DETECTION AND CLASSIFICATION OF MARINE MAMMALS USING AN LFAS SYSTEM

S.P. van IJsselmuide and S.P. Beerens
TNO Physics and Electronics Laboratory,
P.O. Box 96864, 2509 JG The Hague, The Netherlands, E-mail: Beerens@fel.tno.nl

ABSTRACT

World wide a concern is emerging about the influence of man-made sound in the sea on marine life, and particularly about high power active sonars systems. Most concern lies with marine mammals, which fully depend on sound in their natural behaviour (foraging, navigation and communication). One of the sonars under debate is the Low Frequency Active Sonar (LFAS). This type of system is designed for long range detection of submarines. It consists of a powerful source and a towed array receiver. Incidents with marine mammals could be avoided if the receiver that is dedicated to detection of submarine echoes, is equipped with Detection, Classification and Localisation capabilities for marine mammals as well.

In this paper the development of a prototype transient detector and classifier for the TNO-FEL LFAS array (named CAPTAS) is described. A broadband beamformer is developed that creates 8 beams (sectors) that are equally wide over the whole frequency band. A multi-beam LOFAR display is presented. On the normalised data a Page's test detector is applied that is "optimum" for signals with unknown duration. Detected transients are sent to a classifier that tries to discriminate between biological and man-made or natural transients. Time-frequency analysis is performed and in the resulting time-frequency plot structures are determined by means of cluster analysis after which the sound is classified. Detection results of the prototype are very good, the Classification module is under development and the Localisation module is part of future research. Part of this research is sponsored by the Royal NetherLands Navy (RNLN).

RÉSUMÉ

L'impact des sons d'origine artificielle sur l'écosystème sous-marin soulève un intérêt mondial croissant. Plus particulièrement, cet intérêt se porte sur l'impact des systèmes sonar actifs à forte puissance sur les mammifères marins, dont le comportement est entièrement basé sur l'utilisation du son (aussi bien pour s'alimenter, s'orienter ou communiquer). Un des systèmes sonar concernés est le Sonar Actif à Basses Fréquences (LFAS). Ce type de système est conçu pour la détection longue distance de sous-marins. Il consiste généralement d'une source puissante et d'une antenne de réception remorquée. Les accidents causés par l'interaction de ces systèmes sur les mammifères marins pourraient être évités si l'antenne réceptrice dédiée à la détection d'échos de sous marins était munie de capacités de Détection, Classification, et de Localisation (DCL) des mammifères marins environnants.

Cet article décrit un prototype de détecteur/classificateur de transitoires développé pour l'antenne LFAS de TNO-FEL, l'antenne CAPTAS). Un algorithme de formation de voies est appliqué sur l'ensemble de la bande de fréquence, créant 8 voies de largeur égale (secteurs). Une visualisation LOFAR multi-voies est alors proposée et les données normalisées sont soumises à un détecteur élaboré a partir d'un test de Page, optimal pour les signaux de durée indéterminée. Les transitoires détectés sont transmis à un classificateur qui tente de discriminer les signaux suivant leur origine biologique, artificielle ou naturelle: après une analyse temps-fréquence, les images obtenues sont soumises à une analyse de clusters. Les structures temps-fréquence résultant de ce traitement permettent alors de classifier le son précédemment détecté. Les résultats de détection sont excellents, le classificateur progresse rapidement et le développement d'un algorithme de localisation est amorcé. Cette recherche est en partie sponsorisée par la Marine Royale Néerlandaise (RNLN).

1. INTRODUCTION

World wide a concern is emerging about the effects of anthropogenic (man-made) noise in the marine environment. At present most concern lies with marine mammals [1], [2], [3], [4], [5], *i.e.* cetaceans (whales and dolphins) and pinnipeds (seals, etc.), but there is also an increasing interest in effects on sea turtles and fish [6].

1.1 The problem

Marine mammals fully depend on sound in their natural behaviour (foraging, navigation and communication). For these animals, knowledge of the physiological effects of anthropogenic noise on the auditory system is more developed than for 'lower' animal species. However, precise knowledge on acoustic disturbance and/or injury of marine mammals is still very limited, and the same holds for detailed information on marine mammal hearing systems. Intense sound can have severe negative effects on marine animals. The effects may vary between 'audible', via 'change in behaviour' and 'severe disturbance' up till 'hearing injury/death'.

Figure 1: A stranded juvenile Fin whale, found in North France by W.C. Verboom (photo by M. Verboom).

One of the sonars under debate is the powerful low-frequency sound source of Low Frequency Active Sonar (LFAS) systems. Besides a sound source, these systems also consist of a towed array receiver. Incidents with marine mammals could be minimised or even avoided if this receiver is equipped with Detection, Classification and Localisation (DCL) capabilities for marine mammals.
The development of a prototype marine mammal detector and classifier is described in this paper. A first version of the detector was already used during the combined TNO-FEL/NURC ADULTS 2003 trial in the Mediterranean, where many whales and dolphins were encountered. Using passive acoustic monitoring as developed in this project, together with adequate mitigation measures should minimise the impact of LFAS on marine mammals.

1.2 Mitigation measures: the solution?

It is clear that mitigation procedures to reduce the impact of anthropogenic noise are at least recommendable to protect marine life. Also in the Netherlands mitigation measures are defined for active sonar. These procedures aim to prevent any damage in the hearing system of marine mammals in the vicinity of military sonar equipment. Three types of measures are commonly applied in mitigation procedures:

Careful mission planning: Before planning a mission in which an active sonar is operated, it is verified whether the area is inhabited by marine mammals in that season.

Monitoring of marine mammals in the best possible way before using the sonar: Not using the sonar if marine mammals are present is a very efficient mitigation measure. But, how do we know whether marine mammals are present? Marine mammals can be monitored in two ways:

- *Visual* monitoring can be successful, but it is problematic at night (although the use of infra-red is considered). Besides, marine mammals spent most of their hours underwater, hidden from eyesight. At high sea-state even a well-trained whale watcher can easily misses a sighting.
- *Passive sonar* can help to detect and is probably the most promising monitor. However, not all species of marine mammals produce sound (some types of pinnipeds), while other mammals produce sound outside the frequency band of the sonar (Cuvier's beaked whales). It is not known if all endangered species vocalise. Moreover, passive sonar does not (directly) provide the animal's range, which is important in all mitigation measures.

Ramp-up schemes to scare marine life away: Slowly raising the source level, so that the animal can swim away and keep the received sounds to acceptable levels (well below the *temporary threshold shift* level of the animal) may work. It prevents the mammal from being ear-damaged; however, it may still impact on the animal's natural behaviour.

Currently at TNO-FEL a tool (named SAKAMATA [7]) is under development that supports all three mitigation measures. Additionally this tool supports the sonar operator with passive acoustic monitoring.

1.3 Outline of this paper

In this paper the development of a transient detector with prototype classifier for the LFAS array of TNO-FEL (named CAPTAS) is described. In Section 2 a broadband beamformer is discussed that creates 8 beams (sectors) that are equally wide over the whole frequency band. Results are presented on a multi-beam LOFAR display. On the normalised data a power-law/Page's test detector is applied that is robust for signals with unknown frequency content and duration.

Detected transients are sent to a classifier that tries to discriminate between biological and man-made or natural transients. The proposed method is based on pattern recognition in the time-frequency plot (which is a visualisation of the time-frequency distribution). The time-frequency distribution of a transient signal gives valuable information on the nature of the signal. Its bandwidth, duration and other spectral and temporal characteristics can be derived from the time-frequency plot, from now on denoted as *tf*-plot. Other common names for the *tf*-plot are "spectrogram", "LOFAR-gram" or just "gram".

Section 3 describes how time-frequency analysis is performed by means of conventional short-time FFT processing. In the resulting *tf*-plots, structures are determined by means of image processing (clustering). Dedicated cluster analysis classifies the sound as biological or mechanical. In the former case, it is also specified whether the mammal is large or small and whether it is a baleen whale or toothed whale.

