Robust detection of dolphin whistles in noisy data

Wang Dizhi

Tutor: Ramon Miralles

Guillermo Lara

Bachelor's Thesis presented at Escuela Técnica Superior de Ingenieros de Telecomunicación of Universitat Politècnica de València,

Academic Year 2019-2020

Valencia, 5 de March de 2020

Abstract

A fast automatic detection technique for dolphin whistle detection is studied in this work. Starting from a whistle detection algorithm developed by the iTEAM and used in the SAMARUC passive acoustic monitoring device, we have worked to enhance its detection probability while at the same time keeping a low false alarm ratio. The existed detection system is composed of a pre-detector and a whistle detector. The former is obtained as the combination of the PTR detector and STA/LTA detector (Miralles, 2012), whereas the latter is based in Gillespie's (20132) work and includes algorithms to remove pulsed noise and continuous noise. We have focused in the whistle detector and we have incorporated some enhancements in this part. Using manual annotation of dolphin whistles as standard criteria, the recall rate of new algorithm is 75.4%, which is 501.7% higher than the recall rate of old algorithm. The precision rate of new algorithm is 97.7% and it is 95% higher than the precision rate of old algorithm. Although this algorithm does not apply pre-detection, this paper discussed the application of pre-detection techniques as more as possible, based on the try during the development of the new detector.

Resumen

En este trabajo se estudia una técnica de detección automática rápida para la detección de silbidos de delfín. Partiendo de un algoritmo de detección de silbidos desarrollado por el iTEAM y utilizado en el dispositivo de monitorización acústica pasivo SAMARUC, hemos trabajado para mejorar su probabilidad de detección y al mismo tiempo mantener una baja tasa de falsa alarma. El sistema de detección existente está compuesto por un predetector y un detector de silbidos. El primero se obtiene como la combinación del detector PTR y el detector STA / LTA (Miralles, 2012), mientras que el segundo se basa en el trabajo de Gillespie (20132) e incluye algoritmos para eliminar el ruido impulsivo y el ruido continuo. Nos hemos centrado en el detector de silbidos y hemos incorporado algunas mejoras en esta parte. Utilizando la anotación manual de silbidos de delfín como criterio estándar, la tasa de recuperación del nuevo algoritmo es del 75,4%, que es 501,7% más alta que la tasa de recuperación del algoritmo antiguo. La tasa de precisión del nuevo algoritmo es 97.7% y es 95% veces mayor que la tasa de precisión del algoritmo anterior. Aunque finalmente este algoritmo no emplea ningún predetector, este documento estudia la aplicación de técnicas de predetección basándose en los intentos preliminares de incluir estos en el diseño del nuevo detector.

Resum

En este trabajo se estudia una técnica de detección automática rápida para la detección de silbidos de delfín. Partiendode un algoritmo de detección de silbidos desarrollado por el iTEAM y utilizado en el dispositivo de monitorización acústica pasivo SAMARUC, hemos trabajado para mejorar su probabilidad de detección y al mismo tiempo mantener una baja tasa de falsa alarma. El sistema de detección existente está compuesto por un predetector y un detector de silbidos. El primero se obtiene como la combinación del detector PTR y el detector STA / LTA (Miralles, 2012), mientras que el segundo se basa en el trabajo de Gillespie (20132) e incluye algoritmos para eliminar el ruido impulsivo y el ruido continuo. Nos hemos centrado en el detector de silbidos y hemos incorporado algunas mejoras en esta parte. Utilizando la anotación manual de silbidos de delfín como criterio estándar, la tasa de recuperación del nuevo algoritmo es del 75,4%, que es 501,7% más alta que la tasa de recuperación del algoritmo antiguo. La tasa de precisión del nuevo algoritmo es 97.7% y es 95% veces mayor que la tasa de precisión del algoritmo anterior. Aunque finalmente este algoritmo no emplea ningún predetector, este documento estudia la aplicación de técnicas de predetección basándose en los intentos preliminares de incluir estos en el diseño del nuevo detector.

Index

Capítulo 1. Introductory

1.1 Background of the dolphin whistle detection

With the increasing utilization of marine resources, such as expansion in shipping, exploration of oil and gas, and advanced development of all kinds of infrastructures, anthropogenic noises are also increasing in the marine environment (Figure 1). Those noises of different intensity and frequency can result in chronic and acute impacts on marine organisms. The studies about how anthropogenic noises impact marine organisms focus mainly on the adult fish and marine mammals. Those studies show that marine noises can cause auditory masking, leading to cochlear damage, changes individual and social behaviour, altered metabolisms, hampered population recruitment and can subsequently affect the health and service functions of marine ecosystems.

To help reduce such kind of negative effect as more efficient as possible, we need to understand the undesired effect of these noises and then try to create policies to achieve a good environmental state of our seas and oceans. Acoustic monitoring can play a major role in it. The objective of the present project is to evaluate different dolphin whistle detection algorithms and propose those which give better accuracy in noisy environments. For this purpose, the group has a large database obtained with the collaboration of Spanish Institute of Oceanography.

Acoustic monitoring is also used to determine whether or not a sound source is affecting the behavior of marine animals. For example, dolphins and beluga whales have been found to shift the frequency of their clicks to avoid noise in the normal frequency range of their echolocation (Au *et al*, 1993). This indicates that noise may be affecting the animals by reducing the efficiency of their echolocation. However, the animals are able to compensate for this noise by changing the structure of their echolocation clicks. Acoustic monitoring has also been used to show that the average length of the humpback whale song increased during and following SURTASS-LFA sonar transmissions, although much of the variation in song length was not related to the transmissions (Fristrup *et al*, 2003).

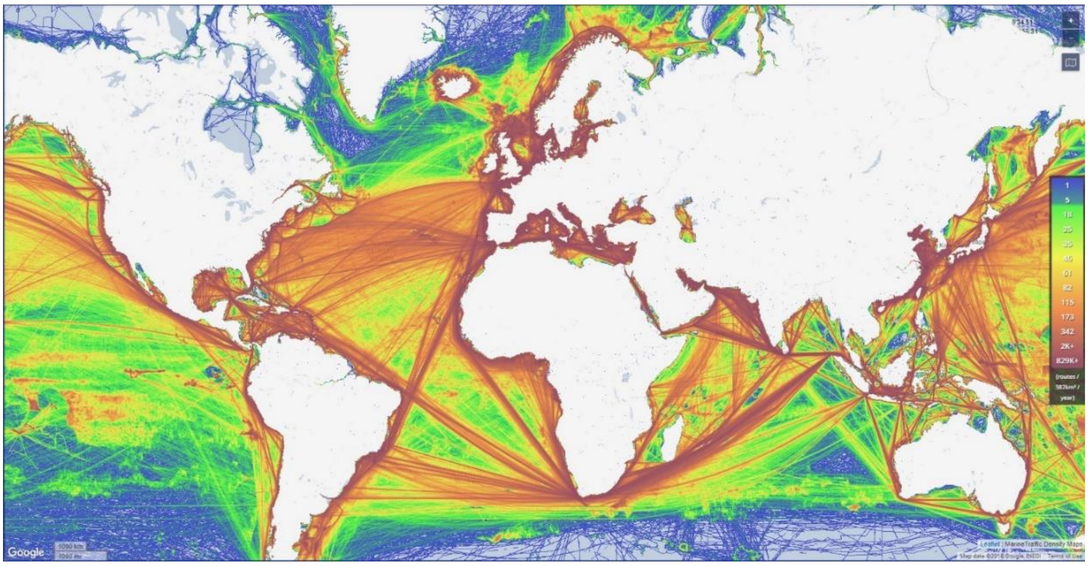


Figure 1: Marine Traffic lastingly collects data on ship positions (MarineTraffic, 2020)

Every year, a large number of marine mammal vocalizations are recorded and studied for various purposes, such as studying behavior and situational associations, animal detection and location, and census. Most of these natural signals of interest are non-stationary waves, ideal for analysis using time-frequency representations (Mallawaarachchi, 2007). With such a huge data set, it is almost impossible for the operator to analyze by looking at the spectrogram or listening to the recording (Gillespie, 2013). Also, because those records of marine mammal usually sustain for at least a month, it is important to develop a method that can process such a big database quickly and efficiently.

