Hindawi
Mobile Information Systems
Volume 2022, Article ID 4466037, 6 pages
https://doi.org/10.1155/2022/4466037

Research Article

Detection Method for Echolocation Clicks Based on LSTM Networks

Dexin Duan

First Institute of Oceanography, MNR, Qingdao 266061, China

Correspondence should be addressed to Dexin Duan; ddx@fio.org.cn

Received 24 January 2022; Revised 15 February 2022; Accepted 24 February 2022; Published 19 March 2022

Academic Editor: Chia-Huei Wu

Odontocete clicks are a kind of short-time echolocation signal with high frequency. Research on the detection method of clicks is helpful for the accurate detection of marine mammal vocalization, so as to better protection of marine mammals. A method based on image processing is proposed to detect odontocete echolocation clicks. The collected data are divided into fixed-length frames and generate spectrograms. The spectrograms are filtered to remove noise and enhance the line-shape clicks. Considering that the echolocation signals are like lines in time-frequency domain, line detection is subsequently used to obtain the precise position of the lines. Finally, a Long Short-Term Memory network was trained to obtain a detector to distinguish clicks. The performance of the proposed method was evaluated using real audio recordings. The experimental results indicate that comparing with the traditional energy detector method, the proposed algorithm shows higher recall and precise under low signal-to-noise ratio (SNR). The proposed method can provide technical support in odontocete survey to accurately determine the species and better for marine bioacoustics study.

1. Introduction

Cetaceans generally can transmit clicks, which is a short-time pulse echolocation signal [1] and can be used for positioning and foraging activities [2]. Compared with communication signals, clicks have a higher frequency [3]. Accurate detection of clicks is conducive to determine the existence and appearance of cetaceans and the study of its population and biological ecology [4]. The research on click detection methods is helpful to detect cetaceans and then better protect marine mammals.

In this paper, a method of detecting clicks based on image processing is proposed. Firstly, the collected signals are divided into frames to obtain the spectrogram of each frame, and then, the spectrograms are filtered to highlight the line-shaped clicks. After filtering, the line detection is applied and extract the features. Finally, a Long Short-Term Memory network was trained to obtain a detector to distinguish clicks of cetaceans.

2. Dataset

There are three species of marine mammals' sound data used to generate dataset which are Mesoplodon densirostris, Globicephala melas, and Eubalaena japonica. Acoustic data of Mesoplodon densirostris, Globicephala melas, and Eubalaena japonica came from MobySound [9]. The MobySound is an open source marine mammal acoustics' database that provides sound data from a wide range of marine mammals in different seas around the world.

The data files of the above three animals are 24-bit audio files in WAV format. The data files of each animal only

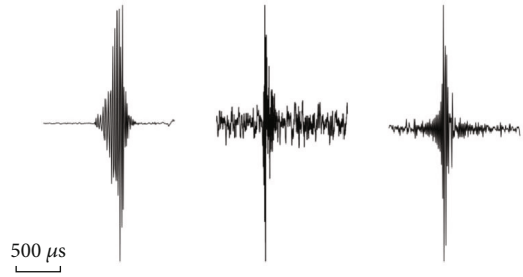

Figure 1: Waveform examples of echolocation signals from different animals.

Table 1: Data used in the experiment.

Species	Sampling rate/ Hz	Duration/ s	Number of clicks
Mesoplodon densirostris	96 000	1 800	779
Globicephala melas	96 000	900	559
Eubalaena japonica	192 000	394	225

contain echolocation signals of a single species. The waveform of three animals' clicks are shown in Figure 1. All clicks have been annotated through manual processing. All data information are shown in Table 1.

