, 제안된 CNN 모델은 다양한 성능 지표에서 우수한 결과가 나타나 돌고래 3종의 자동 분류의 적용 가능성을 확인하였다.

분석 결과 성능지표는 Case 2에서 정확도 85.29%, F1-score 84.65%로 Case1과 차이는 크지 않았지만 가장 높게 나타났다. CNN 모델은 입력 데이터의 특징을 추출하여 패턴을 학습하므로, 신호처리 과정을 통해 입력 데이터의 노이즈를 제거하면 모델의 성능이 더 높아질 것으로 예상하였다. 그러나 본 연구 결과에서는 윤곽선 검출 기법까지 활용한 Case 3의 성능은 원본 이미지보다 더 낮은 성능이 나타났다. 이는 윤곽선 검출 기법이 과도하게 노이즈를 제거하며 각각의 돌고래가 가지는 휘슬음의 특징도 같이 약해져 상대적으로 성능이 낮아 지거나, 윤곽선 검출 기법에 저주파의 일부 강한 노이즈가 포함되어 신호의 특징과 혼동되었을 수도 있으나 정확한 원인에 대해서는 추가적인 검토가 필요하다. 고래 종에 따른 성능 지표는 참돌고래, 큰돌고래, 흑범고래 순서로 나타났다. 특히, 참돌고래는 학습 데이터의 양도 많았지만 다른 돌고래의 주파수 대역보다 휘슬음 신호가 상대적으로 길게 나타나서 성능지표가 높게 나타난 것으로 생각된다.

스크립스 해양 연구소를 비롯한 선진 해양 연구기관에서는 해양포유류를 추적하고 식별하기 위하여, PAM을 통해 수천만개 이상의 고래류 음향 신호를 구축하여 머신 러닝 기반의 알고리즘이 개발되어 적용하고 있다(Shamir et al., 2014; Usman et al., 2020). 국내에서는 고래류의 음향 신호에 대한 데이터가 충분하지 않아, 인공지능 기법을 활용한 자동 탐지에 대한 연구만 수행되었으며 고래류 자동 분류에 적용한 사례는 없었다. 본 연구에서도 웹에서 공개되어 있는 자료를 활용하였으나 제한적인 자료를 이용하여 증강 기법을 이용하여 학습 및 분류를 수행하였다. 향후에는 국내 연안 해역에 서식하는 고래 소리를 측정하여 데이터를 추가적으로 확보할 계획이며, 지속적인 연구를 통해 고래류 종에 대한 확장 탐지 및 분류의 정확성을 증가시킬 예정이다. 본 연구를 기반으로 고래류에 대해 PAM 기법이 더 효율적으로 활용된다면 종 분류, 서식 해역, 분포 특성, 자원량 추정 등 고래류 생태 특성에 많은 정보를 제공할 것으로 판단된다.

## 사 사

본 연구는 2024년 정부(방위사업청)의 재원으로 국방기술진흥연구소(KRIT)의 지원을 받아 수행된 연구임(KRIT-CT-22-056, 과제명: 해양생물음 재현 알고리즘 및 신호 모델링 기법 연구).