Dolphins (family: Delphinidae) are an extremely vocal mammalian family and vocal communication plays an important role in mediating social interactions (Herzing, 2000; Janik, 2009). The focus of this work will be specially on the enhancement of existed methods used for dolphin whistle detection. The classic way to detect dolphin whistles are finding peaks in the spectrogram of audio data and connect those peaks together to see whether it leads a contour that should indicate a whistle (Gillespie, 2013). There are also some ways using particle filter to estimate the contour based on the peaks (Roch 2011) or using Gaussian mixture probability hypothesis density (GM-PHD) filter track the spectral peaks which can be a whistle contour (Gruden, 2016). Some more complicate ways include phase tracker technique which develop the analysis not based on the spectrogram but based on a local analysis of time, frequency, and phase coherence (Ioana, 2010) and applying graph searching techniques to detect a contour in the peaks of spectrogram (Roch 2011).

The method we enhanced is the classic way based on the work of Gillespie, which starts from a time frequency representation, find the maximum peaks, and then fit a curve or spline to contours. One of the biggest problems which influence strongly the accuracy of detection results comes from the complex noises in the marine environment. De-noising time-frequency representations is one of the most vital steps. The most common noises include broad clicks (Short-duration transient noise), constant tones and harmonic noises (harmonics which are contained by whistles themselves). With the purpose to detect whistles, we focus mainly on remove clicks, constant tones and improve the quality of spectrograms.

All the data of dolphin whistles we used comes from the Polytechnic University of Valencia and the iTEAM research group database. The data collection has been obtained through several projects funded by the EU, by the Spanish Ministry and by the Spanish Institute of Oceanography. Data was acquired using a system entirely developed by the iTEAM – UPV researchers named SAMARUC. The recordings were done in different Marine protected areas in the Meditteranean and Cantabrian Seas.

More specifically, this work focus on the enhancement of the existed detection system of dolphin whistles named SAMARUC, which is developed by iTEAM. The existed detector combines Gillespie’s 6 step method detection with an added pre-detector to increase the speed of the process. But the result is not good enough. Applying the old detector (to make a clear difference between the existed detector and the enhanced detector, this work will use old detector to mention the existed detector and new detector to point enhanced detector), it can only detect (recall) 0.15% of human identified sounds, and the false alarming is 50%.

1.2 Old methods of Whistle Detection

Out of practical consideration that the detection should not take too long to help biologists find the whistles in the database, old detector added a pre-detector to decrease the time of applying the algorithm based on Gillespie’s method. The 2 pre-detectors that have been proved effective are STA/LTA detector and PTR detector (Miralles, 2017). Old detector takes the raw audio as the input of both PTR detector and STA/LTA detector. Then collect the active events that are detected by the 2 separate pre-detectors and remove some duplicate events. Only detected active events can be sent to the formal detector to check whether they are real whistles or not. Using Gillespie’s 6 step method, we can find some possible contours which can be the candidates of dolphin whistles. With a score system which focus on the duration of those contours, we can pick which should be a whistle and which should not. The process of old detector shows in Figure 2 and some more explanation about the details of PTR detector, STA/LTA detector and Gillespie’s method shows in section 1.2.1 and 1.2.2.

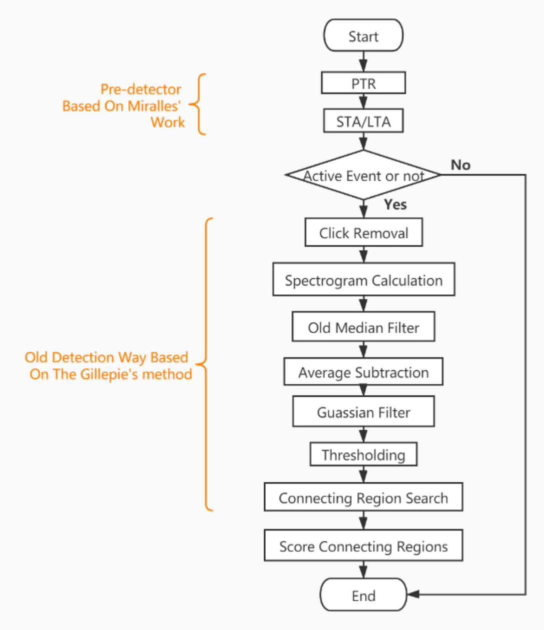


Figure 2: Old Algorithm to Detect Dolphin Whistles

1.2.1 Miralles' Pre-detection method

Based on Miralles' work, the existing whistle detector uses 2 different pre-detectors. One is short term average/ long term average (STA/LTA) and the other one is the pulsed to tonal ratio (PTR).

(1) STA/LTA detector: STA / LTA is easily implemented as the ratio of the mean square value of the signal in the short time window to the mean square value of the signal in the long time window. Using such a ratio, we can detect the pulse events in the short time within the long time windows.

$$STA_i = \frac{1}{N_S}\sum_{n=i}^{i+N_S-1} X^2(n) \text{ (Miralles, 2012)}$$

$$LTA_i = \frac{1}{N_L}\sum_{n=i-N_L}^{i-1} x^2(n)\text{(Miralles, 2012)}$$

, where N_S is the short time window and N_L is the long term window.

(2) PTR: The PTR of the segment of the discrete time signal $x_i(n)$ can be defined as the average power of the pulse component of the sound P_{pulsed} divided by the average power of the tonal component of the sound P_{tonal}. It can detect the events which is in a horizontal shape.

$$\text{PTR(dB)} = 10 \times \log\left[\frac{P_{pulsed}}{P_{tonal}}\right]\text{(Miralles, 2012)}$$

Which can be easily computed using the cosine transform as indicated in the following equation:

$$\text{PTR(dB)} = 10 \times \log\left[\frac{\sum_{u=0}^{N_1-1} F(u,0)^2}{\sum_{v=0}^{N_2-1} F(0,v)^2}\right]\text{(Miralles, 2012)}$$

1.2.2 Gillespie's Detection Method

In detection part we adopted the method used by Gillespie to detect the whistles in the spectrograms. This method provided by Gillespie is composed by 6 steps which are:

1) Click removal

Eliminate clicks from the signal in the time domain by first measuring the mean value m of the signal and the standard deviation (SD) of the signal *x (t)* in each 512 data blocks

$$X_i' = X_i \times W_i \text{ (Gillespie, 2013)}$$

$$W_i = \frac{1}{(1+\left(\frac{(x_i-m)}{thresh\times SD}\right)^{power})} \text{ (Gillespie, 2013)}$$

2) Spectrogram calculation,
3) Spectrogram noise removal,

   a) Median filter

   For each point in the FFT data, take 61 points (5.7 kHz) around the point (i.e. 30 on either side) and find the median of these 61 points. The median value is then subtracted from the original data. y_f is the power spectrum data at frequency bin *f*.

$$l_{y_f} = 10 \times \log 10(y_f) \text{ (Gillespie, 2013)}$$

$$l_{y'_f} = l_{y_f} - median\left(l_{y_{f-30}} : l_{y_{f+30}}\right) \text{ (Gillespie, 2013)}$$

   b) Average subtraction

   To remove constant tones from the spectrogram, calculate a running average background $\mathrm{b}_{t,f}$ at each time and frequency

$$b_{t,f} = \alpha \times l'_{y_{t,f}} + (1-\alpha) \times b_{t-1,f} \text{ (Gillespie, 2013)}$$
$$l_{y''_{t,f}} = l_{y'_{t,f}} - b_{t-1,f} \text{ (Gillespie, 2013)}$$

   c) Gaussian smoothing kernel

   The spectrogram is then smoothed by convolving it with a Gaussian smoothing kernel

$$l_{y'''} = l_{y''} \times G \text{ (Gillespie, 2013)}$$
$$\mathrm{G} = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix} \text{ (Gillespie, 2013)}$$

4) Thresholding,

   All data points in the spectrogram below this threshold (at least 8 dB) are set to 0. In fact, all that remains is a binary graph of the points above and below the point.

