Article

Classification of Hydroacoustic Signals Based on Harmonic Wavelets and a Deep Learning Artificial Intelligence System

Dmitry Kaplun [1] , **Alexander Voznesensky** [1] , **Sergei Romanov** [1] , **Valery Andreev** [2] **and Denis Butusov** [3,*]

[1] Department of Automation and Control Processes, Saint Petersburg Electrotechnical University "LETI", Saint Petersburg 197376, Russia; dikaplun@etu.ru (D.K.); a-voznesensky@yandex.ru (A.V.); saromanov@etu.ru (S.R.)

[2] Department of Computer-Aided Design, Saint Petersburg Electrotechnical University "LETI", Saint Petersburg 197376, Russia; vsandreev@etu.ru

[3] Youth Research Institute, Saint Petersburg Electrotechnical University "LETI", Saint Petersburg 197376, Russia

[*] Correspondence: dnbutusov@etu.ru; Tel.: +7-950-008-7190

check for updates

Abstract: This paper considers two approaches to hydroacoustic signal classification, taking the sounds made by whales as an example: a method based on harmonic wavelets and a technique involving deep learning neural networks. The study deals with the classification of hydroacoustic signals using coefficients of the harmonic wavelet transform (fast computation), short-time Fourier transform (spectrogram) and Fourier transform using a kNN-algorithm. Classification quality metrics (precision, recall and accuracy) are given for different signal-to-noise ratios. ROC curves were also obtained. The use of the deep neural network for classification of whales' sounds is considered. The effectiveness of using harmonic wavelets for the classification of complex non-stationary signals is proved. A technique to reduce the feature space dimension using a 'modulo N reduction' method is proposed. A classification of 26 individual whales from the Whale FM Project dataset is presented. It is shown that the deep-learning-based approach provides the best result for the Whale FM Project dataset both for whale types and individuals.

Keywords: harmonic wavelets; classification; kNN-algorithm; deep neural networks; machine learning; Fourier transform; short-time Fourier transform; wavelet transform; spectrogram; confusion matrix; ROC curve

1. Introduction

The whale was one of the main commercial animals in the past. Whalers were attracted by the huge carcass of this animal—from one whale they could get much more fat and meat than from any other marine animal. Today, many of its species have almost been driven to extinction. For this reason, they are listed in the IUCN Red List of Threatened Species [1]. Currently, the main threat to whales is an anthropogenic factor, expressed in violation of their usual way of life and pollution of the seas. To ensure the safety of rare animals, the number of individuals must be monitored. Within the framework of environmental monitoring programs approved by governments and public organizations of different countries, cetacean monitoring activities are carried out year-round using all of the modern achievements in data processing [2]. Monitoring includes work at sea and post-processing of the collected data: determining the coordinates of whale encounters, establishing the

composition of the group, and photographing the animals for subsequent observation of individually recognizable individuals.

Systematic observation of animals presents scientists with the opportunity to learn about how mammals share the water area among themselves, to collect data on age and gender composition [3]. An important task is to find out where the whales come from and where they then go to in the winter, to track their routes of movements. You must also be able to determine which population the whales belong to.

Sounds made by cetaceans for communication are called "whale songs". The word "songs" is used to emphasize the repeating and melodic nature of these sounds, reminiscent of human singing. The use of sounds as the main communication channel is due to the fact that, in an aquatic environment, visibility can be limited, and smells spread much slower than in air [4]. It is believed that the most complex songs of humpback whales and some toothless whales are used in mating games. Simpler signals are used all year round and perhaps serve for day-to-day communication and navigation. Toothed whales (including killer whales) use emitted sounds for echolocation. In addition, it was found that whales that have lived in captivity for long can mimic human speech. All these signals are transmitted to different distances, under different water conditions and in the presence of a variety of noises. Additionally, stable flocks have their own dialects, i.e., there is wide variability in the sounds made by whales, both within the population and between populations. Thus, sounds can be used to classify both whale species and individuals. The task of classifying whales by sound has been solved by many researchers for different types of whales in different parts of the world, using various methods and approaches, the most popular being signal processing algorithms [5,6] and algorithms based on neural networks [2,7–10]. Neural-network-based approaches present different architectures, models and learning methods. In [2], the authors developed and empirically studied a variety of deep neural networks to detect the vocalizations of endangered North Atlantic right whales. In [7], an effective data-driven approach based on pre-trained convolutional neural networks (CNN) using multi-scale waveforms and time-frequency feature representations was developed in order to perform classification of whale calls from a large open-source dataset recorded by sensors carried by whales. The authors of [8] constructed an ensembled deep learning CNN model to classify beluga detections. The applicability of basic CNN models is also being explored for the bio-acoustic task of whale call detection, such as with respect to North Atlantic right whale calls [9] and humpback whale calls [10].

This paper considers two approaches to hydroacoustic classification, taking the sounds made by whales as examples: on the basis of harmonic wavelets and deep learning neural networks. The main contributions of our work can be summarized as follows. The effectiveness of using harmonic wavelets for the classification of hydroacoustic signals was proved. A technique to reduce the feature space dimension using a 'modulo N reduction' method was developed. A classification of 26 individual whales is presented for the dataset. It was shown that the deep-learning-based approach provides the best result for the dataset both for whale types and individuals.