2. DETECTION OF MARINE MAMMAL TRANSIENTS

Detection of marine mammals within the *danger-zone* of a sonar system is essential in avoiding exposure of those marine mammals to high-level sounds. The danger-zone is defined as the area where receive levels on animals are higher than "acceptable". What is "acceptable" in this respect is still under heavy debate, but more and more legislation is formulated. The radius of the danger-zone strongly depends on the hearing sensitivity of the species present and on the used sonar source. Typically danger-zones have a range on the order of 0 to 5 nautical mile.

The idea is that active sonar systems should have a sub-system that warns the operator for the presence of marine mammals within the danger-zone. The problem with detecting marine mammals is the wide range of species, where each species produces different sounds with different duration, frequency band and source level. For example, the very large baleen whales produce low frequency calls, around 10-20 Hz, which can last for several minutes. The much smaller porpoises produce very short clicks in the order of less than a millisecond and frequencies up to 160 kHz. As an example, *tf*-plots are shown in **Figure 2** for a Humpback whale and a Bottlenose dolphin. Making a robust detector for all these different sounds is a challenging task.

Several papers and reports are of direct relevance to the current work on the passive detection of marine mammals or the more generic problem of transient detection. A TNO paper [8] describes a transient detector developed for the Active Low Frequency (ALF) towed array. The detector was based on energy detection in a *tf*-plot. Another, very interesting report [9] and paper [10] from the NATO Undersea Research Centre (NURC) describes the combination of a power-law integrator [11] and a Page's test [12] for the passive detection of marine mammals. The power-law integrator is robust against varying signal bandwidth and the Page's test detector is a robust detector for signals with an unknown duration. This seems to be a very useful method for detecting marine mammals with their wide variety of sounds.

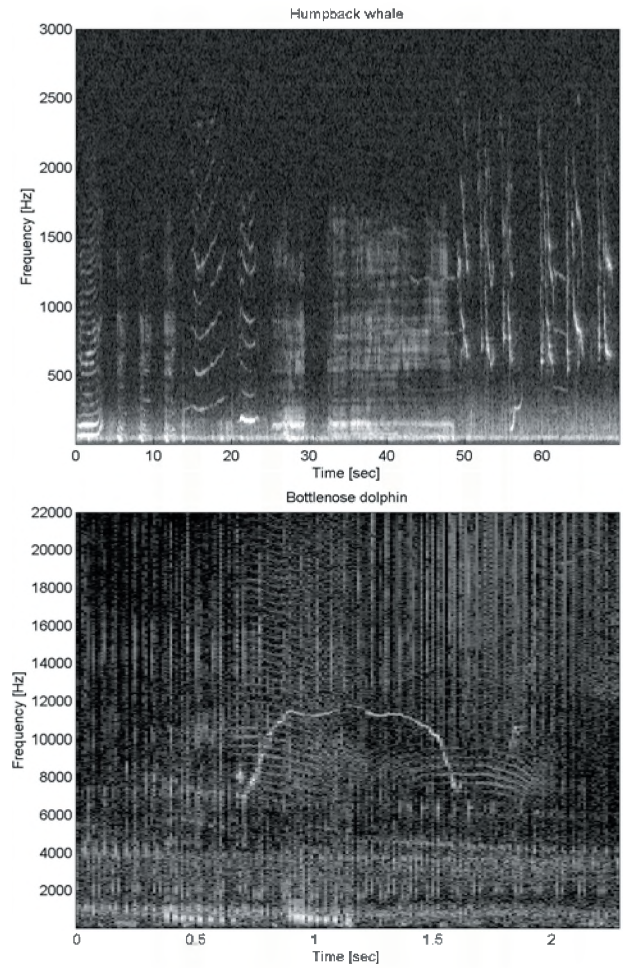

Figure 2: tf-plots of a Humpback whale (top) and a Bottlenose dolphin (bottom); note the very different frequency and time scales.

The remainder of this section describes the development of a prototype marine mammal (or more generic transient) detector for the CAPTAS towed array. The CAPTAS array is a modern LFAS receiving array with a capability for instantaneous left-right discrimination through the use of hydrophone triplets. The array consists of 64 triplets and operates in the frequency band from 10 to 2080 Hz; see [13] and [14]. This detector is strongly based on the already available processing software and structure developed for Anti Submarine Warfare (ASW) tasks (active/passive detection of submarines). A small schematic overview of the proposed marine mammal detector is shown in ***Figure 3***.

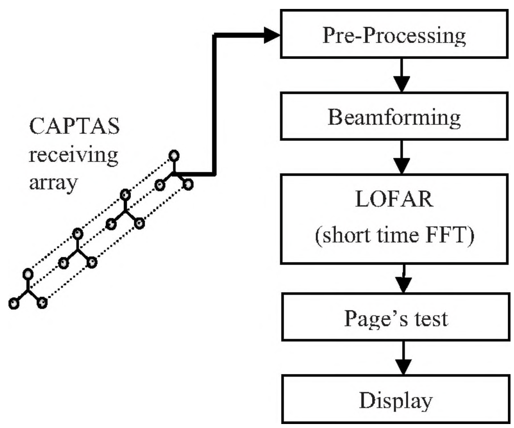

Figure 3: Structure of the proposed marine mammal detector.

The two major processing steps, pre-processing-beamforming and the detection processing, are described in Sections 2.1 and 2.2 respectively. This section ends with a short performance evaluation, based on recorded dolphin sounds and low frequency clicks, in Section 2.3.

2.1 Pre-processing and Beamforming

Pre-processing is the preparation of the hydrophone data so that they can be beamformed. It consist of the following steps:

- Detection and reparation of malfunctioning hydrophones,
- Roll stabilisation of the triplet structure,
- Fourier transformation of the hydrophone signals into the frequency domain.

These pre-processing steps are the same as used in ASW processing.

The beamformer developed for the marine mammal detector is rather different from the beamformer used in the ASW processor. Beamforming, the coherent summation of hydrophones signals, is normally used to improve the Signal-to-Noise Ratio (SNR) resulting in better detection of sound sources. Furthermore, it provides information on the bearing (direction of arrival) of the sound source.

For the detection of marine mammals we are not overly interested in maximising the SNR, since marine mammals generally make more noise than submarines and are relatively easy to detect. However, information on the (horizontal) direction of the target is very useful, *e.g.* for cueing visual observers in the right direction, and is essential for the choice of an appropriate mitigation procedure. When for instance mammals are detected in the forward sectors they will probably close in on the sonar system. This calls for other actions than when the mammals are in the aft sectors where the distance between sonar and mammal increases in time.

Beamforming is therefore an essential step to start with. However, a complicating factor is the wideband nature of the signals, which cover the total frequency band of the CAPTAS array (10-2080 Hz). Applying a straightforward Delay and Sum Beamformer (DSBF) to this frequency band results in a frequency dependent angular resolution [15]. This has several practical drawbacks like the large number of beams that have to be made at the higher frequencies, while at lower frequencies these beams will overlap. Furthermore, a large number of beams will require a highly automated detection process since the amount of beamformed data will be far too large to be presented on a screen.

A proposed solution for these problems is the use of a constant beamwidth beamformer. This beamformer has a frequency dependent array shading which keeps beamwidth constant for all frequencies. Mathematically, beamforming can be seen as a spatial filter, to which standard filter theory applies. To achieve constant beamwidth, a complex-Remez filter design algorithm [16] was used to compute the filter coefficients (equivalent to array shading coefficients) for each frequency. Applying these filter coefficients results in a constant beamwidth output, as shown in ***Figure 4***.

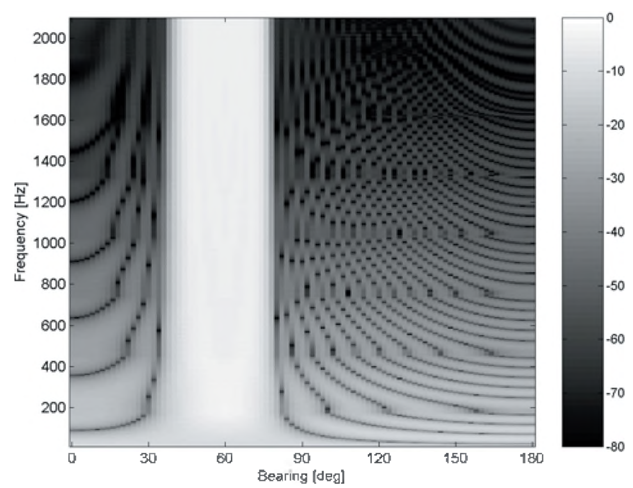

Figure 4: Constant beamwidth beamformer response for the total frequency band and a sector at 60°.