## References

- Allen AN, Harvey M, Harrell L, Jansen A, Merkens KP, Wall CC, Cattiau J and Oleson EM. 2021. A convolutional neural network for automated detection of humpback whale song in a diverse, long-term passive acoustic dataset. *Front Mar Sci* 8, 607321. <https://doi.org/10.3389/fmars.2021.607321>.
- André M, Van Der Schaar M, Zaugg S, Houégnigan L, Sánchez AM and Castell JV. 2011. Listening to the deep: Live monitoring of ocean noise and cetacean acoustic signals. *Mar Pollut Bull* 63, 18-26. <https://doi.org/10.1016/j.marpolbul.2011.04.038>.
- Antichi S, Austin M, May-Collado LJ, Urbán R J, Martínez-Aguilar S and Vilorio-Gómora L. 2023. Differences in the whistles of two ecotypes of bottlenose dolphins from the Gulf of California. *JASA Express Lett* 3, 051201. <https://doi.org/10.1121/10.0019502>.
- Bergler C, Smele SQ, Tyndel SA, Barnhill A, Ortiz ST, Kalan AK, Cheng RX, Brinklov S, Osiecka AN, Tougaard J, Jakobsen F, Wahlberg M, Noth E, Maier A and Klump BC. 2022. ANIMAL-SPOT enables animal-independent signal detection and classification using deep learning. *Sci Rep* 12, 21966. <https://doi.org/10.1038/s41598-022-26429-y>.
- Bittle M and Duncan A. 2013. A review of current marine mammal detection and classification algorithms for use in automated passive acoustic monitoring. In: *Proceedings of Acoustics*. McMinn T, ed. Victor Harbor, Australia, 8.
- Burtenshaw JC, Oleson EM, Hildebrand JA, McDonald MA, Andrew RK, Howe BM and Mercer JA. 2004. Acoustic and satellite remote sensing of blue whale seasonality and habitat in the Northeast Pacific. *Deep Sea Res II Top Stud Oceanogr* 51, 967-986. <https://doi.org/10.1016/j.dsr2.2004.06.020>.
- Carlón-Beltrán Ó, Vilorio-Gómora L, Urbán J, Martínez-Aguilar S and Antichi S. 2023. Whistle characterization of long-beaked common dolphin (*Delphinus delphis bairdii*) in La Paz Bay, Gulf of California. *PeerJ* 11, e15687. <https://doi.org/10.7717/peerj.15687>.
- Caruso F, Dong L, Lin M, Liu M, Gong Z, Xu W, Alonge G and Li S. 2020. Monitoring of a nearshore small dolphin species using passive acoustic platforms and supervised machine learning techniques. *Front Mar Sci* 7, 267. <https://doi.org/10.3389/fmars.2020.00267>.
- Choi KH, Yoon YG, Kim S, Kim H, Cho JW, Bae HS and Park K. 2019. Analysis of echolocation click signals of Indo-Pacific bottlenose dolphin (*Tursiops aduncus*) in Jeju Island. *J Acoust Soc Kr* 38, 56-65. <https://doi.org/10.7776/ASK.2019.38.1.056>.
- Choi S, Yoon YG, Sohn H, Lee K and Park KJ. 2017. Whistle characteristics of captive beluga whale *Delphinapterus leucas*. *Korean J Fish Aquat Sci* 50, 799-805. <https://doi.org/10.5657/KFAS.2017.0799>.
- Clark CW, Borsani JF and Notarbartolo-Di-sciara G. 2002. Vocal activity of fin whales, *Balaenoptera physalus*, in the Ligurian Sea. *Mar Mamm Sci* 18, 286-295. <https://doi.org/10.1111/j.1748-7692.2002.tb01035.x>.
- Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J and Houlsby N. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *ICLR 2020*, 1-22. <https://doi.org/10.48550/arXiv.2010.11929>.
- Frainer G, Dufourq E, Fearey J, Dines S, Probert R, Elwen S and Gridley T. 2023. Automatic detection and taxonomic identification of dolphin vocalisations using convolutional neural networks for passive acoustic monitoring. *Eco Inform* 78, 102291. <https://doi.org/10.1016/j.ecoinf.2023.102291>.
- Gibb R, Browning E, Glover-Kapfer P and Jones KE. 2019. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods Ecol Evol* 10, 169-185. <https://doi.org/10.1111/2041-210X.13101>.
- Grandini M, Bagli E and Visanim G. 2020. Metrics for multi-class classification: An overview. *arXiv*, 2008.05756v1. <https://doi.org/10.48550/arXiv.2008.05756>.
- Gridley T, Martin MJ, Slater J, Roux JP, Swift RJ and Elwen SH. 2020. Towed passive acoustic monitoring complements visual survey methods for Heaviside's dolphins *Cephalorhynchus heavisidii* in the Namibian Islands Marine Protected Area. *Afr J Mar Sci* 42, 495-506. <https://doi.org/10.2989/1814232X.2020.1848925>.
- Hatamizadeh A, Yin H, Heinrich G, Kautz J and Molchanov P. 2023. Global context vision transformers. In: *Proceedings of the 40th International Conference on Machine Learning*. PMLR 202, 12633-12646. <https://doi.org/10.48550/arXiv.2010.11929>.
- Huang G, Liu Z, Van Der Maaten L and Weinberger KQ. 2017. Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, U.S.A., 4700-4708. <https://doi.org/10.1109/CVPR.2017.243>.
- Jeong S, Chung W, Shin S, Kim D, Kim J, Byun G and Lee D. 2023. Development of dolphin click signal classification algorithm based on recurrent neural network for marine environment monitoring. *Geophys Geophys Explor* 26, 126-137. <https://doi.org/10.7582/GGE.2023.26.3.126>.
- Kim H, Jung J, Kang D, Simanungkalit F and Kang M. 2024. Occurrence and acoustic characteristics of finless porpoise (*Neophocaena asiaorientalis*) off the south coast of Korea using hydrophone and visual surveys. *J Korean Soc Fish Ocean Technol* 60, 258-268. <https://doi.org/10.3796/KS-FOT.2024.60.3.258>.
- Kim JS, Yoon YG, Han DG, La HS and Choi JW. 2022. Classification of bearded seals signal based on convolutional neural network. *J Acoust Soc Kr* 41, 235-241. <https://doi.org/10.7776/ASK.2022.41.2.235>.