5) Connecting region search,

   Differently to what it was proposed, we used ‘bridge’ which is a kind of morphological operations in Matlab to connect the regions. Details are given in section 2.2.

6) Crossing, Merging, and branching region

   It is related to the whistle classification and due to the fact that the main task in this work is to detect accurately this part’s operation was discarded.

Capítulo 2. Improvement of New Detector

The enhancements of the new detector are focus on 2 scale, which are the larger accurate detection rate compared with the old detector and the quicker speed of the whole process (Figure 3).

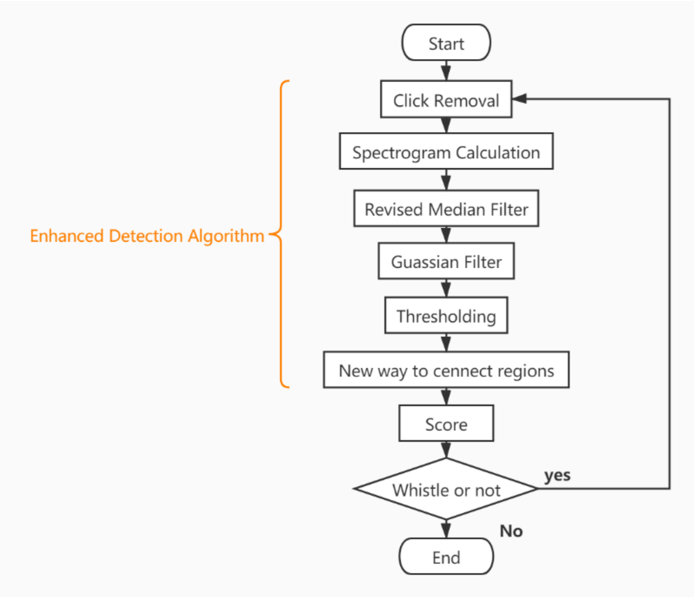


Figure 3: Proposed algorithm for the detection of whistles.

2.1 Revised Median Filter

We gave up the original coded median filter and chose to use the median filter from Math works but still used the form that calculate the median value based on 30 points in the previous time interval and next time interval of the point (30 points for each side). The spectrogram of the raw audio shows in Figure 4. The original median filter didn't consider the time so in the 0.5s segment, the amplitude value of the whole row is the same. According to Figure 5, we can see that the old detector can't recall whistles successfully. But the new median detector calculates the value of all pixels in the 0.5 seconds so we can see that the new detector accurately recovers the whistle contour after the de-noising steps (Figure 6). The purpose of applying median filter is to remove the broadband clicks which can influence the whistle detection severely. However, it can also remove the constant noise and some flat whistles.

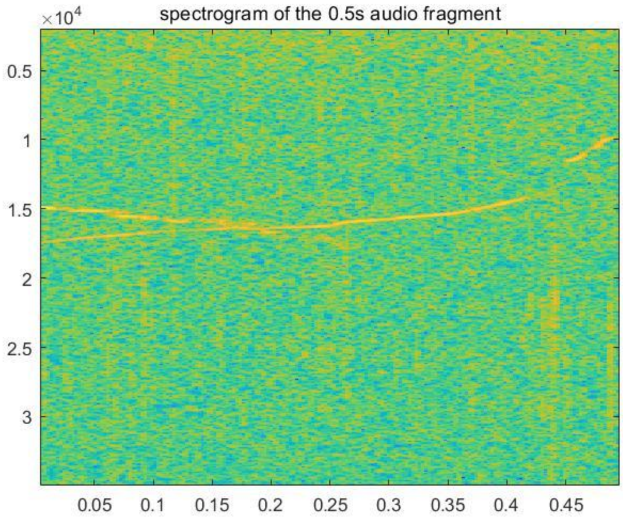


Figure 4: Raw Spectrogram without any de-noising step

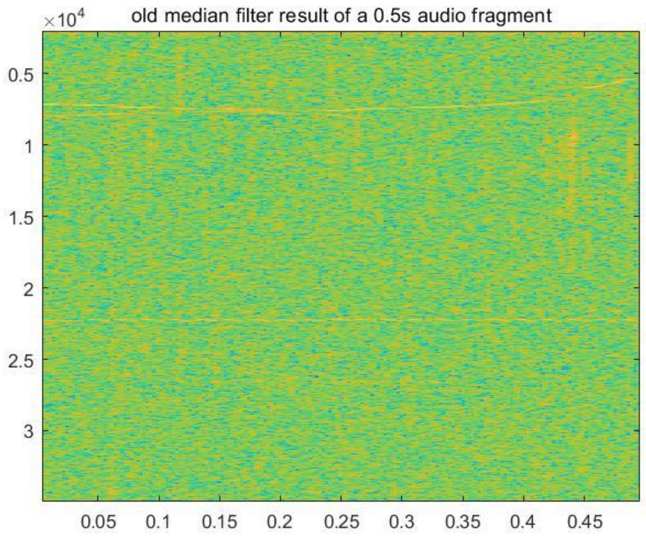


Figure 5: Median filtered by the old median filter

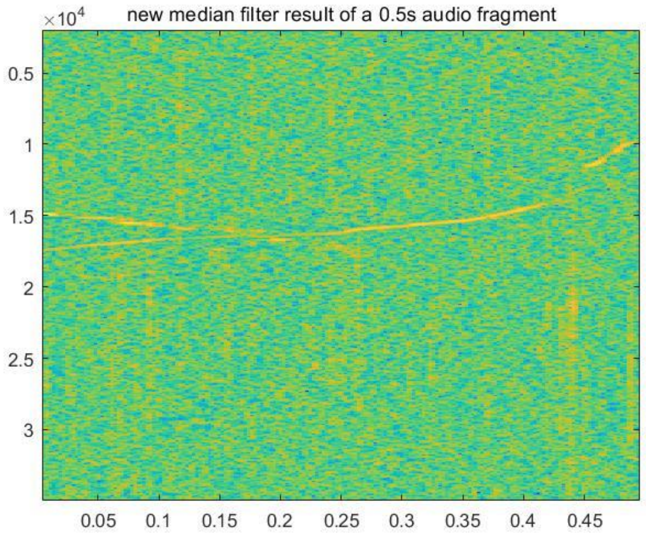


Figure 6: Median filtered by the new median filter

The following figures show how median filter remove the broadband clicks. The effect of removing broadband clicks can't be seen by the changes of spectrogram because the comparison between the background and the clicks are still obvious. Figure 7 and Figure 8 shows the spectrogram before de-noising and after de-noising. According to the 2 figures, the clicks in 0.1s and 0.45s are not removed in our naked eyes. But what happened is the de-noising steps to reduce the value of the spectral peaks in the clicks, makes value comparison between the background and the clicks smaller. We can see the effect of new median filter by the binary maps as Figure 9 and 10. In Figure 9, even with a stricter threshold (the threshold of the points can be counted as a peak), we can see the clicks in 0.45s clearly. However, after the median filter, we can see that the new binary map doesn't include any clicks.