The remainder of this paper is organized as follows. In Section 2, we briefly describe hydroacoustic signal processing and review related works on it. In Section 3, we introduce details of the harmonic wavelets and their application to the processing of hydroacoustic signals. In Section 4, we review the kNN algorithm for classification based on harmonic wavelets and present experimental results to verify the proposed approach. In Section 5, experimental results are presented to verify the approach for classification based on neural networks and machine learning. In Section 6, we discuss the results and how they can be interpreted from the perspective of previous studies and of the working hypotheses. Future research directions also are highlighted. Finally, we present the conclusions in Section 7.

2. Hydroacoustic Signal Processing

Before classifying hydroacoustic signals, which are sounds made by whales in an aquatic environment, they must be pre-processed, as the quality of the classification will depend on their quality. Hydroacoustic signal processing includes data preparation, as well as the use of further

algorithms allowing the extraction of useful signals from certain directions. Preliminary processing includes de-noising, estimation of the degree of randomness, extraction of short-term local features, pre-filtering, etc. Preprocessing affects the process of further analysis within a hydroacoustic monitoring system [11–13]. Even though the preprocessing of hydroacoustic signals has been studied for a long time, there are several unresolved problems, namely: working in conditions of a priori uncertainty of signal parameters; processing complex non-stationary hydroacoustic signals with multiple local features; and analysis of multicomponent signals. Another set of problems is represented by effective preliminary visual processing of hydroacoustic signals and the need for a mathematical apparatus for signal preprocessing tasks.

Current advances in applied mathematics and digital signal processing along with the development of high-performance hardware allow the effective application of numerous mathematical techniques, including continuous and discrete wavelet transforms. Wavelets are an effective tool for signal preprocessing, due to their adaptability, the availability of fast computational algorithms and the diversity of wavelet bases.

Using wavelets for hydroacoustic signal analysis provides the following possibilities [14,15]:

1. Detection of foreign objects in marine and river areas, including icebergs and other ice formations, the size estimation of these objects, hazard assessment based on analyzing local signal features;
2. Detection and classification of marine targets based on the analysis of local signal features;
3. Detection of hydroacoustic signals in the presence of background noise;
4. Efficient visualization and processing of hydroacoustic signals based on multiscale wavelet spectrograms.

Classification is an important task of modern signal processing. The quality of the classification depends on the noise level, training size and testing datasets, and the algorithm. It is also important to choose classification features and determine the size of the feature space. The classification feature is the feature or characteristic of the object used for classification. If we classify real non-stationary signals, it is important to have informative classification features. Among such features are wavelet coefficients.

3. Harmonic Wavelets

Wavelet transform uses wavelets as the basis functions. An arbitrary function can be obtained from one function ("mother" wavelet) by using translations and dilations in the time domain. The wavelet transform is commonly used for analyzing non-stationary (seismic, biological, hydroacoustic etc.) signals, usually together with various spectral analysis algorithms [16,17].

Consider the basis of harmonic wavelets whose spectra are rectangular in the given frequency band [15,16]. Harmonic wavelets are usually represented in the frequency domain. Wavelet-function (mother wavelet) can be written as:

$$\Psi(\omega) = \begin{cases} \frac{1}{2\pi}, & 2\pi \leq \omega < 4\pi \\ 0, & \omega < 2\pi, \omega \geq 4\pi \end{cases} \Leftrightarrow \psi(x) = \int\limits_{-\infty}^{\infty} \Psi(\omega)e^{i\omega x}d\omega = \frac{e^{i4\pi x} - e^{i2\pi x}}{i2\pi x} \tag{1}$$

There are some techniques that allow us to decompose input signals using different basic functions: wavelets, sine waves, damped sine waves, polynomials, etc. These functions form the atom dictionary (basis functions) and each function is localized in the time and frequency domains. Often the dictionary of atoms is full (all types of functions are used) and redundant (the functions are not mutually independent). One of the main problems in these techniques is the selection of basic functions and

dictionary optimization to acheive optimal decomposition levels [17]. Decomposition levels for wavelets can be defined as:

$$\Psi_{jk}(\omega) = \begin{cases} \frac{1}{2\pi} 2^{-j} e^{-\frac{i\omega k}{2^j}}, \ 2\pi 2^j \le \omega < 4\pi 2^j \\ 0, \ \omega < 2\pi 2^j, \omega \ge 4\pi 2^j \end{cases}$$

$$\psi_{jk}(x) = \psi(2^j x - k) = \int_{-\infty}^{\infty} \Psi_{jk}(\omega) e^{i\omega x} d\omega = \frac{e^{i4\pi(2^j x - k)} - e^{i2\pi(2^j x - k)}}{i2\pi(2^j x - k)} \tag{2}$$

where j is decomposition level and k is dilation.

Very often, wavelets are basis functions because of their useful properties [14] and the potential to process signals in the time-frequency domain. The Fourier transform of a scaling function can be written as:

$$\Phi(\omega) = \begin{cases} \frac{1}{2\pi}, \ 0 \le \omega < 2\pi \\ 0, \ \omega < 0, \omega \ge 2\pi \end{cases} \Leftrightarrow \phi(x) = \int_{-\infty}^{\infty} \Phi(\omega) e^{i\omega x} d\omega = \frac{e^{i2\pi x} - 1}{i2\pi x} \tag{3}$$

We can formulate the following properties of harmonic wavelets, which relate them with other classes of wavelets:

- Harmonic wavelets have compact support in the frequency domain, which can be used for localizing signal features.
- There are fast algorithms based on the fast Fourier transform (FFT) for computing wavelet coefficients and reconstructing signals in the time domain.