3. Method

3.1. Image Processing. Firstly, the acoustic data of three marine mammals used in the experiment were divided into several short frames with fixed frame length of 1 s. Then, the spectrograms were made by using STFT and saved as gray scale image. Then the Frangi filter is used to filter the image [10]. Frangi filter has a good effect in detecting tubular signals and is mainly applied in medical aspects, such as blood vessel detection in medical images [11]. In the experiment, the eigenvalues λ_1 and λ_2 of Hessian matrix H of each image are calculated by using this filter, and $\lambda_2 > \lambda_1$ is agreed. Finally, the output value of the filter $V_0(\sigma)$ is calculated according to the eigenvalue which is the gray value of the filtered image. The equation of getting $V_0(\sigma)$ is shown in Equation (1):

$$V_0(\sigma) = \begin{cases} 0, \lambda_1 < 0; \\ e^{-R_B{}^2/2\beta^2}\left(1 - e^{-S^2/2c^2}\right), \text{otherwise}, \end{cases} \quad S = \sqrt{\lambda_1{}^2 + \lambda_2{}^2}; R_B = \frac{\|\lambda_1\|}{\|\lambda_2\|}; \beta = 1, c = 10. \tag{1}$$

Pixels in spectrogram before filtering are displayed with different gray values as shown in Figure 2(a). As shown in Figure 2(b), pixels with gray value of 0 in the filtered spectrograms are displayed as black and nonzero pixels are displayed with different gray values. After filtering, the isolated noise in the spectrograms is filtered out. However, since the clicks and the impulse noise are approximately a straight line in the spectrograms, the impulse noise is still not removed in the spectrograms, which is necessary to further distinguish the clicks and residual noise in the spectrograms. Therefore, the line detection is introduced to accurately determine the starting position of the signal, which is helpful to extract frequency-domain characteristics [12]. The basic equation of line detection is shown in Equation (2):

$$r = x_i \cos\theta + y_i \sin\theta, \tag{2}$$

A point in the $x - y$ plane is a curve in the $r - \theta$ plane, and multiple points on a straight line in the $x - y$ plane map to multiple curves intersecting a point in the $r - \theta$ plane. The intersection number of curves in the $r - \theta$ plane at a point is set as threshold. When the intersection number exceeds the threshold, a straight line in the $x - y$ plane is detected [13, 14]. If the threshold is too low, a large number of noises will be judged as line, which increases the sample size and the calculation cost of the model, increases time to generate the model. However, some clicks with low SNR are discontinuous in the spectrograms. If the threshold is too high, this part of low SNR signals will be missed, leading to the reduction of recall rate. Therefore, different thresholds need to be tried to strike a balance between calculating cost and recall rate. By fine tuning to find the optimal parameter, the threshold is set to 50.

Pixel points with nonzero gray value in the filtered spectrograms Figure 2(b) were mapped to $r - \theta$ plane for line detection, and the results were shown in Figure 2(c). In this experiment, the duration of four animals' clicks is less than 0.002 seconds, so each click in the spectrogram is less than 2 pixels width. If two vertical lines are detected during the line detection, we took the mean value of two lines.

3.2. Long Short-Term Memory Network. In this paper, a Long Short-Term Memory (LSTM) network [12] is used to adaptively distinguish the clicks from other line-shape pulse noise from spectrograms. Recurrent Neural Networks (RNNs) are a deep learning model for learning sequential data [15]. LSTM is an improved RNN network. For a conventional RNN, the hidden state of each layer is achieved by transformation and activation of the former layer. So, the derivative used for back propagation contains the continued product of every step, which could cause gradient vanishing or gradient explosion gradient. So, it is difficult for a conventional RNN to tackle the problem of "long-range dependence" to learn the information contained in long sequence. A diagram of a basic LSTM unit is shown in Figure 3.