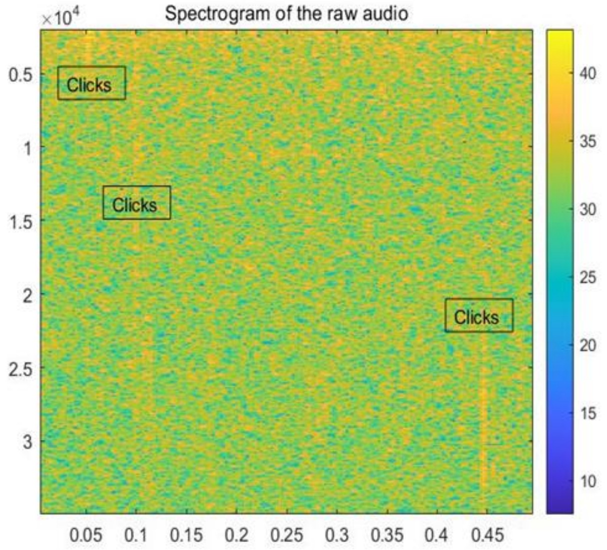


Figure 7: Spectrogram of the raw audio

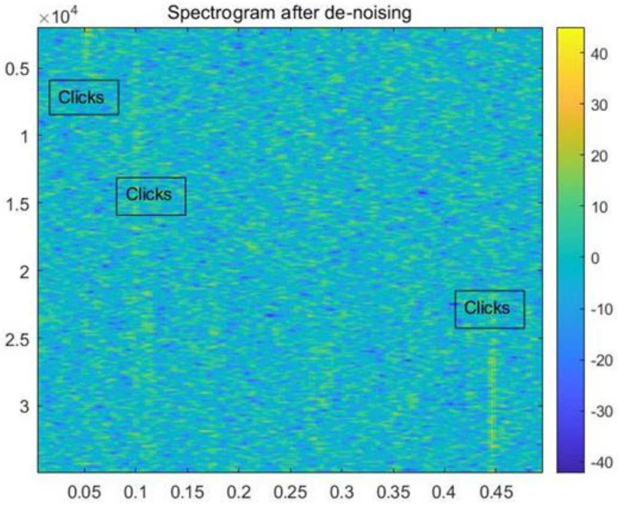


Figure 8: Spectrogram after de-noising

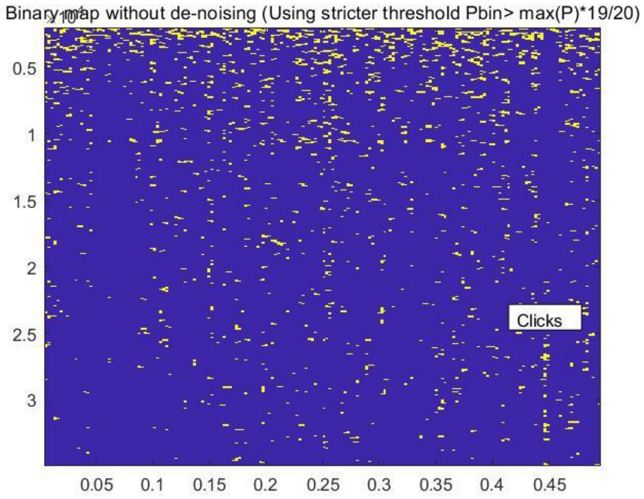


Figure 9: Binary map of the spectrogram without median filter

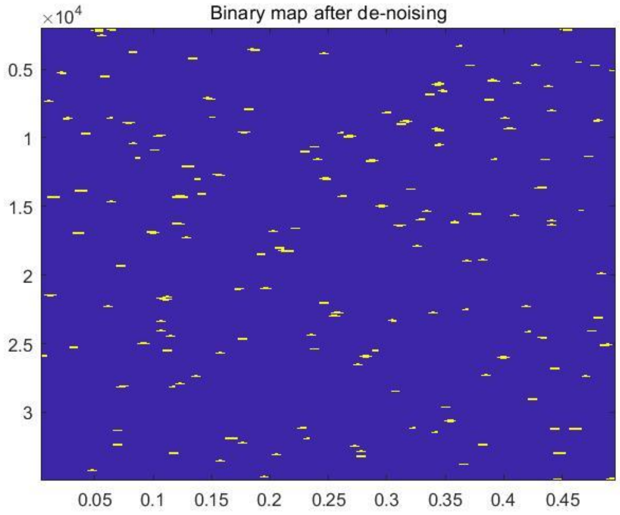


Figure 10: Binary map after de-noising

2.2 New Ways to Connect Regions

We used '*bwmorph'* in Matlab to connect the separate points after de-noising. In the beginning, we used operation named '*bridge*' to achieve the region connection. As one of the morphological operation, '*bridge*' can bridge unconnected pixels, that is, if they have two unconnected non-zero neighbors, set the pixel with a value of 0 to 1. But only with this operation, the result of region connection is not good enough, many pixels that are not apart with one pixels but more pixels can't be connected. So we tried a new set of operations. Firstly, use '*clean*' to remove isolated pixels, then use '*diag*' to eliminate 8-connectivity of the background, then use '*bridge*' and '*bwperim*' to get only the perimeter pixels of objects in the input image, finally, we used '*thicken*' to bold the object by adding pixels outside the object until doing so will cause previously unconnected objects to be connected by 8. The differences of using 2 sets of morphological operation are as Figure 11 and 12:

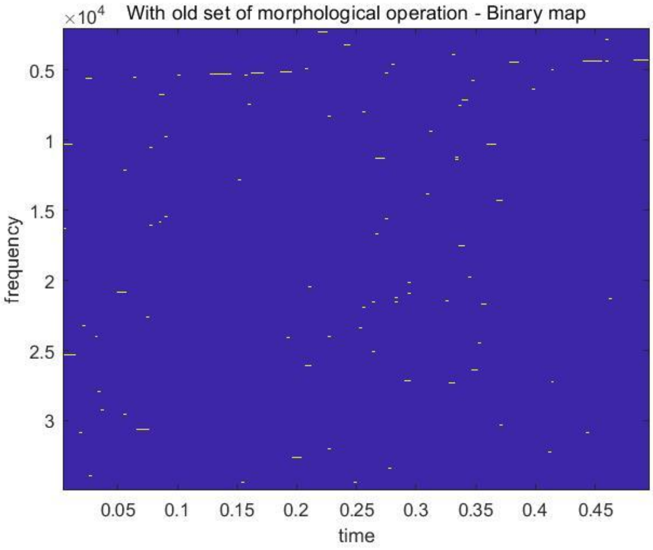


Figure 11: Use the old morphological operators operators

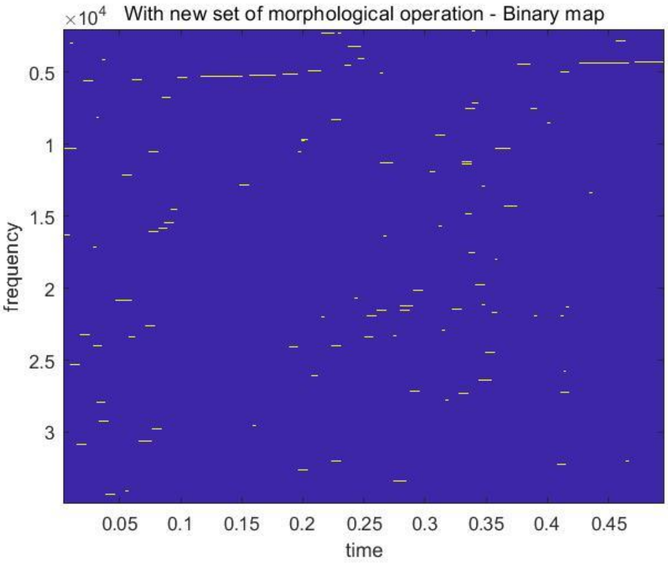


Figure 12: Use the new set of morphological

Another problem occurred in the detection process is that the complex noises in the marine environment such as pink noises and cause layered spectrogram. With noises like pink noises (See Figure 13), the step of click removal will introduce some fake clicks in the shallow layered part of the spectrogram (See Figure 14). When those fake clicks come intensely, they will be treated as a short whistle falsely (See Figure 15, Figure 16). To solve this problem, the new detector judges the orientation of connected region after getting all of those candidate contours in binary map and remove those connected regions of which the absolute value of the orientation is between 80 and 100 degrees. Also, to make sure the number of false alarming low in different kinds of noisy environment, we also adjust our threshold of a minimum duration of a whistle. This operation removes most of the false clicks and decrease the connectivity of those clicks (See Figure16). The false alarming in those marine areas that have strong pink noises like noises decreases from 98 to 1. Table 2 compares the results of the detector before adjustment in threshold and morphological operation (add the judgement of orientation) and the results after those adjustments.