The drawback of harmonic wavelets is their weak localization properties in the time domain in comparison with other types of wavelets. The spectrum in the form of a rectangular wave leads to decay in the time domain as *1/x*, which is not sufficient for extracting short-term singularities in a signal in the time domain.

Wavelet Transform in the Basis of Harmonic Wavelets

Detailed coefficients $a_{jk}, \widetilde{a}_{jk}$ and approximation coefficients $a_{\phi k}, \widetilde{a}_{\phi k}$:

$$a_{jk} = 2^j \int_{-\infty}^{\infty} f(x) \overline{\psi}(2^j x - k) dx \quad \widetilde{a}_{jk} = 2^j \int_{-\infty}^{\infty} f(x) \psi(2^j x - k) dx$$

$$a_{\phi k} = 2^j \int_{-\infty}^{\infty} f(x) \overline{\phi}(x - k) dx \quad \widetilde{a}_{\phi k} = 2^j \int_{-\infty}^{\infty} f(x) \phi(x - k) dx \tag{4}$$

where j is the decomposition level; k is the dilation.

If *f(x)* is a real-valued function, then: $\widetilde{a}_{jk} = \overline{a}_{jk}, \widetilde{a}_{\phi k} = \overline{a}_{\phi k}$.

Wavelet decomposition [14]:

$$f(x) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} a_{jk} \psi(2^j x - k) = \sum_{k=-\infty}^{\infty} a_{\phi k} \phi(x - k) + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} a_{jk} \psi(2^j x - k) \tag{5}$$

Wavelet decomposition using harmonic wavelets [18]:

$$f(x) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \left[a_{jk} \psi\left(2^j x - k\right) = \widetilde{a}_{jk} \overline{\psi}\left(2^j x - k\right) \right]$$

$$= \sum_{k=-\infty}^{\infty} \left[a_{\phi k} \phi(x - k) = \widetilde{a}_{\phi k} \overline{\phi}(x - k) \right] + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} \left[a_{jk} \psi\left(2^j x - k\right) + \widetilde{a}_{jk} \overline{\psi}\left(2^j x - k\right) \right] \tag{6}$$

$$a_{jk} = 2^j \int_{-\infty}^{\infty} f(x) \overline{\psi}(2^j x - k) dx$$

Calculations with the last two formulae are inefficient.

Fast decomposition can be implemented in the following way:

$$a_{jk} = 2^j \int_{-\infty}^{\infty} F(\omega) \frac{1}{2\pi} 2^{-j} e^{\frac{i\omega k}{2^j}} d\omega = \frac{1}{2\pi} \int_{2\pi 2^j}^{4\pi 2^j} F(\omega) e^{\frac{i\omega k}{2^j}} d\omega \approx \int_{2\pi 2^j}^{4\pi 2^j} F(\omega) e^{\frac{i\omega k}{2^j}} d\omega \qquad (7)$$

The substitution is of the following form:

$$n = 2^j + s$$
$$F_{2^j+s} = 2\pi F\left[\omega = 2\pi\left(2^j + s\right)\right] \qquad (8)$$

We can show that:

$$a_{jk} = \sum_{s=0}^{2^j-1} F_{2^j+s} e^{\frac{i2\pi sk}{2^j}} \; k = 0\ldots 2^j - 1; j = 0\ldots n - 1. \qquad (9)$$

$$\widetilde{a}_{jk} = \sum_{s=0}^{2^j-1} F_{N-(2^j+s)} e^{\frac{i2\pi sk}{2^j}} \; k = 0\ldots 2^j - 1; \; j = 0\ldots n - 1.$$
$$\widetilde{a}_{jk} = \bar{a}_{jk} \qquad (10)$$

Thus, the algorithm for computing wavelet coefficients of the octave harmonic wavelet transform [19] of a continuous-time function $f(x)$ can be written in the following way:

1. The original function $f(x)$ is represented by discrete-time samples: $f(n)$, $n = 0 \ldots N{-}1$, where N is of degree 2 (if necessary, we use zero-padding).
2. We calculate the discrete Fourier transform using the fast Fourier transform to obtain a set of complex numbers $f(n)$, $n = 0 \ldots N{-}1$—Fourier coefficients (DFT coefficients).
3. Octave blocks F are processed using the discrete Fourier transform (DFT) to obtain coefficients: $a_{jk} = a_{2^j+k}$. The calculation results for the coefficients are given in Table 1.

Table 1. Distribution of wavelet coefficients among decomposition levels.

Number of Decomposition Level j	Wavelet Coefficients	Number of Wavelet Coefficients
−1	$a_0 = F_0$	1
0	$a_1 = F_1$	1
1	a_2, a_3	2
2	a_4, a_5, a_6, a_7	4
3	$a_8 \ldots a_{15}$	8
…	…	…
j	$a_{2^j} \ldots a_{2^{j+1}-1}$	2^j
…	…	…
$n{-}2$	$a_{N/4} \ldots a_{N/2-1}$	2^{n-2}
$n{-}1$	$a_{N/2}$	1

Further, consider two approaches to classifying bio-acoustic signals. We have used real hydroacoustic signals of whales from the database [20].