- Kim S, Jung S, Kang D, Kim M and Cho S. 2020. Application of the artificial intelligence for automatic detection of shipping noise in shallow-water. *J Acoust Soc Kr* 39, 279-285. <https://doi.org/10.7776/ASK.2020.39.4.279>.
- Küsel ET, Siderius M and Mellinger DK. 2016. Single-sensor, cue-counting population density estimation: Average probability of detection of broadband clicks. *J Acoust Soc Am* 140, 1894-1903. <https://doi.org/10.1121/1.4962753>.
- Lee JH, Kim EH, Lee K, Park KJ, An YR, Kim HW, Sohn H and Choi SG. 2022. Occurrence and spatial distribution of marine mammals by sighting surveys in Korean waters during 2011-2020. *Korean J Fish Aquat Sci* 55, 938-945. <https://doi.org/10.5657/KFAS.2022.0938>.
- Lindseth A and Lobel P. 2018. Underwater soundscape monitoring and fish bioacoustics: A review. *Fishes* 3, 36. <https://doi.org/10.3390/fishes3030036>.
- Li Y, Mao H, Girshick R and He K. 2022. Exploring plain vision transformer backbones for object detection. In: *Computer Vision – ECCV 2022*. ECCV 2022. Lecture Notes in Computer Science, vol 13669. Avidan S, Brostow G, Cissé M, Farinella GM and Hassner T, eds. Springer, Cham, Switzerland. [https://doi.org/10.1007/978-3-031-20077-9\\_17](https://doi.org/10.1007/978-3-031-20077-9_17).
- Marques TA, Thomas L, Martin SW, Mellinger DK, Ward JA, Moretti DJ, Harris D and Tyack PL. 2013. Estimating animal population density using passive acoustics. *Biol Rev* 88, 287-309. <https://doi.org/10.1111/brv.12001>.
- McCullough JLK, Simonis AE, Sakai T and Oleson EM. 2021. Acoustic classification of false killer whales in the Hawaiian Islands based on comprehensive vocal repertoire. *JASA Express Lett* 1, 071201. <https://doi.org/10.1121/10.0005512>.
- McDonald MA and Fox CG. 1999. Passive acoustic methods applied to fin whale population density estimation. *J Acoust Soc Am* 105, 2643-2651. <https://doi.org/10.1121/1.426880>.
- Mellinger DK, Stafford KM, Moore SE, Dziak RP and Matsumoto H. 2007. An overview of fixed passive acoustic observation methods for cetaceans. *Oceanography* 20, 36-45. <https://doi.org/10.5670/oceanog.2007.03>.
- Munger LM, Wiggins SM, Moore SE and Hildebrand JA. 2008. North Pacific right whale (*Eubalaena japonica*) seasonal and diel calling patterns from long-term acoustic recordings in the southeastern Bering Sea, 2000-2006. *Mar Mamm Sci* 24, 795-814. <https://doi.org/10.1111/j.1748-7692.2008.00219.x>.
- Oleson EM, Calambokidis J, Burgess WC, McDonald MA, LeDuc CA and Hildebrand JA. 2007. Behavioral context of call production by eastern North Pacific blue whales. *Mar Ecol Prog Ser* 330, 269-284. <https://doi.org/10.3354/meps330269>.
- Parsons ECM, Baulch S, Bechshoft T, Bellazzi G, Bouchet P, Cosentino AM, Godard-Codding CAJ, Gulland F, Hoffmann-Kuhnt M, Hoyt E, Livermore S, MacLeod CD, Matrai E, Munger L, Ochiai M, Peyman A, Recalde-Salas A, Regnery R, Rojas-Bracho L, Salgado-Kent CP, Slooten E, Wang JY, Wilson SC, Wright AJ, Young S, Zwamborn E and Sutherland WJ. 2015. Key research questions of global importance for cetacean conservation. *Endanger Species Res* 27, 113-118. <https://doi.org/10.3354/esr00655>.