In this case a synthetic beam at 60° has been made with a constant beamwidth for all frequencies between 300 and 2080 Hz. A drawback of this method is the increasing sidelobe levels at lower frequencies. For frequencies below 300 Hz the filter coefficients are all set to one, to avoid excessive sidelobes. This changes the beamformer into an ordinary DSBF, as a consequence the beamwidth starts to increase for frequencies below 300 Hz.

The desired number of beams (look directions) is a compromise between the desired beam resolution, performance and display properties. Initially four beams are formed. In the consequent triplet processing the Port-Starboard (PS) ambiguity is solved, see [13] and [14], and number of beams is doubled. Finally, eight beams are made directed in eight compass directions: 0°, 60°, 90°, 120°, 180°, 240°, 270° and 300°.

The output of this newly designed beamformer is used as an input for the Page's test detector, treated in more detail in Section 2.2.

2.2 Page's test detector

After frequency-domain beamforming the eight beams are converted back to the time-domain by means of an Inverse Fourier Transform (IFFT). To each of these beams a transient detector is applied. ***Figure 5*** shows an overview of the detector. The detector consists of a power-law integrator and a Page's test, which seems to be a good combination for detecting the capricious marine mammal signals. This algorithm is very well described in [9] and this section is largely based on this report.

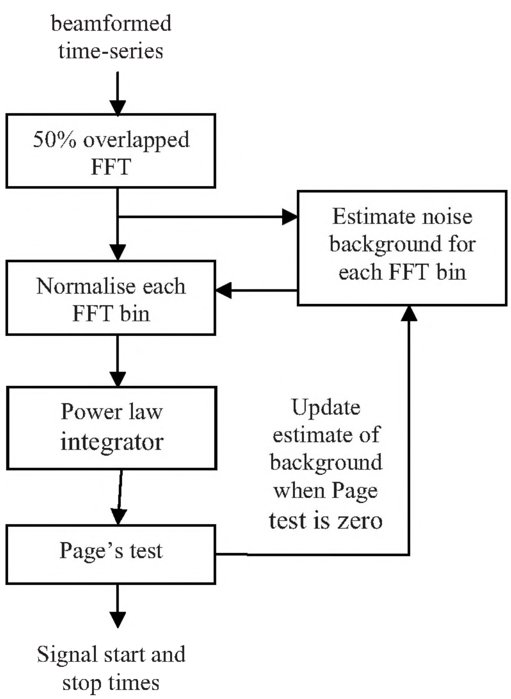

Figure 5: Block diagram of the detection scheme.

Note that in this application the Power law/Page's test detector is applied to beamformed data, but it can be applied to any time-series. In fact it can be applied to almost any receiving system. A good example is the application to sonar buoys in [17]. The following "walk-through" of the detector is based on one channel with acoustic data in the time domain.

As a first step the time-domain data are converted to the frequency domain by means of 50% overlapped short-time Fourier transforms (STFT). The integration time for the Fourier transform is always a compromise between spectral and temporal resolution. Generally a high temporal resolution (short integration time) will improve the detection of short signals like click (bursts) and sweeps.

One of the most important steps in the detector is the estimation of the background noise and interference for each frequency bin. The background consists of (wideband) ambient noise and (narrowband) shipping noise. As shown in ***Figure 5*** the detector exploits the Page's test to isolate data that is believed to be signal free. These data are then exponentially averaged over time using the following equation.

$$\lambda_{t+1} = \alpha\lambda_t + (1-\alpha)X_t \qquad (1)$$

In this equation λ_t is the old and λ_{t+1} the updated estimate of the background, α the time constant for the exponential averager and X_t the latest signal free power spectrum.

The following two steps are the actual normalisation of the power spectrum and application of the power-law integrator to the normalised spectrum. These two steps are shown together in a detailed overview of the Page's test as shown in Appendix A. The power law integrator sums the normalised frequency bins to a scalar, which is an indication of the energy level (for p=1 it is the energy). After a proper normalisation and in a *noise only* case this sum (denoted by Z in Figure) is approximately zero, while during a *signal present* case this sum is positive. Here several thresholds (h_0, h_1, b_0 and b_1) start to play a role. The used Page's test has separate thresholds for the onset detection of the signal (h_0) and the termination detection of the signal (h_1). Associated to these thresholds are biases in order to reduce the sensitivity (b_0 and b_1).

Based on trial and error the bias, threshold and power law parameters were set with the following values:

p - power law (p=2)
h_0 - threshold for start of signal detection ($h_0 = 8$)
b_0 - Page's test bias for start of signal detection ($b_0 = 2$)
h_1 - threshold for end of signal detection ($h_1 = 10$)
b_1 - Page's test bias for end of signal detection ($b_1 = 3$)
α - time constant for exponential averager ($\alpha = 0.99$)

These values were "optimised" for a proper operation on the experimental data evaluated, see Section 2.3. For different receiving systems and/or environments other settings might work better. As a rule of thumb the following guidelines can be used for optimising the different parameters:

- Low values for the power law (p) make the system relative more sensitive to wideband signals, while higher values make the system sensitive to narrow band signals.
- A high value for the power-law (p) and low values for the bias (b_0 and b_1) can make the system very sensitive to the small noise bursts that are always present in the underwater environment.
- Increasing the bias and threshold values decreases the sensitivity so that a signal needs a higher SNR to be detected.

For the current application, the detection of marine mammal vocalisations within the sonar danger-zone, SNR is not a real problem. Therefore, relative high values for the biases and thresholds are chosen.

2.3 Performance evaluation

In this section the previously described beamformer and detector are tested on two recorded marine mammal vocalisations. Both recordings were made using the TNO-FEL CAPTAS triplet array. The first recording has been made during an ASW-LFAS trial in 1999 near the Spanish coast of La Coruña and consists of several "high" frequency dolphin sweeps. This trial was conducted in co-operation with the Royal NetherLands Navy (RNLN) and Thales Underwater Systems (TUS). The second recording consists of several low frequency clicks and was made during another ASW-LFAS trial in the Autumn of 2003 near the coast of Sardinia. This trial was conducted in co-operation with the RNLN and NATO Underwater Research Centre (NURC). The recordings represent two different signals (high frequency opposed to low frequency and different characteristics) and are therefore very suitable for testing the marine mammal detector.

Both trials (and other trials performed in the intervening period) were dedicated ASW trials with the focus on testing new active sonar concepts. During these trials marine mammals were only rarely seen and even more rarely recorded with our towed array. Actually the two presented signals are, up to now, the only known marine mammal vocalisation recorded with the CAPTAS array.

The presented results in the following two subsections are intended to illustrate the functioning of the described beamformer and detector. In the future, a more thorough investigation on the detection performance has to be made, preferably with data from a dedicated marine mammal trial.

2.3.1 Detection of dolphin clicks

During an LFAS trial in 1999 several common dolphins approached the towing vessel within visual range, see ***Figure 6*** for a picture of the dolphins. An example of the acoustic recording made during this approach is shown in Figure 7. The upper panel shows a time-series of a single hydrophone containing several dolphin sweeps and an array artefact, the short and high peak around the 15 second time stamp. The lower panel depicts a *tf*-plot of the same recording. The dolphin sweeps are clearly visible in the upper frequency band (1000-2500 Hz). Also visible are some tonals from the tow ship, the horizontal lines.

Figure 6: Picture of the common dolphins that approached the sonar during the 1999 trial.

Applying the special beamformer of Section 2.1 reveals the direction of the vocalisation. Furthermore, beamforming rejects noise from directions other than the look direction. This is especially helpful in suppressing the tow ship noise, which often dominates the background noise levels.

The output of the beamformer is shown in Figure 8. This figure is a so-called "multi-beam LOFAR". For each of the eight beams a tf-plot is shown with frequency on the horizontal axis and time on the vertical axis. The axes are rotated to make the display look like a more standard LOFAR gram (waterfall) as used in passive sonars.