4. Classification Using the kNN-Algorithm

The classification was based on 14,822 records of whales of two types: 'killer' (4673 records) and 'pilot' (10,149 records). Data for processing was taken from [20]. Research has been conducted for the following signal-to-noise ratios (SNR): 100, 3, 0 and −3 dB. Training of the classifier was based on 85% of records of each class, and testing was based on 15% of records of each class. The following attributes have been used for comparison: the harmonic wavelet transform (HWT) coefficients, the short-time Fourier transform (STFT) coefficients and the discrete Fourier transform (DFT) coefficients.

All records had different numbers of samples (8064–900,771) and different sampling rates. To perform classification, we had to change the lengths of the records so that they equaled 2. To reduce the feature space dimension, we employed the approach based on modulo N reduction [21]. Such an approach allows us to reduce the data dimension when calculating N-point DFT if $N < L$ (L is signal length). The final signal matrix size ($N = 4096$) was 14,822 × 4096.

To reduce the feature space dimension, we also used coefficients of symmetry for the harmonic wavelet transform and the DFT: we used 50% coefficients (matrix: 14,822 × 2048). In the case of using a short-time Fourier transform (Hamming window of the size 256, overlap 50%), the final signal matrix size was 14,822 × 3999.

Below we can see the classification results (Tables 2–13, Figure 1) using the kNN-algorithm [22] for different features and different SNR values.
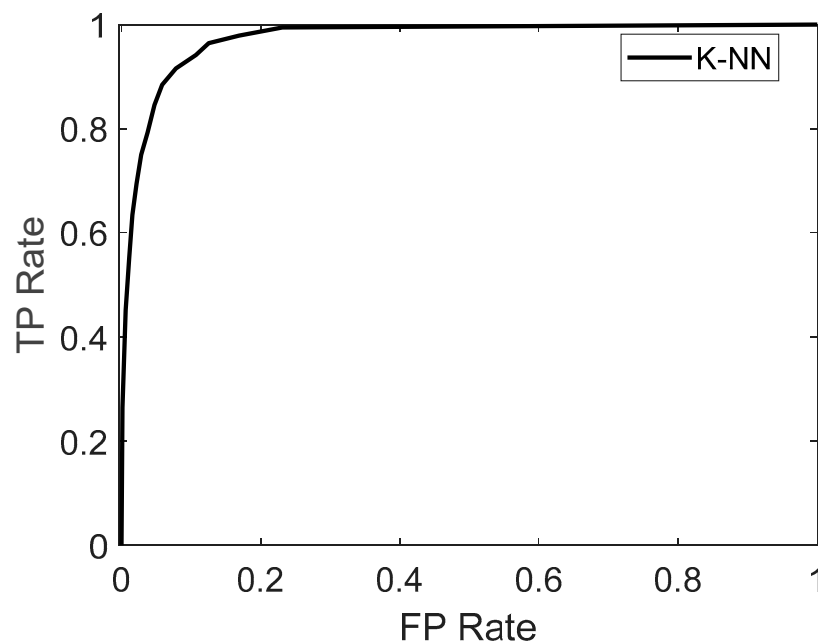
Figure 1. ROC curve of the classification: HWT, SNR = 100 dB.

The classification problem is to attribute vectors to different classes. We have two classes: positive and negative. In this case, we can have four different situations at the output of a classifier:

- If the classification result is positive, and the true value is positive as well, we have a true-positive value—TP.
- If the classification result is positive, but the true value is negative, we have false-positive value—FP.
- If the classification result is negative, and the true value is negative as well, we have a true-negative value—TN.
- If the classification result is negative, but the true value is positive, we have a false-negative value—FN.

We have calculated the following classification quality metrics: precision, recall and accuracy.

$$Precision = \frac{TP}{TP+FP}; \; Recall = \frac{TP}{TP+FN}; \; Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{11}$$

Tables 14–16 contain precision, recall and accuracy for different classification features and different signal-to-noise ratios. Additionally, we can find the average final efficiency score characterizing the use of different classification features.

Table 2. Classification results: HWT, SNR = 100 dB.

Feature: HWT SNR = 100 dB	Killer	Pilot	Totals
Killer	TP = 590	FN = 110	700
Pilot	FP = 85	TN = 1437	1522
Totals	675	1547	2222

Table 3. Classification results: STFT, SNR = 100 dB.

Feature: STFT SNR = 100 dB	Killer	Pilot	Totals
Killer	TP = 591	FN = 109	700
Pilot	FP = 107	TN = 1415	1522
Totals	698	1524	2222

Table 4. Classification results: DFT, SNR = 100 dB.

Feature: DFT SNR = 100 dB	Killer	Pilot	Totals
Killer	TP = 592	FN = 108	700
Pilot	FP = 93	TN = 1429	1522
Totals	685	1537	2222
Totals	813	1409	2222

Table 5. Classification results: HWT, SNR = 3 dB.

Feature: HWT SNR = 3 dB	Killer	Pilot	Totals
Killer	TP = 642	FN = 58	700
Pilot	FP = 171	TN = 1351	1522

Table 6. Classification results: STFT, SNR = 3 dB.

Feature: STFT SNR = 3 dB	Killer	Pilot	Totals
Killer	TP = 642	FN = 58	700
Pilot	FP = 238	TN = 1284	1522
Totals	880	1342	2222

Table 7. Classification results: DFT, SNR = 3 dB.

Feature: DFT SNR = 3 dB	Killer	Pilot	Totals
Killer	TP = 535	FN = 165	700
Pilot	FP = 112	TN = 1410	1522
Totals	647	1575	2222

Table 8. Classification results: HWT, SNR = 0 dB.