From the perspective of external structure, the input and output of LSTM and RNN are exactly the same. They also accept external input X_t and the hidden state h_{t-1} of the previous stage at each step and output a value. However, different from ordinary RNN, LSTM's hidden state has two parts, one is h_t and the other is C_t. The C_t is the main message that passes between the steps. By addition, it is possible to pass C_t over the cell without trouble, so a gradient can travel over long distances. In an LSTM unit, there are mainly three

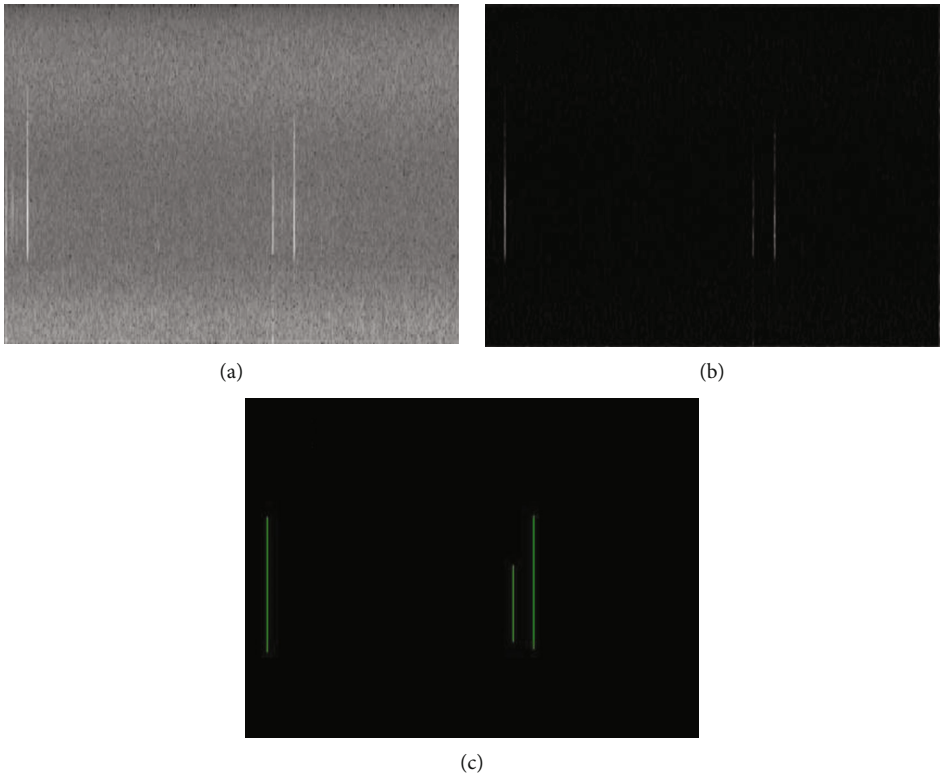

Figure 2: Example of image process procedure.

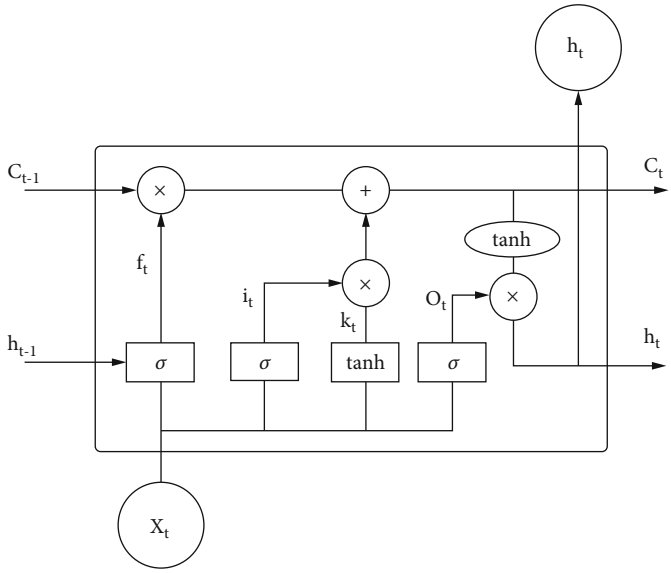

Figure 3: Diagram of a LSTM unit.

different "gates" to control the transmission of information which are forget gate, memory gate, and output gate [16].