The tow ship noise is clearly visible in the Northern direction with several loud tonals that also leak into the other directions. Figure 8 depicts a sub-set of the single hydrophone data shown in Figure 7. The dolphin sweep is clearly visible in the Southern direction. This sweep is also weakly visible in the other directions (leakage through the sidelobes) together with some low frequency rumbles.

Figure 8 shows the intermediate result after pre-processing and beamforming. The next step is the normalization of the beamformed data and application of the power-law/Page's test detector for automated detection and extraction of the signals. The result after normalization is shown in Figure 9. This figure has the same set-up as Figure 8, *i.e.* a multi-beam LOFAR. The difference is the normalization, which equalizes the stationary background noise. Signals (fluctuations in the background) are now clearly visible. The dolphin sweep in the southern direction has been particularly clarified.

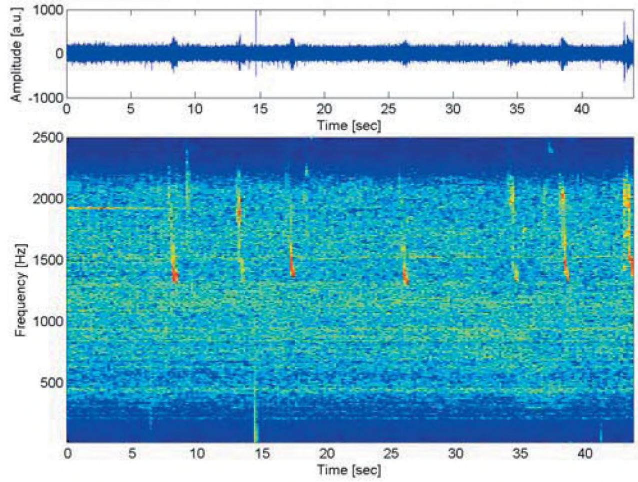

Figure 7: Time-series and tf-plot of single hydrophone data with several dolphin sweeps.

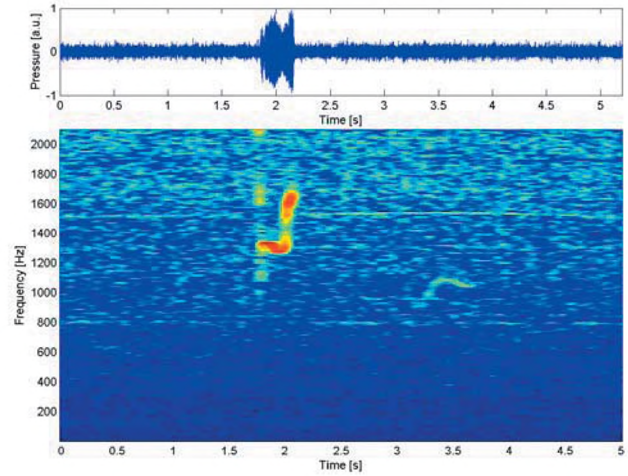

Figure 10: Time-series and "high-resolution" tf-plot of the detected dolphin sweep.

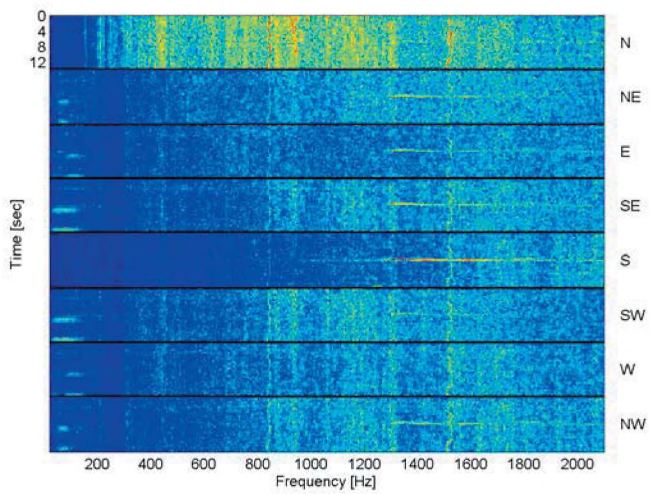

Figure 8: Multi-beam LOFAR display. In each of the eight formed beams 12 seconds of data are depicted with frequency on the horizontal axis and time on the vertical axis. The beam directions (N, NE, etc.) are listed on the right side. A dolphin sweep is visible in the Southern direction, but has also leaked into other directions via the beamformer's sidelobes.

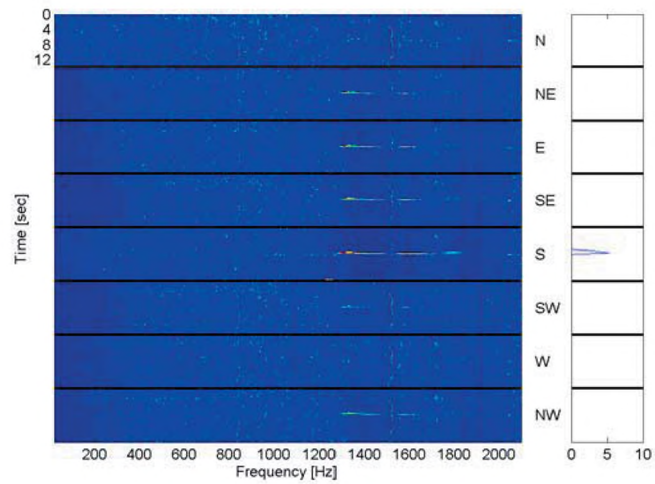

Figure 9: The CAPTAS marine mammal detection display is a multi-beam LOFAR of the normalised data (left) and corresponding Page's test output (right). The blue line on the right side depicts the Page's test output and marks a short signal detection in the Southern (aft) direction.

The right side of Figure 9 shows the output of the power-law/Page's test detector. This detector performs a summation over all frequency bins for each time step. Whenever this summation exceeds the detection onset threshold (h_0), a signal is detected. The thresholds are set using trial and error so that the detector is not sensitive for small noise bursts but still detects the low amplitude transients. In this case, the dolphin sweeps in the southern direction are detected.

After the detection of a transient, the start and stop times of this transient are known and the transient can be stored. This isolation of the signal is very helpful for further analysis (classification and localisation).

Figure 10 depicts a high-resolution *tf*-plot of the dolphin sweep detected in ***Figure*** 9. This *tf*-plot was made using the stored beamformed data. Compared to the *tf*-plot before beamforming, in Figure 7, it is readily apparent that the signal to noise ratio has been significantly increased. This increase in SNR makes it easier to classify the transient and to detect at longer ranges.

2.3.2 Detection of low frequency clicks

For the second example, we have used recordings made during another LFAS trial conducted in the Mediterranean Sea in the autumn of 2003. During this trial a prototype of the described detector was used for monitoring of the underwater environment and the detection of transients. During one of the passive experiments, several low frequency transients were detected. An example of recorded, single hydrophone, data is shown in Figure 11. In this figure, a time-series and tf-plot for a 45-second snapshot are shown. The low frequency transients are present in this data but not clearly visible. Around the 10, 17 and 35 second time stamps some very weak and low frequency transients are visible. They are masked by both tow ship noise and acoustic transmissions from an active low frequency source, which was used for performance evaluation tasks during the passive experiment. These acoustic transmissions consisted of broadband noise in the 1000 to 2000 Hz frequency band and three tonals at 1000, 1100 and 1200 Hz. However, both noise sources are only present in the forward sector.

After beamforming the signal to noise ratio of the transients has improved dramatically as can be seen in Figure 12. This figure depicts the beamformer output for a snapshot around the 17 second time mark in Figure 11. The spatial filtering of the beamforming seems highly effective. Several clicks are now visible, especially around the Eastern and Western direction. For these low frequency transients Port-Starboard discrimination cannot be obtained with an array designed for 1-2 kHz. However, the transients in the Eastern direction are slightly stronger than the transients in the Western direction. Other dominating features are the three transmitted tonals, which leak into all directions due to their relative high transmission levels.