Table 1: Comparison of the result before and after the adjustments

	Before	After
Detected Events	103	43
False Alarming	98	1

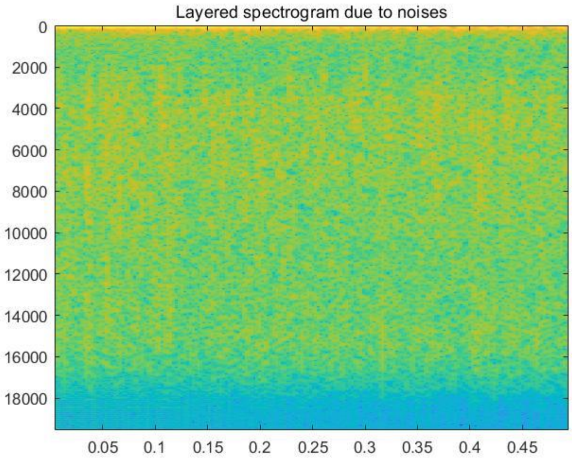


Figure 13: Spectrogram with pink noises

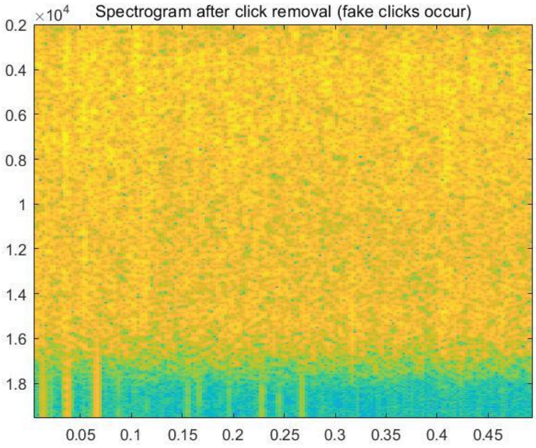


Figure 13: Spectrogram after click removal

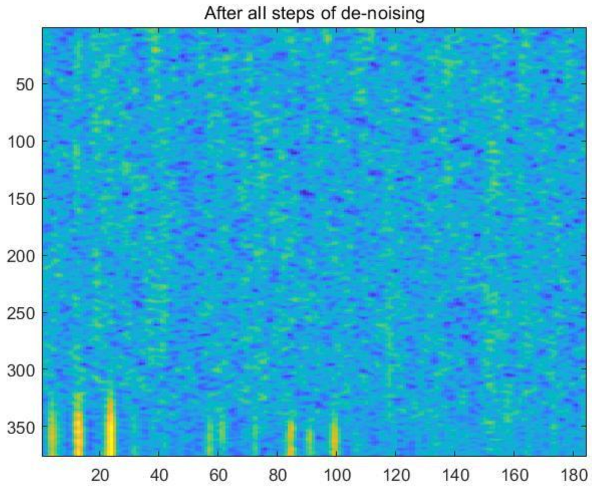


Figure 14: Spectrogram after de-noising

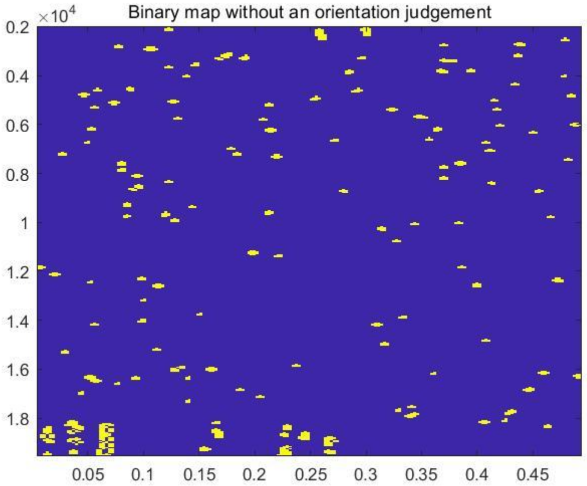


Figure 15: Binary map without orientation judgement

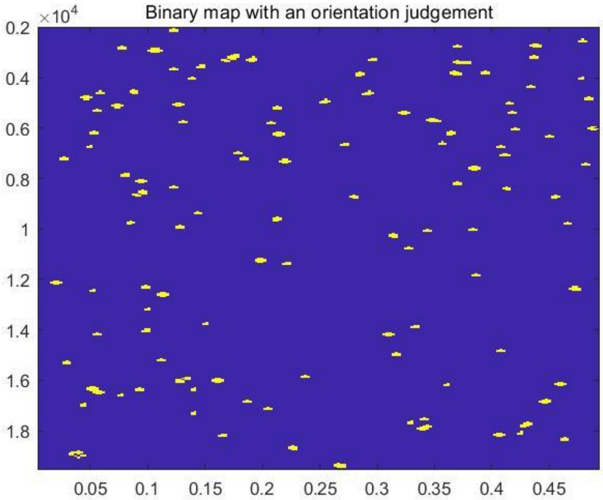


Figure 16: Binary map with orientation judgements

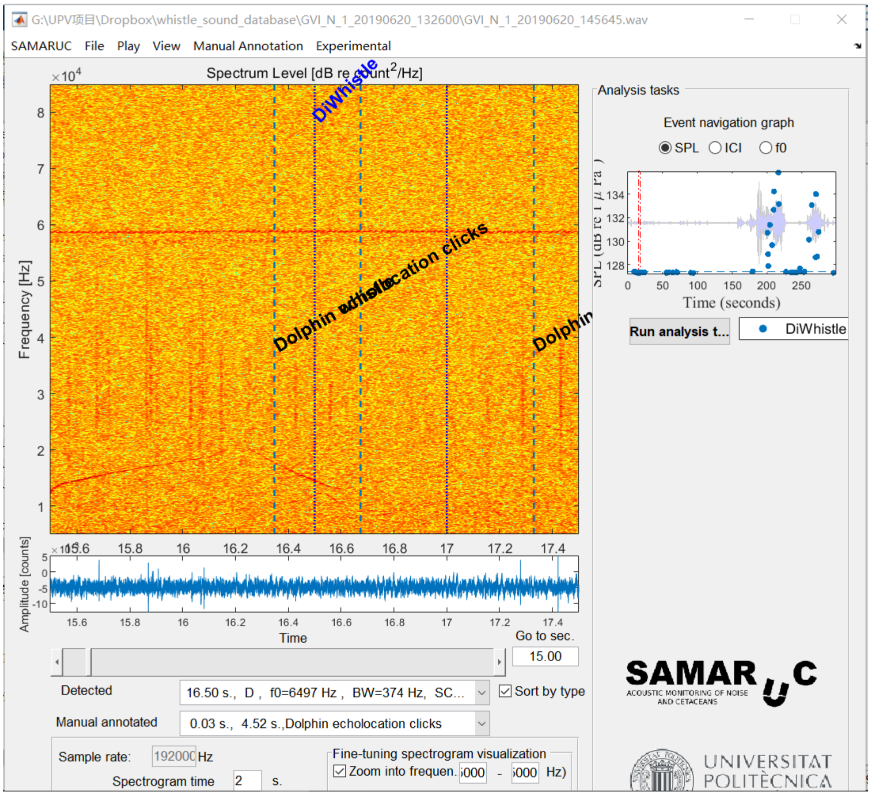
2.3 Complete A Whistle Contour

One of the biggest problems in the whistle detection is due to the fact we are analysing the recordings in time blocks of a given length. It is quite frequent that a whistle lasts for 2 or 3 blocks and the detector only detects the whistle in those blocks where the whistle has high intensity and longer duration. The weak and short parts (typically the start and end of the whistle) are ignored by the detectors. Our new detector solved this problem by detecting the previous and next segments around the whistles. After all whistles are detected, we locate those whistles' position in the raw audio, then find the previous and next segments around the whistles. We used the same algorithm to detect whether there exists a whistle or not but with a lower threshold in the time duration. Because the probability that these fragments include a whistle is relatively high. The lower threshold is adjusted according to the performance. The algorithm is described using pseudocode in the following table.

```
Algorithm:
  For all audio segments xi do
    If  xi.type=whistle;
      Position 1 = xi.start-window; Position 2 = xi;
      Position 3 = xi.end;          Position 4 = xi.end+window;
      Do whistle detection at new fragments of x(position 1,position 2)
          If  duration > durationThreshold
              x(position1,position2).type =whistle;
          end
      Do whistle detection at new fragments of x(position 3, position4)
          If  duration > durationThreshold
              x(position3,position4).type =whistle;
          end
      end
end
```

Some examples of the results are given in the Figures 17 and 18. The blue marks "Diwhistle" call out the whistles that are detected by the new detector. The black marks include "Dolphin echolocation clicks" and "Dolphin whistles" are the manual annotation. Marks that are manually annotated by human operators are the standard criteria of the accuracy in this paper. All the automatic detection results that are overlapped with the manual annotation are treated as accurate detection results. Those results that are only marked by the detector will be treated as false alarming.

