Application of NLMS, SPRT, and CUSUM on Manatee Calls Segmentation

Pengyu Zhang, Graduate Student, University of Florida

***Abstract*—This project implements Normalized Least Mean Squares (NLMS) algorithm, Sequential Probability Ratio Test (SPRT), and Cumulative Sum (CUSUM) on manatee calls segmentation problem. Analysis of the frequency properties of the training data, as well as the prediction model diagram are presented. Related hyperparameters of models are provided. Confusion Matrix is used for evaluating the performance of both NLMS and CUSUM. Problems of segmentation related to non-stationary signals when applying NLMS, SPRT, and CUSUM are discussed.**

***Index Terms*—Normalized Least Mean Squares, Speech Segmentation, Non-stationary signal.**

I. Introduction

NLMS is widely used because it has been shown that the NLMS can utilize different step sizes based on time-varying inputs, which provide NLMS faster speed of convergence and smallest disturbance during the process of weight Iteration. SPRT and CUSUM are pure statistic methodologies and typically used for change detection to a random process while does not require adapting any model

In this project, I implement the NLMS firstly to train 11 different models on the given audios, one for noise prediction, the other ten for manatee calls prediction. The manatee calls signal is composed of ten different manatee calls without noises, which are divided by people by hand. The noise audio is just the record of background noise. The testing data is consists of unsegmented manatee calls mixed with noises that are going to be used as a test signal. The block diagram in Fig. 1 shows the system I implement on the segmentation problem. The gate is used as a generator of a square wave in which a value of 1 means the signal is classified as manatee calls, and the value of 0 means it is predicted as background noise. Additionally, a moving average error smoother is also applied in this system to obtain better segmentation accuracy.

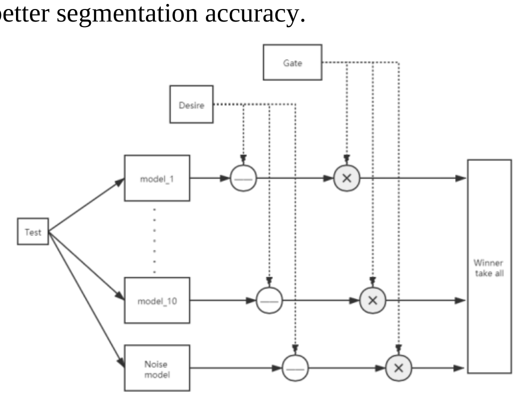


Fig. 1. Detecting & Segmentation System

The second section explains the properties of the training dataset. The third section focuses on the analysis of experiments based on NLMS. The fourth section introduces the result based on CUSUM. The fifth section concludes the results produced by experiments, followed by demonstrating some problems when dealing with segmentation problems in non-stationary signals.

II. Frequency properties of training data

A. Manatee calls

The training dataset of manatee calls are segmented by biologists that represent ten different calls I need to detect in the testing dataset. The following figures show the part of the frequency properties of six of manatee calls:

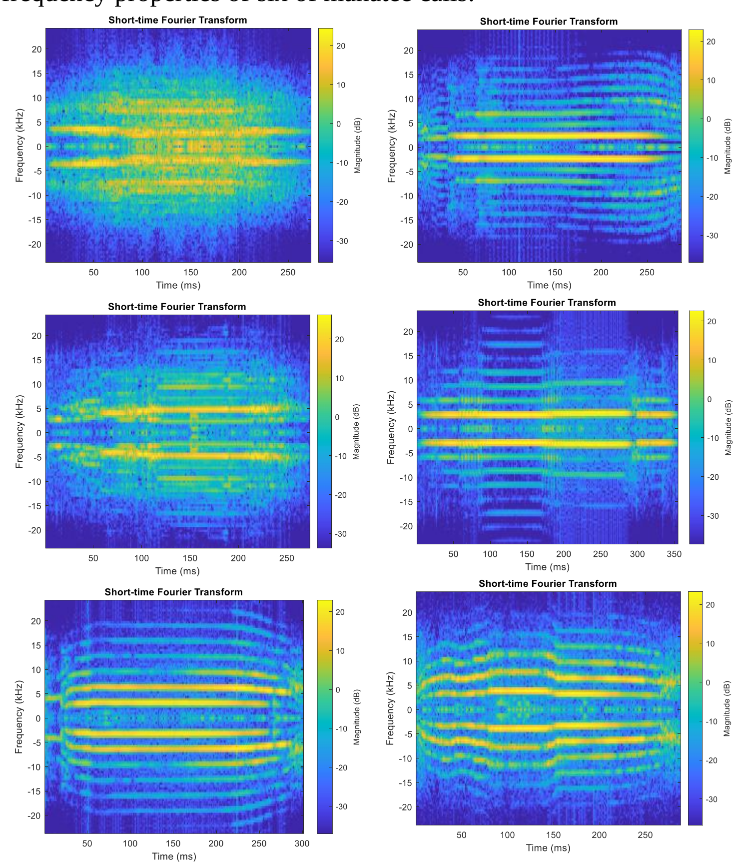


Fig. 2

The frequency properties show the variability of manatee calls, which indicates that one signal model for calls detection might be hard to implement and obtain high accuracy. Therefore, one detection model for one specific call is necessary

B. Noise

Figure 3 shows the frequency response of background noise:

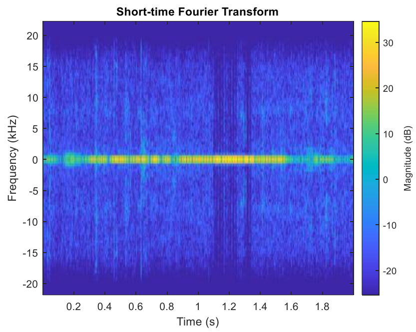


Fig. 3

Compared with the frequency response of manatee calls, the primary frequency of noise is relatively lower. Because the noise signal is not segmented by hand, which means I can only treat it as one overall training data with large variability, a smaller filter order is needed for a more generally detection.

III. Experiments Design & Results Based on NLMS

A. NLMS Methodology With MSE

The NLMS algorithm is mathematically given as formula (1) and formula (2):

$$e(n) = d(n) - \boldsymbol{H}(n)^T \boldsymbol{n}(n) \tag{1}$$

$$\boldsymbol{H}(n+1) = \boldsymbol{H}(n) + \frac{\eta}{\| n(n) \|^2 + \delta} e(n)\boldsymbol{n}(n) \tag{2}$$

Fig. 2. Block diagram of NLMS Adaptive Filter

where $\mathbf{n}(n) = [n(n), n(n-1), n(n-2), \ldots, n(n-L+1)]^T$. η is the learning rate, and δ is a constant that regularizes the denominator in the case of being zero.

B. Hyperparameters

The hyperparameters of each of the manatee calls detection models are shown in Table I:

TABLE I

Filter order	Learning rate	δ
300	1.05	0.001
300	1.5	0.001
300	1.55	0.001
300	1.1	0.001
300	1.255	0.001
300	1.05	0.001
300	1.2	0.001
300	1.18	0.001
300	1.25	0.001
300	1.01	0.001

I experiment with different filter orders in a range between 3 and 300 for manatee calls and find that the performance of models on the training dataset changes little if the filter order is large than 5. But too small filter order cause bad performance on testing data, which indicates that the detection model for manatee calls should be kind of overfitting because we don't know what exactly types of manatee calls exist in the testing dataset.

The hyperparameters of the noise detection model is shown in Table II:

TABLE II

Filter order	Learning rate	δ
3	1.1	0.001

The same thing also occurs when increasing the value of the filter order of noise detection model. The performance of the model almost remains the same if the filter order is larger than 3; the weight track also provides the same information. Therefore, considering the variability of background noise, the filter order I choose finally is 3 to obtain more general detection result instead of a significant value of filter order.