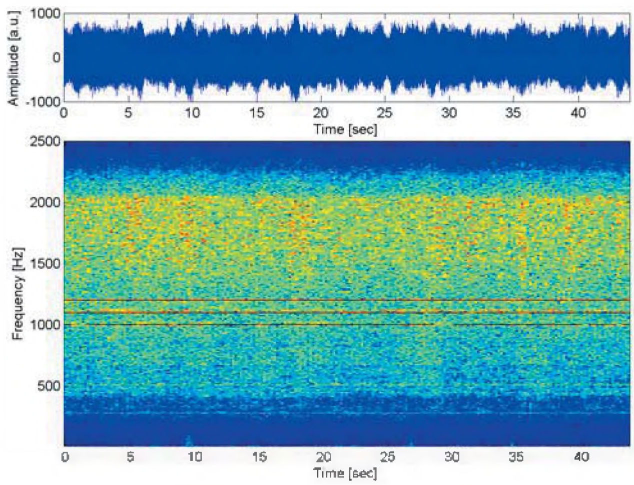

Figure 11: Time-series and tf-plot of a raw hydrophone signal with several very low frequency and barely visible transients.

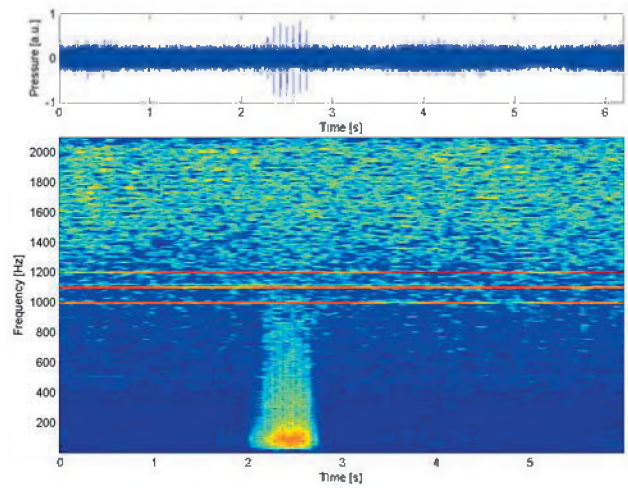

Figure 14: Time-series and high resolution tf-plot of the detected low frequency clicks.

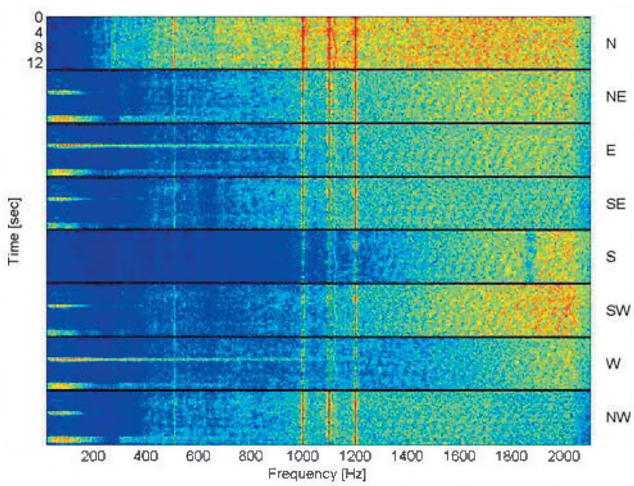

Figure 12: Multi-beam LOFAR plots for the eight formed beams. Low frequency clicks are visible in all directions but clearest in the Eastern (and Western) direction.

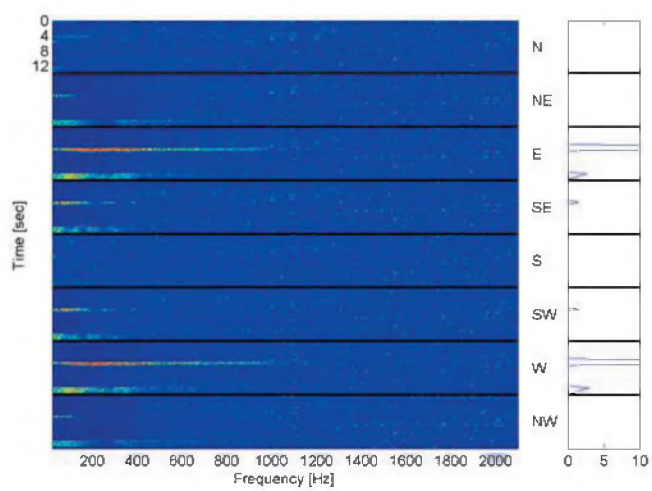

Figure 13: CAPTAS marine mammal detection display. The Page's test output marks several short signal detections especially in the Eastern and Western direction.

The corresponding normalised beamformed data and the power-law/Page's test output are shown in ***Figure*** 13. After normalisation the clicks are even more visible than previously in Figure 12. The power-law/Page's test detector was triggered by the strongest clicks. Lower settings for the different thresholds can make it possible to detect the lower level transients. However, this will go at the expense of an increased false alarm rate. The system then becomes more sensitive to tow speed changes and other noise bursts. A high resolution *tf*-plot and time-series of a detected click is shown in Figure 14. Compared to Figure 11, where the clicks were barely visible, the signal to noise ratio has increased. This makes it easier to classify the transient. Nevertheless, this is still a daunting task for acousticians. Our best guess for now is that it is a large whale (maybe a fin whale or sperm whale).

3. Classification of marine mammal transients

Once a transient is detected, it is important to know whether it is man-made or biological. In the latter case, it is interesting to classify it in more detail. Mitigation measures for large baleen whales are less severe than for small toothed whales, like harbour porpoises.

Two types of classification methods are popular for transients:

- Statistical analysis of time series (higher-order spectra),
- Pattern recognition in *tf*-plot.

We propose the latter method, as the former was unsuccessful in earlier studies. In the following we will work out an example (of harbour porpoise clicks) to demonstrate our prototype classifier, which was trained on several marine mammal recordings that were downloaded from the Internet.