It is quite frequent that a whistle lasts for 2 or 3 blocks and the detector only detects the whistle in those blocks where the whistle has high intensity and longer duration. The weak and short parts (typically the start and end of the whistle) are ignored by the detectors. Without the re-check algorithm which check the previous and next audio fragment of a detected whistle, the detector can only detect one part of the whistle (see the whistle in between 15.5 and 16.7 second in Figure 17). After applying the re-check algorithm, the new detector is able to complete a contour with previous block and next block (see the whistle in between 15.5 and 16.7 second in Figure 18).



G:\UPV项目\Dropbox\whistle_sound_database\GVI_N_1_20190620_132600\GVI_N_1_20190620_145645.wav
SAMARUC File Play View Manual Annotation Experimental
Spectrum Level [dB re count^2/Hz]
$\times 10^4$
Frequency [Hz]
DiWhistle
Dolphin echolocation clicks
Dolphin
Amplitude [counts]
Time
Analysis tasks
Event navigation graph
● SPL ○ ICI ○ f0
SPL (dB re 1 μPa)
Time (seconds)
Run analysis t...
DiWhistle
SAMARUC
ACOUSTIC MONITORING OF NOISE AND CETACEANS
UNIVERSITAT POLITÈCNICA

Go to sec. 15.00

Detected: 16.50 s., D , f0=6497 Hz , BW=374 Hz, SC... ☑ Sort by type

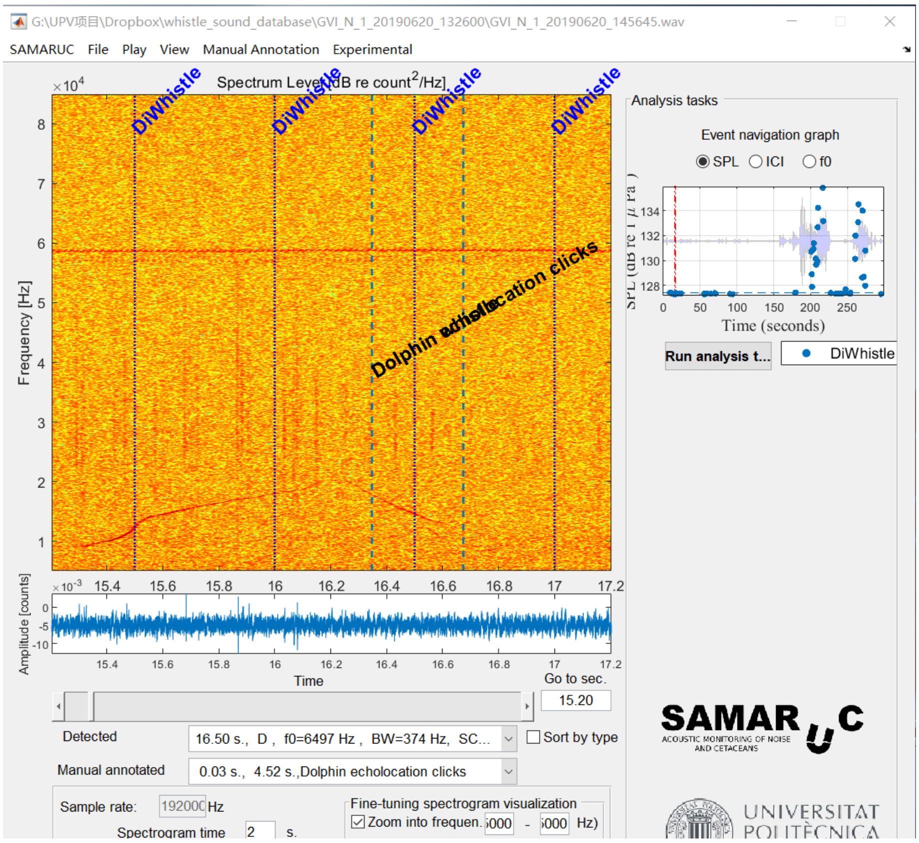
Manual annotated: 0.03 s., 4.52 s.,Dolphin echolocation clicks

Sample rate: 192000 Hz

Spectrogram time 2 s.

Fine-tuning spectrogram visualization: ☑ Zoom into frequen. 5000 - 5000 Hz)

Figure 17: Detected results from old detector



G:\UPV项目\Dropbox\whistle_sound_database\GVI_N_1_20190620_132600\GVI_N_1_20190620_145645.wav
SAMARUC File Play View Manual Annotation Experimental
Spectrum Level [dB re count^2/Hz]
$\times 10^4$
Frequency [Hz]
DiWhistle
DiWhistle
DiWhistle
DiWhistle
Dolphin echolocation clicks
Amplitude [counts]
Time
Analysis tasks
Event navigation graph
● SPL ○ ICI ○ f0
SPL (dB re 1 μPa)
Time (seconds)
Run analysis t...
DiWhistle
SAMARUC
ACOUSTIC MONITORING OF NOISE AND CETACEANS
UNIVERSITAT POLITÈCNICA

Go to sec. 15.20

Detected: 16.50 s., D , f0=6497 Hz , BW=374 Hz, SC... □ Sort by type

Manual annotated: 0.03 s., 4.52 s.,Dolphin echolocation clicks

Sample rate: 192000 Hz

Spectrogram time 2 s.

Fine-tuning spectrogram visualization: ☑ Zoom into frequen. 5000 - 5000 Hz)

Figure 18: Detect results from the new detector which detect the previous and next segments

2.5 Parameter Setting

To get as more whistles as possible, our new detectors decrease the value of the threshold generally. But at the same time, we also have to control the value of threshold so that there will not be many false alarming. The new threshold is adjusted according to the performance of detector. In the old detector, the minimum time duration of the contour is 75 ms. In the new detector, it changed to 45 ms and for those fragments that are the previous or next one of the detected whistles, the lower threshold is 30 ms. Also, we discard the high frequency part (Higher than 35000 Hz) and low frequency part (Lower than 2000 Hz) because dolphins' whistles are limited in the frequency domain of (2k Hz to 35k Hz) (May-Collado & Wartzok, 2008).

The pre-detector in the old detection system aims to reduce the processing time, but the effect of it is proved not very good. Because the detection results of whistle events are not a subset of the active events, actually, many whistle events are not detected as active events, so they were discarded in the pre-detection stages. Thus there is a large amount of whistles are missed in the old system. For the big cut off in dolphin whistles cannot be traded off by the time it decreases, the new detector removes the whole pre-detection part and begin with the formal detector directly. To get as many whistles as possible, our new detectors decrease the value of the threshold generally. But at the same time, we also have to control the value of threshold so that there will not be many false alarming. The new threshold is adjusted according to the performance of detector. In the old detector, the minimum time duration of the contour is 75 ms. In the new detector, it changed to 45 ms and for those fragments that are the previous or next one of the detected whistles, the lower threshold is 30 ms. As we did before we discarded the higher and lower (<2000 Hz) frequency parts (>35000 Hz).