Feature: HWT SNR = 0 dB	Killer	Pilot	Totals
Killer	TP = 669	FN = 31	700
Pilot	FP = 228	TN = 1294	1522
Totals	897	1325	2222

Table 9. Classification results: STFT, SNR = 0 dB.

Feature: STFT SNR = 0 dB	Killer	Pilot	Totals
Killer	TP = 646	FN = 54	700
Pilot	FP = 297	TN = 1225	1522
Totals	943	1279	2222

Table 10. Classification results: DFT, SNR = 0 dB.

Feature: DFT SNR = 0 dB	Killer	Pilot	Totals
Killer	TP = 439	FN = 261	700
Pilot	FP = 145	TN = 1377	1522
Totals	584	1638	2222

Table 11. Classification results: HWT, SNR = −3 dB.

Feature: HWT SNR = −3 dB	Killer	Pilot	Totals
Killer	TP = 674	FN = 26	700
Pilot	FP = 333	TN = 1189	1522
Totals	1007	1215	2222

Table 12. Classification results: STFT, SNR = −3 dB.

Feature: STFT SNR = −3 dB	Killer	Pilot	Totals
Killer	TP = 617	FN = 83	700
Pilot	FP = 336	TN = 1186	1522
Totals	953	1269	2222

Table 13. Classification results: DFT, SNR = −3 dB.

Feature: DFT SNR = −3 dB	Killer	Pilot	Totals
Killer	TP = 294	FN = 406	700
Pilot	FP = 144	TN = 1378	1522
Totals	438	1784	2222

Table 14. Classification results: HWT.

HWT	SNR = 100 dB		SNR = 3 dB		SNR = 0 dB		SNR = −3 dB	
Precision	0.8740	I *	0.7897	II *	0.7458	II *	0.6693	II *
Recall	0.8429	III *	0.9171	I *	0.9557	I *	0.9629	I *
Accuracy	0.9122	I *	0.8969	I *	0.8834	I *	0.8384	I *
Averaged score for three metrics	I		I		I		I	
Final score				I				

* score of a particular metric for each SNR. The "averaged score for three metrics" means that we estimated the average score for three metrics with the same SNR. Then, the final score for each feature (HWT, STFT, DFT) with different SNRs was chosen. We can see that using HWT as features gives the best result.

Table 15. Classification results: STFT.

STFT	SNR = 100 dB		SNR = 3 dB		SNR = 0 dB		SNR = −3 dB	
Precision	0.8467	III *	0.7295	III *	0.6850	III *	0.6474	III *
Recall	0.8443	II *	0.9171	I *	0.9229	II *	0.8814	II *
Accuracy	0.9028	III *	0.8668	III *	0.8420	II *	0.8114	II *
Averaged score for three metrics	III		III		II–III		II–III	
Final score				III				

* score of a particular metric for each SNR. The "averaged score for three metrics" means that we estimated the average score for three metrics with the same SNR. Then, the final score for each feature (HWT, STFT, DFT) with different SNRs was chosen. We can see that using HWT as features gives the best result.

Table 16. Classification results: DFT.

DFT	SNR = 100 dB		SNR = 3 dB		SNR = 0 dB		SNR = −3 dB	
Precision	0.8642	II *	0.8269	I *	0.7517	I *	0.6712	I *
Recall	0.8457	I *	0.7643	II *	0.6271	III *	0.4200	III *
Accuracy	0.9095	II *	0.8753	II *	0.8173	III *	0.7525	III *
Averaged score for three metrics	II		II		II–III		II–III	
Final score				II				

* score of a particular metric for each SNR. The "averaged score for three metrics" means that we estimated the average score for three metrics with the same SNR. Then, the final score for each feature (HWT, STFT, DFT) with different SNRs was chosen. We can see that using HWT as features gives the best result.

5. Classification Using a Deep Neural Network

The classification was based on 14,822 records of whales of two types: 'killer' (4673 records) and 'pilot' (10,149 records). Data for processing were taken from [20], containing sound recordings of 26 whales of two types: killer whale (15 individuals) and pilot whale (11 individuals).

In [23], for this dataset, two classifiers were constructed based on the kNN-algorithm. In the first case, the sounds were classified into a grind or killer whale sounds. For training, 800 whale sounds of each class were used; for testing, 400 of each were used. A classification accuracy of 92% was obtained. In the second experiment, 18 whales were separated from each other. For training, they took 80 records; for testing, they took 20. The classification accuracy was 51%.

In this work, records less than 960 ms long were removed from the dataset. After that, 14,810 records with an average duration of 4 s remained: 10,149 records of the grind and 4661 records of killer whales.

The classifier for both tasks was based on the VGGish model [24], which is a modified deep neural network VGG [25] pre-trained on the YouTube-8M dataset [26]. Cross entropy was used as a loss function. The audio files have been pre-processed in accordance with the procedure presented in [24]. Each record is divided into non-overlapping 960 ms frames, and each frame inherits the label of its parent video. Then log-mel spectrogram patches of 96 × 64 bins are then calculated for each frame. These form the set of inputs to the classifier. The output for the entire audio recording was carried out according to the maximum likelihood for all classes for each segment. As features, the output of the penultimate layer of dimension 128 was taken. More details can be found in the paper [23].