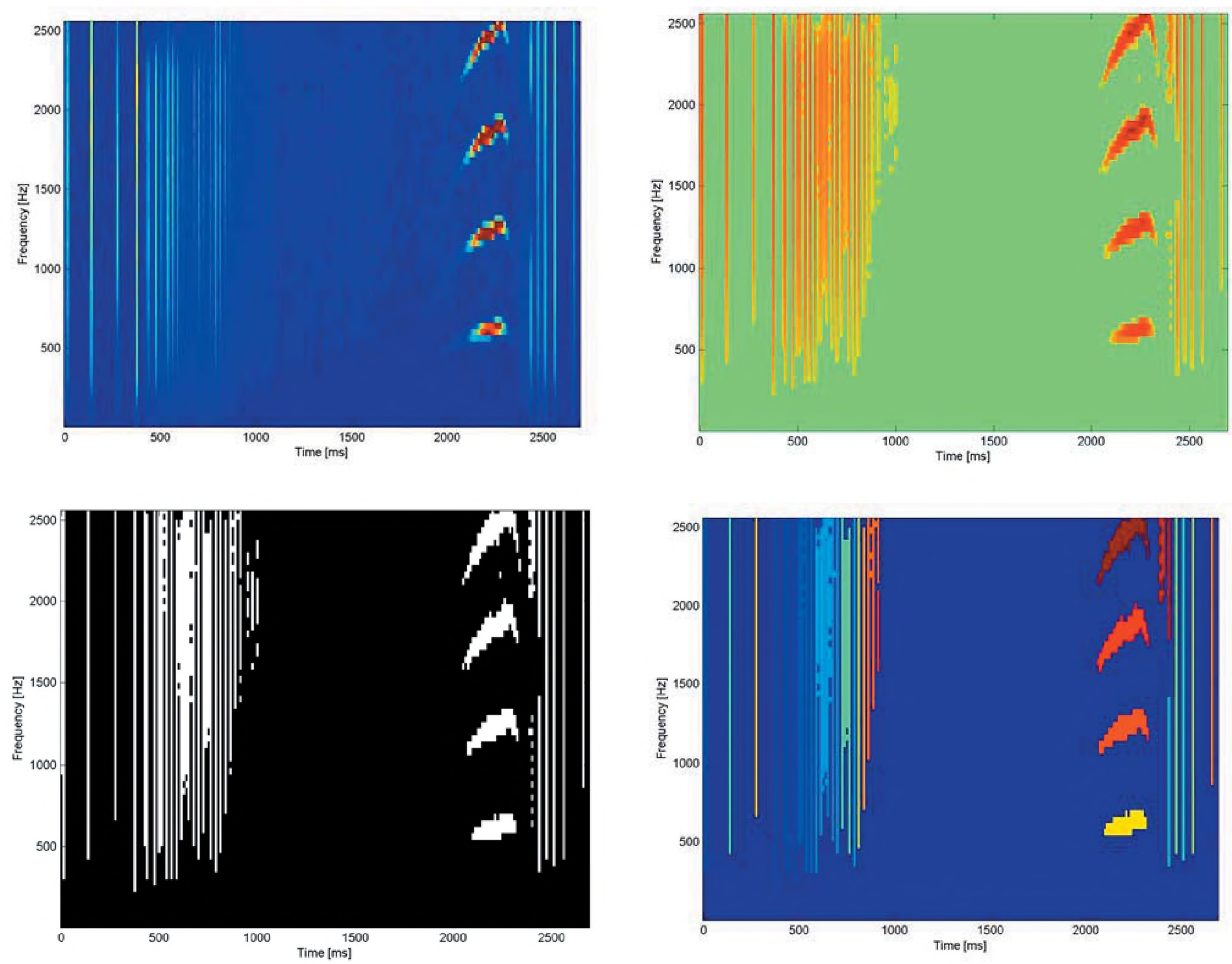

Figure 15: Processing scheme of the tf-plots to enable pattern recognition techniques.
On top left: a raw tf-plot of a porpoise clicks with varying pulse repetition frequency; on top right: a normalised tf-plot; below left: threshold crossings in the normalised tf-plot; below right: clustering of threshold crossings colour-coded by lowest frequency.

3.1 Time-frequency-plots (grams)

Different techniques are often used to compute the time-frequency distribution. The most common used techniques are summarised below:

Short-time FFT processing (STFT) is commonly used to make *tf*-plots. The time-series of the transient is cut into short segments, which are analysed spectrally by means of an FFT. Overlapping (50%) is often used. Sequential spectra are plotted in the gram. The FFT length is an important parameter. It determines the inevitable trade-off between time resolution and frequency resolution. For classification of clicks FFT lengths of 128 (corresponding to 0.025 s) seems suited, but for baleen whale calls longer integration (by a factor 4) is better.

Apart from the standard STFT technique several other methods are worth mentioning. *Cochlea processing* is a technique based on the human–ear. The cochlea in the human inner-ear acts as a logarithmic frequency filter. The technique is very suitable for the identification of human speech and seems suitable for application to other biological sounds as on marine mammal transients. *Wavelet processing* is often mentioned as being optimal for the analysis of transient signals. However, applications of this technique in sonar systems are still pending. This also holds for *Wigner-Ville processing* and other more exotic processing methods.

These innovative techniques all have their own speciality, but none of them proved to be robust for the wide variety of marine mammal sounds (ranging from long, low frequency calls to very short wideband clicks). Therefore we opted to use the simple and robust STFT processing in our prototype classifier until something better comes up. We realise that examples can be found where other processing methods perform better, *e.g.* for low frequency signals of inter animal communication from large baleen whales, the cochleagram more clearly separates the harmonic structures and appears to be the preferred time-frequency distribution. However, the major concern for tactical LFAS systems is for small and medium size odontoces (which are the hunting type of mammals), like harbour porpoises and Cuvier's beaked whales. These animals often produce clicks, which are wideband signals that STFT processing can reasonably deal with. Therefore, it was decided to proceed in this study with ordinary STFT, in which both clicks and calls can be classified.

3.2 Normalisation, thresholding and clustering

Before we can apply pattern recognition techniques, the structures in the *tf*-plot have to be isolated. To achieve this we propose the following processing scheme. See also ***Figure*** 15 for illustrations, where some porpoise clicks are depicted.

3.2.1 Normalisation

This is an important step. The background energy in the *tf*-plot is often distributed rather than uniformly homogeneous. Therefore "whitening" should be applied before structures can be isolated through thresholding. Both temporal and spectral effects cause inhomogeneity: Background noise levels are higher at lower frequencies. Sometimes pre-whitening is already applied in the recording system, but not always. The whitening can be theoretically compensated (-17 log f for sea noise spectrum according to Knudsen [18]). Adaptive methods that measure the actual spectral background (for instance the method used in Page's test) are used in this study. Apart from spectral variation there is also a temporal variation of the background, which is compensated for by automated level control.

In the upper right panel in ***Figure*** 15 a normalised *tf*-plot is shown. Compared to the raw *tf*-plot (upper left) the clicks are more clearly separated, mainly due to the temporal normalisation, which compensates for the higher background between 500 and 1000 ms.

3.2.2 Thresholding

After spectral and temporal normalisation the median is subtracted from the data (such that the noise level is 0) and the data is divided by the maximum in the *tf*-plot (such that the maximum signal level is 1). After this, a threshold can be set. Depending on the data quality its value is on the order of 0.05. (The analysed recordings in our training set were downloaded from Internet and differ in recording quality). In the lower left panel in Figure 15 a *tf*-plot after thresholding is shown. Threshold crossings are groups of vertical lines (clicks) in the first 1000 ms and after 2500 ms. Furthermore four 'islands' are visible around 2150 ms. This harmonic structure is caused by rapidly repeating clicks, for which the repetition time is (much) shorter than the integration time. (In reference [19] an elaborate study on harbour porpoise click trains is presented.)

3.2.3 Clustering

The threshold crossings are clearly grouped. In the clustering procedure all connected points are recognised as a single cluster. This procedure is a standard Matlab® function in the image processing toolbox. We removed all small clusters; signals have either duration or bandwidth, so small clusters are often just noise.

In the lower right panel in ***Figure*** 15 the remaining clusters (51 strongest from a total of 121) in the *tf*-plot are shown. The clusters are numbered starting with the lowest frequency. This means that colour (from blue to red) indicates the lowest frequency in the cluster.

3.3 Pattern recognition

Now that we made clusters, we are left with patterns that need to be recognised in order to classify the signal. The strength of this classification method is that although marine mammal sounds vary a lot in frequency, duration and level, they do not have a lot of different typical patterns. Basically only four typical sounds are produced:

- Clicks
- Moans
- Whistles
- Sweeps

All four of these have easy recognisable (LOFAR) patterns. Clicks are vertical lines. Moans are blobs and always have a harmonic structure. Whistles are thin lines mainly horizontal, and have (weak) harmonics. Sweeps are thin lines with more vertical structure (bandwidth) and sometimes lack harmonics.

For all four typical sounds "recognisers" are developed. These recognisers are built up in similar way. First clusters are reshaped in an automated way by standard image processing techniques. This reshaping is necessary as for instance clicks (vertical lines) are often broken down in several fragments, which can be reconnected by filling techniques. On the other hand moans (islands) tend to be connected by narrow bridges and have to be separated.

Next from the reshaped clusters *features* are determined. These features are elementary properties of the clusters like: length, height, centre of mass, standard deviation, energy content, etc.

Finally these features (or combinations) are compared to standards that are representative for the patterns of the four standard sounds. But before we start, a large false alarm reduction is achieved by recognising air-gun transmission, which is the main source for false detections in the ocean.

3.3.1 Airgun removal

No less than 75% of all detected transients are air-gun transmissions [20]. Air-guns are numerous and both powerful and with low-frequency content such that propagation is favourable for them. Air-guns are easily recognisable in a *tf*-plot, see Figure 16 for an example. They are short and band-limited transients, which are manifested as triangles on the floor of the *tf*-plot. All transients that are classified as air-guns are automatically removed.

3.3.2 Reshaping clusters (Erode-Dilate)

Our strategy is to determine features from the clusters, and compare these to standards for the four classes of signals above. Before we start to determine the features, the clusters are reshaped by means of "erode-dilate" techniques, see [21]. The number of erode-dilate steps is different for each of the four recognisers.