The pre-detector in the old detection system aims to reduce the processing time, but the effect of it is proved not very good. Because the detection results of whistle events are not a subset of the active events, actually, many whistle events are not detected as active events, so they were discarded in the pre-detection stages. Thus there is a large amount of whistles are missed in the old system. For the big cut off in dolphin whistles cannot be traded off by the time it decreases, the new detector removes the whole pre-detection part and begin with the formal detector directly.

Capítulo 3. Result and Conclusion

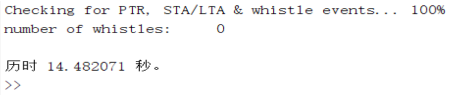
3.1 Results

3.1.1 Evaluation and Test

All the data of dolphin whistles we used comes from the Polytechnic University of Valencia (UPV) iTEAM research group database. The data collection has been obtained through several projects funded by the European Union (EU), by the Spanish Ministry and by the Spanish Institute of Oceanography. Data was acquired using a system entirely developed by the UPV- iTEAM researchers named SAMARUC. The recordings were done in different Marine protected areas in the Mediterranean and Cantabrian Seas. We have tested our algorithm in one database but because of COVID-19, our test based on another big database suspended. After revising some parts to adapt our algorithm to different noisy environment, we tested our algorithm in 2 subsets of 2 different algorithms. One is the subset of the database that we have already tested which names GVI_N_1_20190620, one is another database with different kinds of noise named GOZ_S_1_20180905.

3.1.2 Comparison Between Old and New Detector

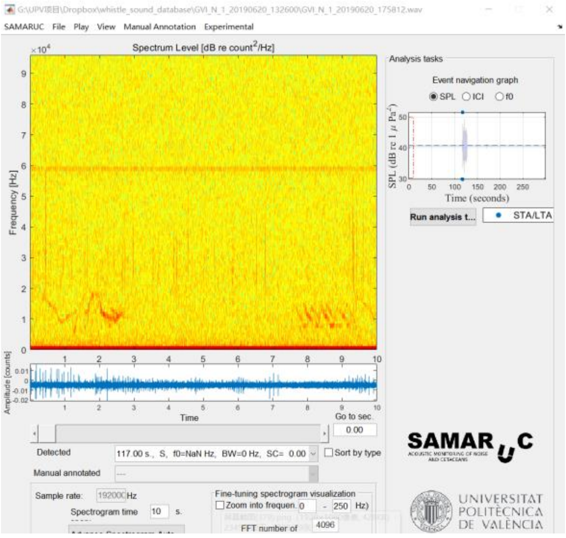
For the same 5 minute wav file, the old detector can only detect 2 active events and those 2 events are not whistles. The new detector got 402 fragments (0.5s each) have whistles. The processing time of old detector is 14.482071s, and the processing time of the new detector is 130.682379s (Figure 19, 21). According to the Figure 3, 5, this file does include many whistles, and only the new detector gets the matched results. To distinguish the whistles that are detected by the new detector, we called it as “DiWhistle”. The part in the top left of the panel shows the spectrogram of this file and we can drag the time bar beneath it to see the spectrogram in the whole 5 minutes. Using the zoom button and adjust the spectrogram time, we can zoom choose whether we want to see a more detailed figures in the time or frequency. In the top right part of the panel, we can see a small time amplitude figure of the file and we can see how the detected events distributed in it. As we said before, humans can easily distinguish the whistles in the spectrogram so we can see that in the first 10 second, there are many whistles, and the only the new detector detects them (Figure 5). The old detector can only detect 2 active events but no whistles (Figure 3).



```
Checking for PTR, STA/LTA & whistle events... 100%
number of whistles:     0

历时 14.482071 秒。
>>
```

Figure 19: Detection result of the old detector of a whole file



G:\UPV项目\Dropbox\whistle_sound_database\GVI_N_1_20190620_132600\GVI_N_1_20190620_175812.wav

Detected: 117.00 s., S, f0=NaN Hz, BW=0 Hz, SC= 0.00

Manual annotated: ---

Go to sec. 0.00

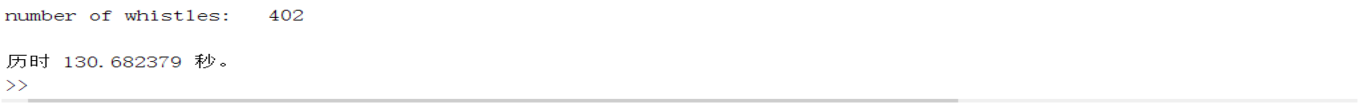
Sample rate: 192000 Hz

Spectrogram time 10 s.

Fine-tuning spectrogram visualization: ☐ Zoom into frequen. 0 - 250 Hz)

FFT number of 4096

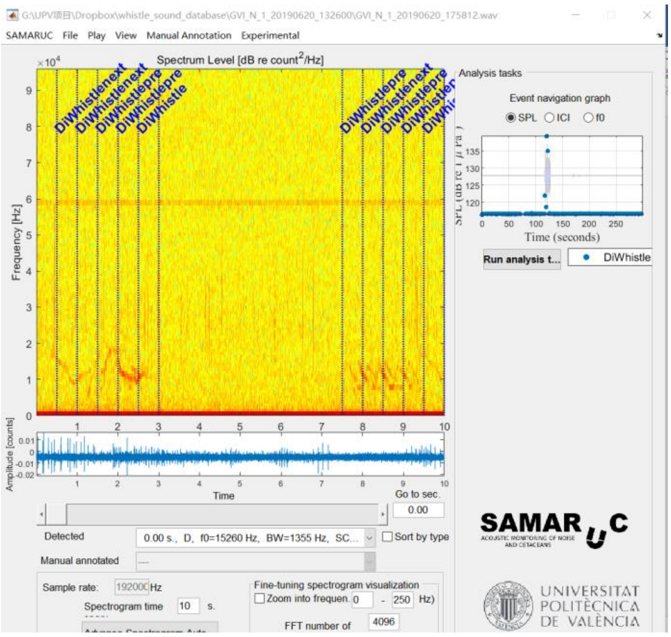
Figure 20: Detection result of the old detector (panel)



```
number of whistles:   402

历时 130.682379 秒。
>>
```

Figure 21: Detection result of the new detector



Detected: 0.00 s., D, f0=15260 Hz, BW=1355 Hz, SC...

Go to sec. 0.00

Sample rate: 192000 Hz

Spectrogram time 10 s.

☐ Zoom into frequen. 0 - 250 Hz)

FFT number of 4096

Figure 22: Detection result of the new detector (panel)

Table 3: Comparison between old and new detector

	Old Detector	New Detector
Detected Fragments	6	3110
False Alarming	3	72
Time Used (s)	358.8	2269.7

3.1.3 General Result

The recall rate of the new detector is 75.4%. It is estimated based on 24 wav files that have manual annotations (It is a difficult task to manually annotate all files in different database). Because those 24 files come from a same database named "GVI_N_1_20190620", the recall rate can variate in different databases. The precision rate of the new detector is calculated based on wav files come from 2 different databases, GVI_N_1_20190620 and GOZ_S_1_20180905. The precision rates of the formal database and of the latter are both 97.7%.

According to the data from Gillespie's (2013) paper (Gillespie, 2011), the overall recall and precision rates for the detector using his method were 88% and 79.6%. For Gillespie's detector, the recall rate ranging from 63.4% to 95.7% and precision from 52.8% to 96.9%. Compare with Gillespie's (2013) detector, our detector maintains an average performance in recalling and a high performance in precision.

Compare with the old detector, the recalling rate of the new one is 501.7 times higher than the old detector and the precision rate is 95% higher than the old one. The main 3 reasons why the old detector doesn't perform well is: 1. The pre-detector can't detect the active events well; 2. The old median filter didn't consider the changes in 0.5s' time interval; 3. The threshold is too big. Although there are some other key improvements that can help the detector work better, these 3 reasons are important to the bad performance of the old detector.