5.1. Experiment 1—Classification by Type

For the first task, we divided the dataset into training and test data in the proportion 85:15; in the training and the test sample there are no sounds from the same whales. The killer whale was designated 0, and the pilot whale was designated 1. Statistics on the training set: 8486–1, 3995–0. Statistics on the test set: 1663–1, 666–0.

The following results were obtained. On the training set, the confusion matrix was:

$$\begin{pmatrix} 3994 & 1 \\ 27 & 8459 \end{pmatrix}$$

1—FP, 27—FN.

On the test set, the confusion matrix was:

$$\begin{pmatrix} 633 & 33 \\ 86 & 1577 \end{pmatrix}$$

33—FP, 86—FN

Recall = 0.95, precision = 0.98 or accuracy = 0.95, AUC = 0.99.

Figure 2 shows the ROC curve for the test set.
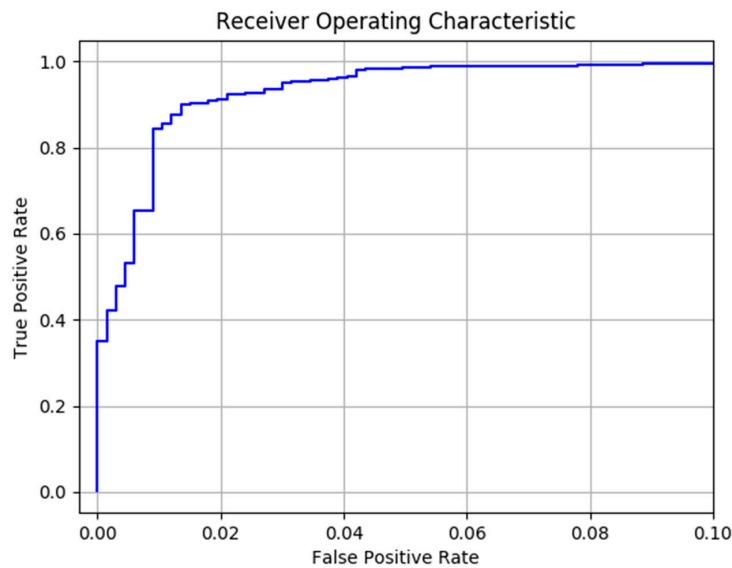
Figure 2. ROC curve for the test set.

5.2. Experiment 2—Classification by Individual

The data was divided into training and test sets in the ratio of 85:15, maintaining the proportions of the classes. As Figure 3 shows, the classes are very unbalanced. Thus, in the training test, for classes 26 and 5, 15 and 12 files are available. For class 20, 3684 files are available (see Figure 3).
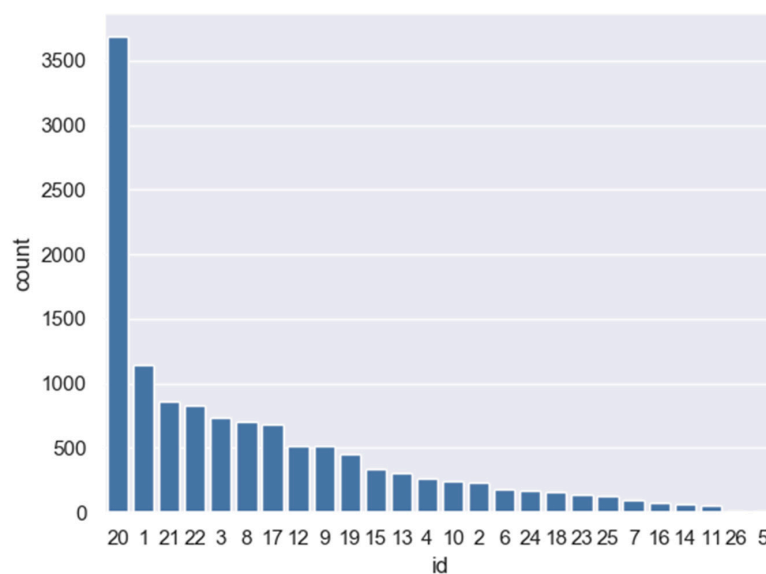
Figure 3. The number of files available for each class (individual). For training, from each class we took 900 files (augmented).

The confusion matrix for the training set is given in Figure 4.

Figure 4. Confusion matrix for the training set.

The confusion matrix for the test set is presented in Figure 5.

Figure 5. Confusion matrix for the test set.