- Putland RL, Ranjard L, Constantine R and Radford CA. 2018. A hidden Markov model approach to indicate Bryde's whale acoustics. *Ecol Indic* 84, 479-487. <https://doi.org/10.1016/j.ecolind.2017.09.025>.
- Shamir L, Yerby C, Simpson R, von Benda-Beckmann AM, Tyack P, Samarra F, Miller P and Wallin J. 2014. Classification of large acoustic datasets using machine learning and crowdsourcing: Application to whale calls. *J Acoust Soc Am* 135, 953-962. <https://doi.org/10.1121/1.4861348>.
- Shin HL, Seo DO, Lee DJ, Hwang DJ, Bae MK and Lee YW. 2002. Characteristics of bottlenose dolphin (*Tursiops truncatus*) whistle. *J Kor Soc Fish Technol* 38, 271-277.
- Širović A, Hildebrand JA, Wiggins SM, Moore SE, McDonald MA and Thiele D. 2003. Seasonality of blue and fin whale calls west of the Antarctic Peninsula. In: *Proceedings of the Oceans 2003*. San Diego, CA, U.S.A., 818. <https://doi.org/10.1109/OCEANS.2003.178423>.
- Stafford KM, Nieuwkerk SL and Fox CG. 1999. An acoustic link between blue whales in the eastern tropical Pacific and the Northeast Pacific. *Mar Mamm Sci* 15, 1258-1268. <https://doi.org/10.1111/j.1748-7692.1999.tb00889.x>.
- Usman AM, Ogundile OO and Versfeld DJ. 2020. Review of automatic detection and classification techniques for cetacean vocalization. *IEEE Access* 8, 105181-105206. <https://doi.org/10.1109/ACCESS.2020.3000477>.
- Weinstein BG. 2018. A computer vision for animal ecology. *J Anim Ecol* 87, 533-545. <https://doi.org/10.1111/1365-2656.12780>.
- White EL, White PR, Bull JM, Risch D, Beck S and Edwards EW. 2022. More than a whistle: Automated detection of marine sound sources with a convolutional neural network. *Front Mar Sci* 9, 879145. <https://doi.org/10.3389/fmars.2022.879145>.
- Winn HE and Winn LK. 1978. The song of the humpback whale (*Megaptera novaeangliae*) in the West Indies. *Mar Biol* 47, 97-114. <https://doi.org/10.1007/BF00395631>.
- Yoon YG, Sohn H, Park KJ, Choi YM, Kim S and Choi JW. 2017. Study of acoustic characteristics of common dolphins *Delphinus delphis* in the East Sea. *Korean J Fish Aquat Sci* 50, 406-412. <https://doi.org/10.5657/KFAS.2017.0406>.
- Yoon YG, Yang W and Choi JW. 2021. Preliminary study of passive acoustic monitoring of finless porpoises *Neophocaena asiaeorientalis* around the Southwest offshore wind farm in Korea. *J Acoust Soc Kr* 40, 537-545. <https://doi.org/10.7776/ASK.2021.40.6.537>.
- Ziegenhorn MA, Frasier KE, Hildebrand JA, Oleson EM, Baird RW, Wiggins SM and Baumann-Pickering S. 2022. Dis-

criminating and classifying odontocete echolocation clicks in the Hawaiian Islands using machine learning methods. PLoS One 17, e0266424. <https://doi.org/10.1371/journal.pone.0266424>.

Zimmer WMX. 2011. Passive acoustic monitoring of cetaceans. Cambridge University Press, Cambridge, U.K.