Forget gate controls which part from C_{t-1} should be forgotten by the current LSTM unit. The output f_t of forget gate is shown in Equation (3):

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right). \tag{3}$$

The σ is the sigmoid activation function, its output falls between 0 and 1. The output of a forget gate is a matrix which has same shape of C_{t-1}, this matrix will multiply C_{t-1} to decide which part to forget. The input of a forget gate is the external input X_t and the hidden state h_{t-1} of the previous stage.

Memory gate controls which part from C_{t-1} should be memorized by the current LSTM unit. The output i_t of

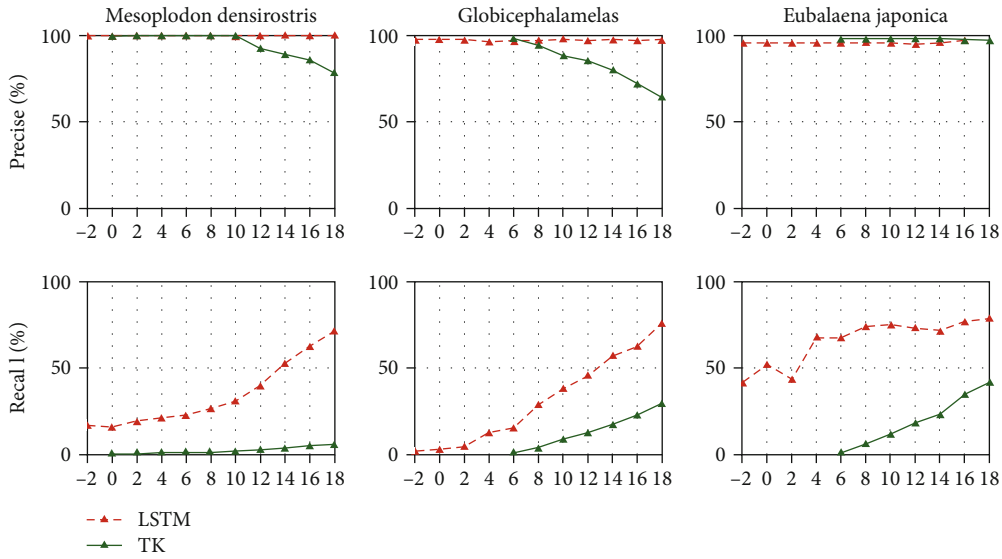

Figure 4: Performance of two methods in different SNR.

forget gate is shown in Equation (4):

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) k_t = \tan h(W_C \cdot [h_{t-1}, x_t] + b_C). \tag{4}$$

Output C_t is the combined effect of forget gate and memory gate as shown in Equation (5).

$$C_t = f_t \times C_{t-1} + i_t \times k_t. \tag{5}$$

The final output of a LSTM unit h_t is shown in Equation (6).

$$\begin{aligned} O_t &= \sigma(W_o[h_{t-1}, x_t] + b_o), \\ h_t &= O_t \times \tan h(C_t). \end{aligned} \tag{6}$$

In the experiment, the lines in spectrograms obtained in 2.1 were fed to the LSTM network. Every line was considered as a sequential data and every point in the sequence corresponding to a pixel of the line in spectrogram. All the lines were fed to the network from the pixel corresponding to the lowest frequency to the pixel corresponding to highest frequency. The amplitude of ever point is the STFT result. According to the annotation of data, the clicks are labeled as 1 and the nonclicks are labeled as 0. The obtained lineshape input and label form the dataset to train the LSTM network. In this way, the network could learn the advance feature of the clicks. 90% of the dataset are the training sets and the other 10% of the dataset are the testing sets. The detection models of these three animals are trained separately. After the training, the testing sets are sent to the trained model to get the output as the result of the proposed model.

3.3. TK Algorithm. TK algorithm is the most widely used click detection algorithm. This algorithm is usually used for the detection of sperm whale clicks [7], and its average precise rate on the sperm whale clicks can reach 94.05%, which has been proved to be highly effective.