The accuracy of the classification of individuals in percent on a test sample is presented in Figure 6. Blue lines indicate the true-positive value, orange lines indicate false-positive value.
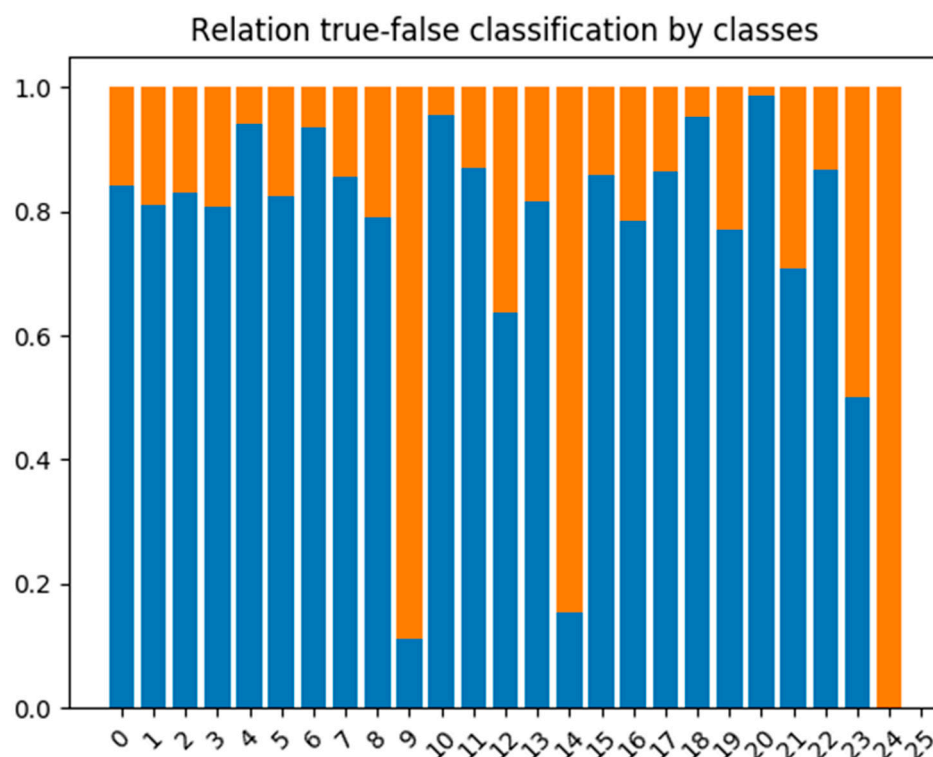

Figure 6. Classification accuracy for 26 whales.

As can be seen, the 25th (whale ID 26) class never predicts. Only for the 9th (whale ID 10), 14th (whale ID 15), 24th (whale ID 25) classes was the classification accuracy below 60%; for all the others it was higher. For some classes, classification accuracy is higher than 95%.

6. Discussion

Classification of whale sounds is a challenging problem that has been studied for a long time. Despite great achievements in feature engineering, signal processing and machine learning techniques, there still remain some major problems to be solved. In this paper, we used harmonic wavelets and deep neural networks. The results of the classification of whale types and individuals by means of deep neural networks are better than in previous works [23] with this dataset, but accuracy in the classification of types using harmonic wavelets as features and in the classification of individuals using deep neural networks should be increased. In further studies, we will use a Hilbert–Huang transform [27] and adaptive signal processing algorithms [28] to generate features.

For improvement of individual classification, two approaches can be suggested. The first combines data augmentation with other architectures of the neural network, but this will lead to large computational costs. The second approach is to use technology for simple and non-iterative improvements of multilayer and deep learning neural networks and artificial intelligence systems, which was proposed some years ago [29,30]. Our further research in the classification of hydroacoustic signals will be related to these two approaches. We also intend to test these approaches by adding noises at different SNRs, as we have done for harmonic wavelets.

7. Conclusions

In our paper, we considered the harmonic wavelet transform and its application to classifying hydroacoustic signals from whales of two types. We have provided a detailed representation of

the mathematical tools, including fast computation of the harmonic wavelet transform coefficients. Classification results analysis allows us to draw conclusions about the reasonability of using harmonic wavelets when analyzing complex data. We have established that the smallest classification error is provided by the k-NN algorithm based on the harmonic wavelet transform coefficients.

The analysis (Table 17 Figures 5 and 6) illustrates the superiority of using a neural network for the Whale FM Project dataset in comparison with known work [23] and a kNN-classifier for the classification problem [31]. However, it is worth noting that the implementation of a neural network of such a complicated structure requires significant computational resources.

Table 17. Analysis of different approaches to bioacoustic signal classification of whale types (pilot and killer).

	Recall	Precision
Neural network	0.95	0.98
kNN-algorithm Feature: HWT SNR = 100 dB	0.84	0.87

Classification of 26 individual whales from the Whale FM Project dataset was proposed, and better results in comparison with previous works were achieved [23].

The proposed approach can be used in the study of the fauna of the oceans by research institutes, environmental organizations, and enterprises producing equipment for sonar monitoring. In addition, the study showed that the same methods can be used for speech processing and classification of underwater bioacoustic signals, which will subsequently allow the creation of effective medical devices based on these methods.

References

1. Orcinus Orca (Killer Whale). Available online: https://www.iucnredlist.org/species/15421/50368125 (accessed on 16 March 2020).
2. Shiu, Y.; Palmer, K.J.; Roch, M.A.; Fleishman, E.; Liu, X.; Nosal, E.-M.; Helble, T.; Cholewaik, D.; Gillespie, D.; Klinck, H. Deep neural networks for automated detection of marine mammal species. *Sci. Rep.* **2020**, *10*, 607. [CrossRef] [PubMed]
3. Dréo, R.; Bouffaut, L.; Leroy, E.; Barruol, G.; Samaran, F. Baleen whale distribution and seasonal occurrence revealed by an ocean bottom seismometer network in the Western Indian Ocean. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **2019**, *161*, 132–144. [CrossRef]
4. Bouffaut, L.; Madhusudhana, S.; Labat, V.; Boudraa, A.; Klinck, H. Automated blue whale song transcription across variable acoustic contexts. In Proceedings of the OCEANS 2019, Marseille, France, 17–20 June 2019; pp. 1–6.
5. Bouffaut, L.; Dréo, R.; Labat, V.; Boudraa, A.-O.; Barruol, G. Passive stochastic matched filter for antarctic blue whale call detection. *J. Acoust. Soc. Am.* **2018**, *144*, 955–965. [CrossRef] [PubMed]
6. Bahoura, M.; Simard, Y. Blue whale calls classification using short-time Fourier and wavelet packet transforms and artificial neural network. *Digit. Signal Process.* **2010**, *20*, 1256–1263. [CrossRef]
7. Zhong, M.; Castellote, M.; Dodhia, R.; Ferres, J.L.; Keogh, M.; Brewer, A. Beluga whale acoustic signal classification using deep learning neural network models. *J. Acoust. Soc. Am.* **2020**, *147*, 1834–1841. [CrossRef] [PubMed]