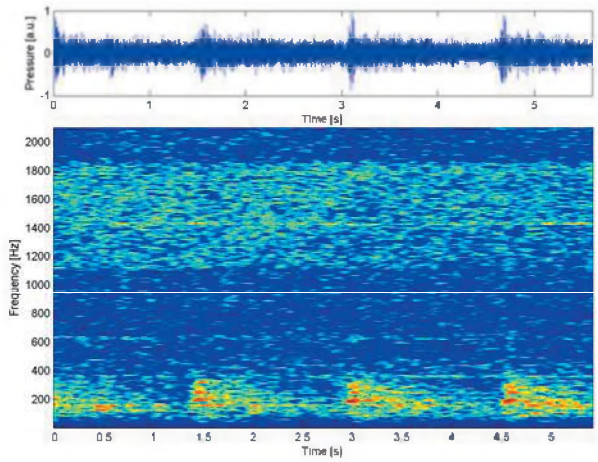

Figure 16: Tf-plot of air-gun transmissions.

For *clicks* the sequence starts with dilate steps in vertical direction followed by erode steps. This will fill the gaps between segments of a broken line. Furthermore we erode in horizontal direction to remove reverberation that tends to attach consecutive clicks. For *whistles* a similar procedure is followed, but horizontal and vertical are interchanged. For *moans* it is important to get rid of artificial vertical connections (due to imperfect FFT filtering) to separate the islands. Here erode steps in both direction are useful. *Sweeps* that have a 2-dimensional structure are best left alone. The number of erode-dilate steps is an important tuning parameter. It depends on the quality of the data and on the pre-processing. In general it can be remarked that it is better to use a low threshold and apply many erode steps, than to use high thresholds and dilate steps.

3.3.3 Classification

In order to classify a detected sound the measured features of a cluster are compared to those of the standard sounds. Below an abbreviated procedure is given:

1. Clicks; we demand the cluster to have small aspect and considerable height. Apart from single cluster features, we also check whether the cluster is repetitive, *i.e.* we check if a group of lines is present.
2. Moans; we demand the cluster to be compact (medium aspect) and repetitive in frequency.
3. Whistles; we demand the aspect to be large and the filling to be poor. There is a check for harmonics, which concludes whether the whistle is biological or man-made.
4. Sweeps; as for whistles, but with a special demand for the third moments for obtain skewness.

When a structure is not recognised as an air-gun, or any of our list the transient is unclassified.

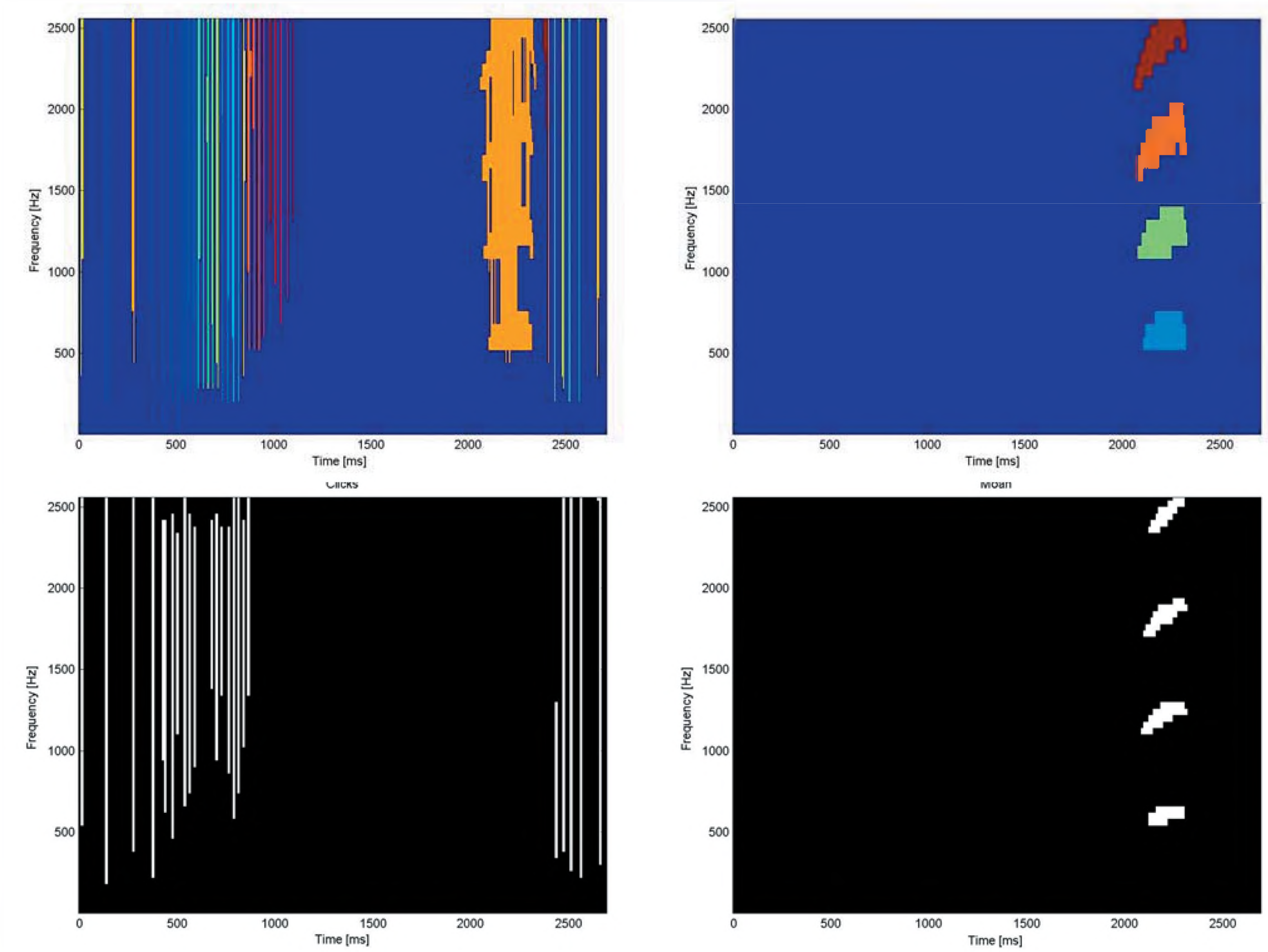

Figure 17: Example of erode-dilate processing on clicks (left) and harmonics (right). Input is the final plot of Figure 15 with clusters. For clicks erode-dilate aims at filling the gaps between the fragments. Here DDDEEE in vertical direction is applied. For the harmonics the moan detector is used. Here the bridges between the islands should be broken. Here ED in vertical and in horizontal is applied. The results are shown in the upper panels. Below pattern recognition is applied, which recognises the clicks on the left and the harmonics on the right.

4 CONCLUSIONS

The proposed transient or marine mammal detector can be separated into two basic steps. The first step is pre-processing and beamforming. This step is used to improve the signal to noise ratio and to obtain direction information on the detection. For this purpose a new type of beamformer is developed, with a constant beamwidth in the full frequency band.

The second step is automatic detection of the wide variety of marine mammal vocalisations. This is achieved by the combined use of a power-law and Page's test algorithms. The power-law integrator is robust against varying signal bandwidth while the Page's test detector is a robust detector for signals with an unknown duration.

This combination of sector beamforming and power-law/Page's test detector seems to be very promising in detecting marine mammal vocalisations; see also [17]. In an application to an LFAS array it proved possible to detect high frequency dolphin sweeps as well as low frequency clicks from a large whale during sea trials.

The classifier is still under development. Algorithms are implemented, but tests of the classifier on recorded transients are pending. The amount of useful CAPTAS data are still limited. Some marine mammal transients from the Internet were gathered and the score on those was fair (especially for clicks and moans), but not exceptional. However, the quality of these recordings differs a lot (in noise level, filters, etc.). The algorithms are sensitive to the exact settings of the detector and therefore tuning of classification parameters for arbitrary WAV-files is cumbersome. If sufficient CAPTAS data is available a well-trained classifier could be developed as the proposed algorithms seem quite robust. A final step would be the inclusion of a localiser. The animal's range is an essential parameter in mitigation measures. Ideas for this are being developed [22].