C. Error smoother

We expect that the high value of the square wave concentrates explicitly on the period that we hear the manatee calls. Therefore if the density of high value is high in one period, we should take the period as manatee calls. But the density of high value can be easily affected by the switching between 1 and 0. Therefore an error smoother is needed to take advantage of memory to help smooth the square wave.

Given that the manatee calls usually last at least 0.1s, I implement a moving average window as the error smoother to provide memory information. Without the averaging process, the output is going to be very noisy, so that it's hard to distinguish the manatee calls from the square value.

Based on the moving average process, I further set up a threshold to control the square wave generation while comparing the error between each output obtained from each model. The value of the threshold is 0.079.

D. Segmentation result

Figure 4 presents the segmentation results, and figure 5 is the True manatee calls distribution:

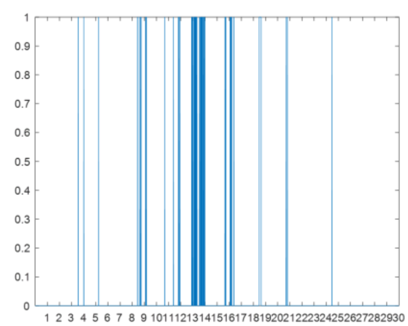


Fig. 4

A large window of smoothing can sometimes introduce extra noise. However, a smaller window might be lack of accuracy. There because I choose a window of 300 samples, and consider that manatee calls last in a range of 0.1s to 0.3s if the value of 1 of the square wave occurs more than 2 in a variety of 0.2 seconds, It's reasonable to classify the signal as manatee calls. So It's clear to decide on how many manatee calls I detect and correctly classify. The true classification and predicted result, as well as a confusion matrix, are shown in Table III and Table IV:

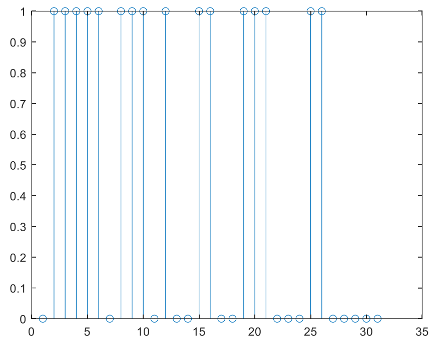


Fig. 5

TABLE III

Time /second	True	*Predicted*
0	0	0
1	1	0
2	1	0
3	1	1
4	1	1
5	1	1
6	0	0
7	1	0
8	1	1
9	1	1
10	0	1
11	1	1
12	0	1
13	0	1
14	1	1
15	1	0
16	0	1
17	0	0
18	1	1
19	1	0
20	1	0
21	0	1
22	0	0
23	0	0
24	1	1
25	1	0
26	0	0
27	0	0
28	0	0
29	0	0
30	0	0

TABLE IV

		True Condition		Total Number
		p	n	
Predicted Condition	p`	7	3	10
	n`	4	2	6
Total Number		11	5	

The accuracy is defined as the ratio between (TP+TN) and (P+N). Under the condition that the threshold is set as 0.079, the accuracy is (7+2)/(11+5), which is 56.25%. One problem exists is that the baseline is just obtained by hearing the audio by myself. Therefore it's hard to say the "True" in table III is 100% accurate, which will directly affect the confusion matrix.

IV. Results Based on SPRT and CUSUM

A. SPRT Methodology and result

The Sequential Probability Ratio Test aims to obtain the statistic of the training data by implementing maximum likelihood estimation. In this project, two different probability density distributions are expected; one is estimated from manatee calls, the other is estimated from background noise. The goal of SPRT is to compare the log-likelihood ratio between two candidate probability density distributions and the pdf obtained from the testing data. The formula is given as below:

$$L_{ij}(n) = \log\left[{p_i\big(X(n)\big)}\Big/{p_j\big(X(n)\big)}\right]$$

Because of the MLE requires a sequence of data, and we should also make an assumption that this segment of data is stationary, the segmentation of SPRT should be chosen carefully. Therefore, I choose 30000 samples, which is close to 0.7s in time domain, as the length of each segment with a step of 3000 to generate multiple segments. A threshold is also needed to control the format of square wave. I choose 0.01 as the value of threshold. The square wave generated by SPRT is as follows:

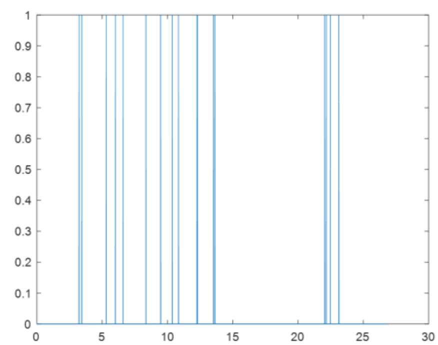


Fig. 6

Given that we have already known the true distribution of manatee calls in time domain from figure 5, a similar process is going to be done to calculate confusion matrix:

TABLE V

		True Condition		Total Number
		p	n	
Predicted Condition	p`	7	2	9
	n`	4	3	7
Total Number		11	5	

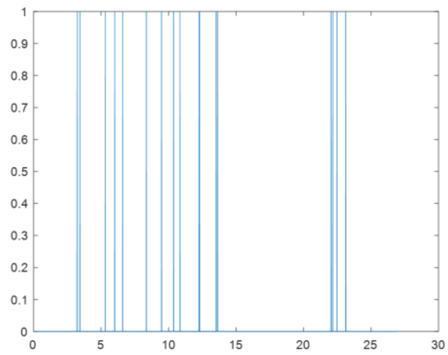
The accuracy of SPRT is (7+3)/(11+5), which is 62.5%.

B. CUSUM Methodology and result

SPRT takes advantages of MLE to contruct statistical information from each signal but with an assumption that the signal is locally stationary. The CUSUM provide method to track the change of the region instead of only focusing on small length of segment. The main idea behind the CUSUM is to reset the log-likelihood relatively to its minimum value. The calculation formula is as follows:

$$\triangle L_{ji}(n) = L_{ji}(n)$$
$$-min\left[L_{ji}(n-1), L_{ji}(n-2), \ldots L_{ji}(n-N)\right]$$

The parameter N represent the number of log-likelihood test in CUSUM, which is 20000 specifically. The threshold is set as



0.01. The length of segment is still 30000 samples with a step of 3000. The final output is shown in figure 7.

Fig. 6

I implement the same criterion again to obtain the confusion matrix as Table VI shows:

TABLE VI

		True Condition		Total Number
		p	n	
Predicted Condition	p`	7	2	9
	n`	4	3	7
Total Number		11	5	

The accuracy of detection almost remains the same as that of SPRT, which is 62.5%

V. Conclusion and Discussion

A. Conclusion

Because there doesn't exist a good criterion that can help me to evaluate the performance instead of hearing the audio and make decision by people themselves. The manatee calls distribution in time domain may vary due to different. Taking the three accuracies into consideration, it's clear to conclude that they center around 60%. But the NLMS and CUSUM is apparently required more computational resources, and the former need a moving average smoother, the latter need to recursive compute the log-likelihood.

It's valuable to mention that the filter order of NLMS profoundly affects the performance. I tried to increase the order of detecting noise model during the experiments but obtained pretty lousy performance. If the signal is changing all the time, like the background noise, we should consider using smaller filter order to get more general prediction.

B. The problem in segmentation assignment

When applying the SPRT or CUSUM, the length of segment we choose is important. The prerequisite of SPRT is the signal should be locally stationary. The CUSUM is more robust than SPRT to non-stationary segment; however, it still uses the information of MLE. Practically, we can implement the autocorrelation function to judge whether the segment uint can be assumed as a stationary signal. In this project, the goal is to detect every call, so choosing a length of the window that equals the length of samples corresponds to each call can exactly include one call at a time might be helpful.