8. Zhang, L.; Wang, D.; Bao, C.; Wang, Y.; Xu, K. Large-Scale Whale-Call Classification by Transfer Learning on Multi-Scale Waveforms and Time-Frequency Features. *Appl. Sci.* **2019**, *9*, 1020. [CrossRef]

9. Smirnov, E. North Atlantic Right Whale Call Detection with Convolutional Neural Networks. In Proceedings of the ICML 2013 Workshop on Machine Learning for Bioacoustics, Atlanta, GA, USA, 16–21 June 2013.

10. Dorian, C.; Lefort, R.; Bonnel, J.; Zarader, J.L.; Adam, O. Bi-Class Classification of Humpback Whale Sound Units against Complex Background Noise with Deep Convolution Neural Network. Available online: https://arxiv.org/abs/1703.10887 (accessed on 12 March 2019).

11. Hodges, R.P. *Underwater Acoustics: Analysis, Design, and Performance of Sonar*; John Wiley & Sons: London, UK, 2010.

12. Kaplun, D.; Klionskiy, D.; Voznesenskiy, A.; Gulvanskiy, V. Digital filter bank implementation in hydroacoustic monitoring tasks. *PRZ Elektrotechniczn* **2015**, *91*, 47–50. [CrossRef]

13. Milne, P.H. *Underwater Acoustic Positioning Systems*; Gulf Publishing Co.: Houston, TX, USA, 1983.

14. Mallat, S. *A Wavelet Tour of Signal Processing*, 3rd ed.; Academic: New York, NY, USA, 2008.

15. Klionskiy, D.M.; Kaplun, D.I.; Gulvanskiy, V.V.; Bogaevskiy, D.V.; Romanov, S.A.; Kalincev, S.V. Application of harmonic wavelets to processing oscillating hydroacoustic signals. In Proceedings of the 2017 Progress in Electromagnetics Research Symposium—Fall (PIERS—FALL), Singapore, 19–22 November 2017; pp. 2528–2533.

16. Newland, D.E. Harmonic wavelet analysis. Series A. *Proc. R. Soc. Lond.* **1993**, *443*, 203–225.

17. Kaplun, D.; Voznesenskiy, A.; Romanov, S.; Nepomuceno, E.; Butusov, D. Optimal Estimation of Wavelet Decomposition Level for a Matching Pursuit Algorithm. *Entropy* **2019**, *21*, 843. [CrossRef]

18. Newland, D.E. *An Introduction to Random Vibrations, Spectral & Wavelet Analysis*, 3rd ed.; John Wiley & Sons: New York, NY, USA, 1993.

19. Newland, D.E. Harmonic wavelets in vibrations and acoustics. *Philos. Trans. R. Soc. A* **1999**, *357*, 2607–2625. [CrossRef]

20. Whale FM Project. Available online: https://whale.fm (accessed on 3 April 2020).

21. Orfanidis, S.J. *Introduction to Signal Processing*; Prentice Hall: Upper Saddle River, NJ, USA, 1996.

22. Mitchell, T. *Machine Learning*; McGraw-Hill: New York, NY, USA, 1997.

23. Shamir, L.; Yerby, C.; Simpson, R.; von Benda-Beckmann, A.M.; Tyack, P.; Samarra, F.; Miller, P.; Wallin, J. Classification of large acoustic datasets using machine learning and crowdsourcing: Application to whale calls. *J. Acoust. Soc. Am.* **2014**, *135*, 953–962. [CrossRef] [PubMed]

24. Hershey, S.; Chaudhuri, S.; Ellis, D.P.W.; Gemmeke, J.F.; Jansen, A.; Moore, C.; Plakal, M.; Platt, D.; Saurous, R.A.; Seybold, B.; et al. CNN architectures for largescale audio classification. In Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017.

25. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2015**, arXiv:1409.1556v6.

26. A Large and Diverse Labeled Video Dataset for Video Understanding Research. Available online: https://research.google.com/youtube8m/ (accessed on 3 April 2020).

27. Huang, N.E.; Shen, S.S.P. *Hilbert-Huang Transform and Its Applications*; World Scientific: Singapore, 2005; 350p.

28. Voznesensky, A.; Kaplun, D. Adaptive Signal Processing Algorithms Based on EMD and ITD. *IEEE Access* **2019**, *7*, 171313–171321. [CrossRef]

29. Tyukin, I.Y.; Gorban, A.N.; Prokhorov, D.V.; Green, S. Efficiency of Shallow Cascades for Improving Deep Learning AI Systems. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–8.

30. Gorban, A.N.; Burton, R.; Romanenko, I.; Tyukin, I.Y. One-trial correction of legacy AI systems and stochastic separation theorems. *Inf. Sci.* **2019**, *484*, 237–254. [CrossRef]

31. Marsland, S. *Machine Learning an Algorithmic Perspective*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2016.