REFERENCES

[1] *R.W Richardson et al*, Marine mammal and noise, *Academic Press, San Diego, 1995.*

[2] *National Research Council (2000),* Marine mammals and low frequency sound (progress since 1994), *National Academic Press, Washington DC.*

[3] *K.M. Fristrup, L.T. Hatch and C.W. Clark,* Variation in humpback whale (Megaptera novaeangliae) song length in relation to low-frequency sound broadcasts, *JASA 113(6), pp341-3424, June 2003.*

[4] *D.M.F. Chapman and D.D. Ellis,* The elusive decibel: thoughts on sonars and marine mammals, *Canadian Acoustics 26(2), pp29-31, 1998.*

[5] *C. Erbe and D.M. Farmer,* A software model to estimate zones of impact on marine mammals around anthropogenic noise, *JASA 108(3), pp1327-1331, Sept 2000.*

[6] *R.D. McCauley, J Fewtrell & A.N. Popper, (2003*), High intensity anthropogenic sound damages fish ears, *JASA 113 (1).*

[7] *F.P.A. Benders, S.P. Beerens & W. Verboom,* SAKAMATA: the ideas and algorithms behind it, *Proceedings of the ECUA 2004 conference, July 2004.*

[8] *G.H. Hotho,* Adaptation of a submarine transient detector to the detection of whale sounds, *Proeedings of UDT Europe 1999.*

[9] *D.A. Abraham,* Passive acoustic detection of marine mammals, *Saclantcen Memorandum SM-351, April 2000.*

[10] *D.A. Abraham, R.J. Stahl,* Rapid detection of signals with unknown frequency content using Page's test, *Proceedings of 1996 Conference on information Sciences and Systems, March 1996, pp. 809-814.*

[11] *A.H. Nuttall,* Detection performance of power-law processors for random signals of unknown location, structure, extent, and strength, *Tech. Rep. 10751, Naval Underwater Systems Center, September 1994.*

[12] *E.S. Page,* Continuous inspection schemes, *Biometrika, vol. 41, pp 100-114, 1954.*

[13] *G.W.M. van Mierlo, S.P. Beerens, R. Been, Y. Doisy and E. Trouve,* Port-Starboard discrimination on hydrophone triplets in active and passive towed arrays, *Proceedings of UDT Europe 1997, pp. 176-181, 1997.*

[14] *S.P. Beerens, R. Been, J Groen, Y. Doisy and E. Noutary*, Adaptive Port-Starboard beamforming for triplet arrays, *Proceedings of UDT Pacific, 63-68, 2000.*

[15] *R.O. Nielsen*, Sonar signal processing, *Artech house 1991.*

[16] *L.J. Karam., and J H. McClellan,* Complex Chebyshev Approximation for FIR Filter Design,. *IEEE Trans. on Circuits and Systems II, March 1995, pp. 207-216.*

[17] *R. Laterveer, S.P. Beerens, S.P. van IJsselmuide, M. Snyder & J Newcomb*, A high frequency transient detector suitable for monitoring marine mammal activity, *Proceedings of UDT Europe, 2004.*

[18] *R.J Urick,* Principles of underwater sound, *Peninsula publishing, 0-932146-62-7, 1983.*

[19] *W.C. Verboom, R.A. Kastelein,* Structure of harbour porpoise (Phocoena phocoena) click train signals, *De Spil Publishers, ISBN 90-72743-07-5, 1997.*

[20] *NOAA:* http://www.noaa.gov/whales/

[21] *Robert M. Haralick and Linda G. Shapiro,* Computer and Robot Vision, vol. I, *Addison-Wesley, 1992, pp. 158-205.*

[22] *S.P. Beerens, S.P. van IJsselmuide & M.K. Robert,* Localisation of Marine Mammals using an LFAS array , *Proceedings of the ECUA 2004 conference, July 2004.*

APPENDIX A: Flow chart of described power-law/Page's test detector.

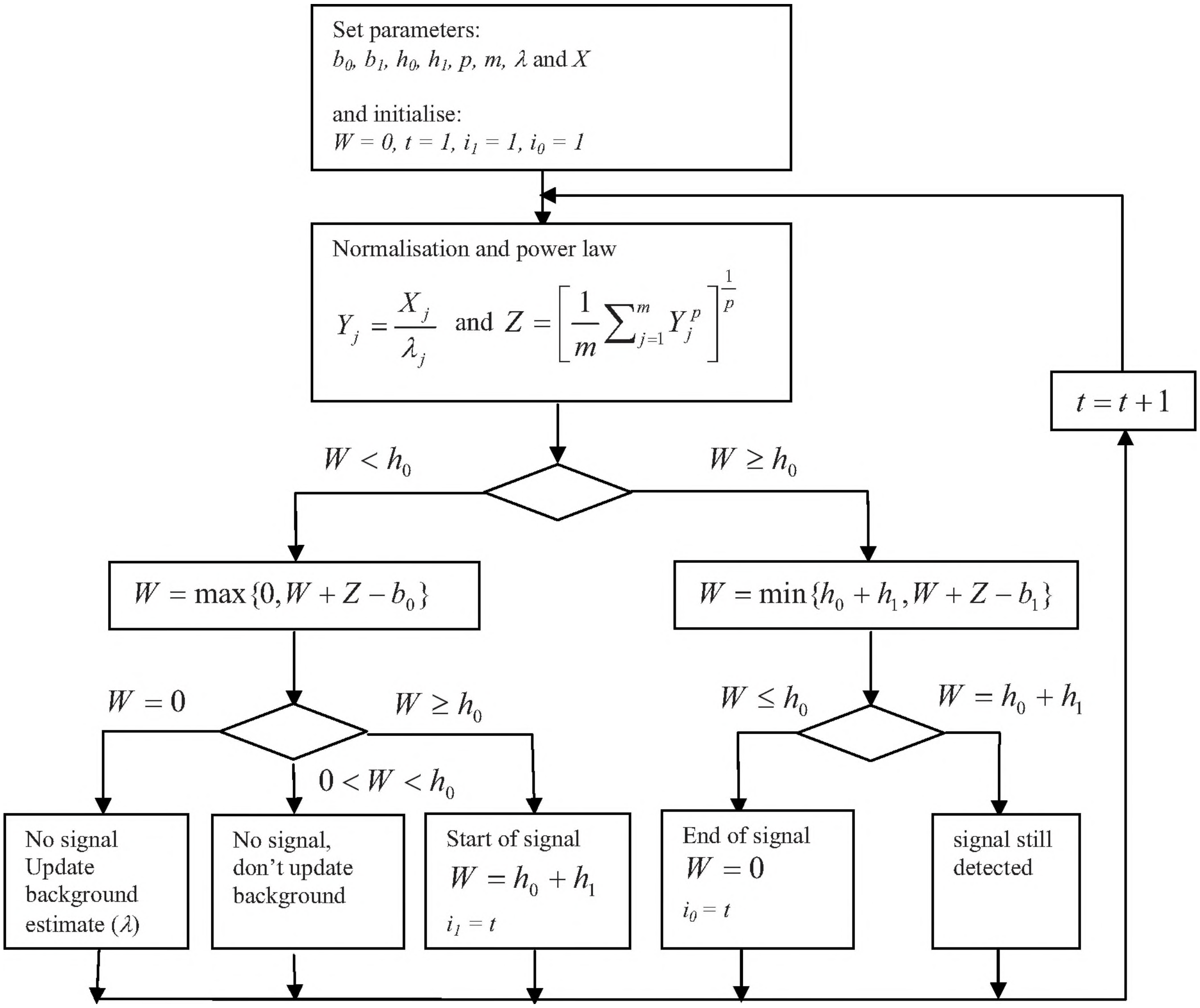

Figure A1: Block diagram of the Page's test detector.

Description of used variables

p - power law ($p \geq 1$)
h_0 - threshold for start of signal detection
b_0 - Page's test bias for start of signal detection
h_1 - threshold for end of signal detection
b_1 - Page's test bias for end of signal detection
α - time constant for exponential averaging of power spectra ($0 < \alpha < 1$)
Y - normalised power spectrum
Z - power-law output
W - Page's test statistic (is 0 if no signal and h_0+h_1 if signal detected)
i_1 - start index of signal detected
i_0 - stop index of signal detected
j - frequency bin index number