According to the background noise of different data, different SNR thresholds are set by manually checking the data. After frame processing, the signal-to-noise ratio (SNR) of each frame is calculated. The frame with higher SNR than the threshold is selected as the candidate frame. The TK algorithm uses three continuous sampling points to calculate instantaneous energy and detect clicks [7]. The TK energy operator Ψ is defined as follows:

$$\Psi[x(n)] = x^2(n) - x(n+1) \times x(n-1). \tag{7}$$

In each selected candidate frame, Equation (7) can be used to calculate the TK operator output value. The maximum value of TK operator in each candidate frame is recorded as clicks, and its specific position is located; then, the detection performance are counted, which are the precise rate and recall rate.

3.4. Experiment Platform. The experiment was carried out on a laptop operating Windows 10, with 16 GB RAM available and an Intel(R) Core(TM) i7-8565U CPU @ 1.80 GHz processor. The data used in the experiment is normalized first, and Gaussian white noise is added to the original data by controlling the amplitude coefficient to obtain the test data with different SNR. The experiment was then carried out through the steps described in Sections 2.1 and 2.2.

4. Result and Discussion

Comparing the precise rate and recall rate of the proposed method with TK algorithm as shown in Figure 4, the recall rates of the two algorithms are positively correlated with the SNR. However, the proposed method had higher recall rate than TK algorithm, this is because the TK algorithm used the threshold detection. When the SNR is lower than a certain threshold, the recall rate falls to zero and no click

can be detected. For example, when the SNR of the data of Mesoplodon densirostris is 0 dB or lower, and the SNR of the data of Globicephala melas and Eubalaena japonica is 6 dB or lower, and the TK algorithm cannot detect any clicks. On the contrary, the proposed method can still detect signals with high precise rate under the above condition of low SNR.

For the data of Mesoplodon densirostris and Globicephala melas, the precise rate of TK algorithm will decrease when the SNR is high and decrease to 78% and 65%, respectively, when the SNR is 18 dB. However, the precise rate of the proposed method changes little with SNR (not less than 98%).

The clicks of cetaceans resemble a straight line in the spectrograms. Frangi filter can filter out isolated noise points in the image and highlight linear signals. However, linear detection can detect more linear segment noise in the spectrograms under the condition of low SNR, so the proposed method in this paper has a high recall rate under the condition of low SNR. Linear detection can accurately determine the starting and ending positions of linear signals in the image and calculate the corresponding eigenvalues according to the definition of signal characteristics. We trained a LSTM network to distinguish clicks and nonclicks like pulse noise. Comparing with other method like random forest, LSTM network can adaptively learn the characteristics of clicks, without manually extracting features. This further improved the precise rate of the proposed method.

Although the proposed method has certain advantages in the precise and recall rate, the clicks with a certain length of time are treated as a straight line segment in the spectrograms during linear detection which means the information of time dimension is lost characteristics such as the duration of clicks cannot be further utilized. How to add time dimensional features such as the duration of clicks into the algorithm will be the future optimization direction of the proposed method.

5. Conclusion

This paper firstly introduce Frangi filter and LSTM network to the cetacean clicks' detection. Acoustic data of Mesoplodon densirostris, Globicephala melas, and Eubalaena japonica are used to generate the dataset. By adding noise to generate different SNR conditions, experiment was carried out by comparing the proposed method with traditional algorithms of TK. The results show that the Frangi filter can effectively filter out random noise, and the line detection can precisely detect the line-shape clicks. Combining with the LSTM networks, the detection model can distinguish clicks and remained pulse noise accurately. The proposed method has higher recall rate and maintains higher precise rate at low SNR, which verifies its effectiveness and robustness. Under the condition of low SNR, the rate of the proposed method shows better detection effect, which can provide certain technical support for the research of cetacean acoustic signals. Whether can adding timedimensional information such as clicks pulse interval to achieve classification will be the next research step.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The author(s) declare(s) that they have no conflicts of interest.