Regarding the complete GVI_N_1_20190620 database which includes 2342 wav files, we randomly choose 10 samples of files to see whether it detect most of the whistles or not. We got the results that this new detector can detect most of the whistles and give the indication that whether the audio file has whistle or not.

	Old Detector	Gillespie’s Method	New Detector
Recalling Rate	0.15%	63.4%~95.7%	75.4%
Precision Rate	50%	52.8%~96.9%	97.7%

Table 4: Comparison between 3 detectors

3.2 Conclusion

Automatically detect dolphin whistles in different marine environment with different noises is challenging because of the complex noises composition in the marine environment. In the detection process, generally, there are 2 main kinds of noises that need to be removed carefully. First one is the clicks which is generated by dolphins themselves, which is short duration transient noises and the second one is the constant tone last for a long time which is usually generated by shipping infrastructure. The way that has been proved effective to remove clicks is giving a weight for each signal data *Xi* and the weight is calculated based on the mean value and standard deviation among the nearby audio blocks which centred with the signal data *Xi* Median filter is efficient enough to remove the constant tone, broadband clicks and it can also enhance the peaks in the spectrogram. The median filter we use in the algorithm is a specific median filter which calculate the median value only in the horizontal lines. Using the original spectrogram to subtract the spectrogram that is filtered, those horizontal noises will be removed effectively. This 2 filters are the main filters to remove the noises and we use Gaussian filter then to smooth the spectrogram, which makes the contours more distinguishable.

The next important step of the detection process is connecting regions. After using threshold to produce a binary map of the spectrogram, we use a set of morphological operations on binary images. We try to connect points as more as possible because the intensity of a whistles varies and it reduces the connectivity of contours severely. We make the detector to understand what kind of connective regions should be treated as whistles by applying *regionprops* function in Matlab. The key to find the most proper operations is to use code to describe why humans think some of them are whistles and why we don’t think some of them are whistles.

Variable intensity of a whistle can also cause that we can’t detect a complete whistle which comes across a few blocks. We apply a re-check part in the new detector to check the previous and next blocks of all the whistles using lower threshold because the probability that whistles exist in those places are relatively high

Capítulo 4. Discussion

The time that the automatic detector needs to give a detection result is an important measurement of how practical the detector is. A complete database can include the audio resource of a marine region in a month, if the analysis time is longer than a month, no matter how accurate the detector is, it is not practical. In this paper, we set a fair standard time as the baseline, which is the time that a human operator needed to analyse the spectrogram and mark all the whistle contours. If the time that the detector needed is less than the operator, we treat it as acceptable. The less the time that the detector needed, the more practical the detector is. Typically, the most common way of reducing the processing time is adding a pre-detector, which detects active events and only sends those fragments (duration of this fragments are 0.5s) which contain active events to the detector. However, the inaccuracy of the pre-detector will impair the final detection results severely. The active events that detected by the pre-detector can't conclude all the whistles, which means, a number of whistles can't be detected as active events in the pre-detection part so it will not be detected by our formal detector.

To explore that how to use pre-detector to reduce the processing time and at the same time, to not impair the detection rate, we tried 3 pre-detection algorithms as follows:

1) PTR+STA/LTA

This is the first version of detector, it decreases the time needed largely but at the same time, it also decreases the detection rate. Many whistles can't be detected as an active event in this step. PTR detector will fail if the whistles are not very flat in horizontal line, because if the whistle is in the diagonal shape, the power in the horizontal axis and the power in the vertical axis will not have a big difference. Many of the whistles are not very flat, instead, they are in diagonal shape. The possible reason why STA/LTA detector can't catch the events include whistles is, if there are many whistles in both the short run windows and long run windows, the value of STA/LTA will not be bigger than the threshold (Miralles, 2011).

2) DESA

A more accurate detect ways but the time it needs to process is longer than the formal detector. The Discrete Energy Separation Algorithm (DESA) (Coutrot, 2013) is a model of auditory saliency based on temporal modulation of amplitude and frequency in multiple frequency bands. Multi-band demodulation analysis allows capturing such modulations in the presence of noise, which is often a limiting factor when dealing with complex auditory scenes. It uses a series of band pass filters to get the signals in different banks and then calculate Teager-Kaiser Energy in audio samples of each band. Next, get the average energy in each band, find the maximum of those averages, which is the mean Teager energy (MTE). At the same time, the mean instant amplitude (MIA) and the mean instant frequency (MIF) will be calculated, too. Combining MTE, MIA, MIF, we can get a value of saliency in this frame. Those frames have big saliency value will be treated as active events. The accuracy of this detect algorithm is very high, but the time it needed is very long. It needs nearly 30s to process a 5s audio files.

3) PTR+ revised STA/LTA

Revised STA/LTA means filter the raw audio signals with a band pass filter firstly and then use STA/LTA. The detection rate of this method even worse than the original version. Inspired by the DESA detector, we consider that maybe a band pass filter like Chebyshev filter can get the same effect and if we use it on STA/LTA detector, the time it needed will be less. In the experiment of comparing DESA and Chebyshev II filter using a 5s audio file, the result seems good, but when we applied it to a normal audio files, the detection rate even worse. The active events which include whistles are less.

The longest time for the detector without pre-detecting to analyse a database which includes a month time files are 5 days. The more the whistles in those files, the more time the detector needs. For the database which have the most whistles, the time is 5 days. It is relatively acceptable so we remove the pre-detector. But reducing the processing time is an important direction to improve the detection algorithm.

Capítulo 5. Bibliography

Gillespie, D. *et al*. (2013). Automatic detection and classification of odontocete whistles. *Acoustical Society of America*, 134 (3), p.2427.

Au, W. W. L. (1993). The Sonar of Dolphins. New York, *NY: Springer New York*. Retrieved from http://dx.doi.org/10.1007/978-1-4612-4356-4

Fristrup, K. M., Hatch, L. T., & Clark, C. W. (2003). Variation in humpback whale (Megaptera novaeangliae) song length in relation to low-frequency sound broadcasts. *Acoustical Society of America*, 113(6), 3411.

Gruden, P. & White, P. (2016). Automated tracking of dolphin whistles using Gaussian mixture probability hypothesis density filters. *Acoustical Society of America*, 140, pp. 1981-1991.

Miralles, R. *et al*. (2012). The pulsed to tonal strength parameter and its importance in characterizing and classifying Beluga whale sounds. *Acoustical Society of America*, 131(3) pp.2173-2179.

Miralles, R. *et al.* (2017). On the detection of impulsive and tonal events in passive acoustics monitoring. In: 22nd International Conference on Digital Signal Processing (DSP). August, 2017. Piscataway, NJ.

Mallawaarachchi, A. *et al*. (2008). Spectrogram denoising and automated extraction of the fundamental frequency variation of dolphin whistles. *Acoustical Society of America*, 124(3), pp. 1159-1170.

Herzing, L. (2016). Dolphin Communication and Cognition: Past, Present, and Future. pp.328.

Roch, M. *et al.* (2011). Automated extraction of odontocete whistle contours. Acoustical Society of America, 133(4), pp. 2212-2223.

Ioana, C. *et al.* (2010). Analysis of underwater mammal vocalisations using time-frequency-phase tracker. *Applied Acoustics*. 71.

Coutrot, A. *et al*. (2014). Video viewing: Do auditory salient events capture visual attention. *Annals of Telecommunications*, 169(1), pp. 89-97.

May-Collado, L. & Wartzok, D. (2008). A Comparison of Bottlenose Dolphin Whistles in the Atlantic Ocean: Factors Promoting Whistle Variation. Journal of Mammalogy, 89(5), pp1229-1240.

MarineTraffic, 2020. What Exactly Is The Density Map Layer? [image] Available at: <https://help.marinetraffic.com/hc/en-us/articles/204103758-What-exactly-is-the-Density-Map-Layer-> [Accessed 28 April 2020].