Authors' Contributions

Dexin Duan designed and carried out the experiment mentioned in paper and wrote this manuscript.

References

[1] D. K. Mellinger, K. M. Stafford, S. E. Moore, R. P. Dziak, and H. Matsumoto, "An overview of fixed passive acoustic observation methods for cetaceans," *Oceanography*, vol. 20, no. 4, pp. 36–45, 2007.

[2] B. Møhl, M. Wahlberg, P. T. Madsen, A. Heerfordt, and A. Lund, "The monopulsed nature of sperm whale clicks," *The Journal of the Acoustical Society of America*, vol. 114, no. 2, pp. 1143–1154, 2003.

[3] F. Q. Niu, Y. M. Yang, Z. M. Zhou, X. Y. Wang, S. Monanunsap, and C. Junchompoo, "Echolocation clicks of free-ranging Irrawaddy dolphins (Orcaella brevirostris) in Trat Bay, the eastern Gulf of Thailand," *The Journal of the Acoustical Society of America*, vol. 145, no. 5, pp. 3031–3037, 2019.

[4] T. A. Marques, L. Thomas, J. Ward, N. DiMarzio, and P. L. Tyack, "Estimating cetacean population density using fixed passive acoustic sensors: an example with Blainville's beaked whales," *The Journal of the Acoustical Society of America*, vol. 125, no. 4, pp. 1982–1994, 2009.

[5] T. Akamatsu, D. Wang, K. Wang, and Z. Wei, "Comparison between visual and passive acoustic detection of finless porpoises in the Yangtze River, China," *The Journal of the Acoustical Society of America*, vol. 109, no. 4, pp. 1723–1727, 2001.

[6] J. F. Kaiser, "On a simple algorithm to calculate the'energy'of a signal," in *International conference on acoustics, speech, and signal processing: IEEE*, pp. 381–384, Albuquerque, NM, USA, 1990.

[7] V. Kandia and Y. Stylianou, "Detection of sperm whale clicks based on the Teager-Kaiser energy operator," *Applied Acoustics*, vol. 67, no. 11-12, pp. 1144–1163, 2006.

[8] R. Morrissey, J. Ward, N. DiMarzio, S. Jarvis, and D. Moretti, "Passive acoustic detection and localization of sperm whales (Physeter macrocephalus) in the tongue of the ocean," *Applied acoustics*, vol. 67, no. 11-12, pp. 1091–1105, 2006.

[9] D. K. Mellinger and C. W. Clark, "MobySound: a reference archive for studying automatic recognition of marine mammal sounds," *Applied Acoustics*, vol. 67, no. 11-12, pp. 1226–1242, 2006.

[10] W. Fu, K. Breininger, T. Würfl, N. Ravikumar, R. Schaffert, and A. Maier, "Frangi-Net: a neural network approach to vessel segmentation," 2017, https://arxiv.org/abs/03345.

[11] A. F. Frangi, W. J. Niessen, K. L. Vincken, M. A. Viergever, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *International conference on medical image computing and computer-assisted intervention*, pp. 130–137, Springer, 1998.

[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[13] R. O. Duda and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Communications of the ACM.*, vol. 15, no. 1, pp. 11–15, 1972.

[14] C. Galamhos, J. Matas, and J. Kittler, "Progressive probabilistic Hough transform for line detection," in *Proceedings 1999 IEEE computer society conference on computer vision and pattern recognition (Cat No PR00149)*, pp. 554–560, Fort Collins, CO, USA, 1999.

[15] L. R. Medsker and L. Jain, "Recurrent neural networks," *Design and Applications*, vol. 5, pp. 64–67, 2001.

[16] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: continual prediction with LSTM," *Neural computation*, vol. 12, no. 10, pp. 2451–2471, 2000.