DETECTION OF LEOPARD SEAL (*HYDRURGA LEPTONYX*) VOCALIZATIONS USING THE ENVELOPE-SPECTROGRAM TECHNIQUE (TEST) IN COMBINATION WITH A HIDDEN MARKOV MODEL

Holger Klinck, Lars Kindermann and Olaf Boebel
Alfred Wegener Institute for Polar and Marine Research, Am Alten Hafen 26, 27568 Bremerhaven, Germany
Homepage: www.awi.de/acoustics
Holger.Klinck@awi.de

ABSTRACT

This paper describes a technique for the automated detection of leopard seal (*Hydrurga leptonyx*) vocalizations. Automatic detection of leopard seal calls within the Antarctic underwater soundscape is difficult because (a) the calls are frequently of low amplitude (b) the call duration is highly variable and (c) the frequency band overlaps with those of many other marine mammal vocalizations. However, humans easily distinguish leopard seal vocalizations from those of other marine mammals because of the calls' distinctive sound, which is a result of the pulsed structure of the leopard seal vocalizations. To exploit the unique temporal evolution of the pulse repetition rate (PRR) in high (HDT) and low (LDT) double trills, the Envelope-Spectrogram Technique (tEST) was developed. The extracted PRR feature allows detection of the target vocalizations even against a background of other marine mammal vocalizations. To handle the high variability of the calls' duration, the tEST algorithm was combined with a Hidden Markov Model (HMM) which is particularly well adapted to handle temporal variability. The developed HMM based detection algorithm worked rather reliably. The detection rate over a 4 day test period was high (72 %) although the signal to noise ratio (SNR) was low (< 10 dB). The number of false positive detections (12 %) was tolerable. Most of the false positives occurred during the period when R/V Polarstern was approaching the recording station and the SNR was temporarily < 0 dB. The detector worked 3 times faster than real-time and is therefore suitable for both off line biological research and time critical in-the-field applications, such as the detection of the presence of leopard seals in the context of human diver operations.

SOMMAIRE

Cet article décrit une technique de détection automatique des vocalisations du Léopard de Mer (*Hydrurga leptonyx*). La détection des sons émis par le Léopard de Mer à travers le bruit de fond sous marin est difficile parce que (a) les émissions sont fréquemment de basse amplitude (b) la durée des émissions est hautement variable et (c) les vocalisations sont dans la même bande que celle utilisée par de nombreux autres mammifères marins. Cependant, l´homme est facilement en mesure d´identifier les vocalisations émises par le Léopard de Mer de celles des autres mammifères, grâce à la pulsation particulière de ces émissions. Pour exploiter cette caractéristique unique de l'évolution temporelle du taux de répétition des pulsations (PRR) des doubles trilles hauts (HDT) et graves (LDT), la technique du spectrogramme de l'enveloppe (tEST) a été développée. Les caractéristiques PRR du signal permettent la détection des vocalisations recherchées même en présence de celles d´autres mammifères marins. Pour résoudre les problèmes dus à la haute variabilité des durées d´émission, l'algorithme tEST a été combiné avec le modèle des chaines de Markov (HMM), particulièrement bien adapté pour appréhender les variations temporelles. Cet algorithme de détection basé sur les HMM s'est révélé relativement performant. Le taux de détection sur une période d´essai de quatre jours a été élevé (72 %) malgré un faible rapport signal sur bruit (SNR) (< 10 dB). Le nombre de détections positives erronées (12 %) était tolérable. La plupart des détections erronées se sont produites lorsque le navire de recherche R/V Polarstern s´est approché de la station d´enregistrement, diminuant ainsi le SNR (< 0 dB). Le détecteur travaillant trois fois plus vite que le temps réel, il est de fait utilisable aussi bien pour les analyses de données post récolte, que pour une utilisation directe sur le terrain, comme par exemple la détection de la présence de Léopards de Mer lors d´opérations de plongée sous-marine.

1. INTRODUCTION

The leopard seal (*Hydrurga leptonyx*) represents one of three Antarctic pack ice seal species. Leopard seals are solitary living animals, feeding on krill, squid, fish, penguins and other seal species (Reeves et al., 2002). As a top predator of circumpolar distribution, the leopard seal plays an important role in the Antarctic ecosystem. The population size is estimated at around 200000 animals (Reeves et al., 2002). Research on this species is restricted due to the Antarctic pack ice region being accessible for humans usually only during the short austral summer period.

Figure 1: Leopard seal (*Hydrurga leptonyx*) on floating sea ice.

Underwater, leopard seals are known to be rather vocal - at least during polar summer when most of the research was conducted. Stirling and Siniff (1979) described high vocalization rates of male leopard seals during the breeding season (November to January). Females show above-average vocalization rates during sexual receptivity (Rogers et al., 1996). Hence, passive acoustic monitoring offers the unique possibility to investigate the species without a need of direct access. PALAOA - the Perennial Acoustic Observatory in the Antarctic Ocean, (Boebel et al., 2006) is an autonomous recording station operated by the Alfred Wegener Institute (AWI), Germany, on the Ekström Ice Shelf close to the German Neumayer Base, providing underwater recordings from the Atlantic sector of the Southern Ocean. Since January 2006, PALAOA records the Antarctic underwater soundscape quasi-continuously. The station's audio system allows broadband data acquisition with sampling rates of up to 192 kHz and 24 bit resolution. So far more than 6400 hours of acoustic data (as at September 2007) were accumulated. The recorded sounds are transmitted in real-time to the AWI in Germany, allowing real-time access and analysis of the acoustic data.

Extracting the signals of interest - in this case the leopard seal vocalizations - from the resulting 2.5 TBytes of data, is challenging. Obviously, human "observers" will not be able to manage this task, but rather, numerical detection algorithms need to be developed to perform an automated, computer based search. The resulting time series of calls will then form the data base for ecological studies with focus on diel patterns, diurnal and seasonal variability and their interrelation with the changing physical environment.

Apart from these scientific applications, the development of detection algorithms for leopard seal vocalizations can also help to increase the safeness of research divers in the Southern Ocean. Several encounters between human divers and leopard seals have been reported throughout the last decades (Muir et al., 2006). The most serious incident occurred in July 2003 at the British Rothera Station, located at the Antarctic Peninsula, when a scientist was killed by a leopard seal. As a consequence of this accident, acoustic monitoring prior to and during diving activities are used by AWI as risk mitigation method for diving activities. To this end, robust and fast (at least real-time) detection algorithms are needed to screen the hydro-acoustic recordings.

2. THE ACOUSTIC ENVIRONMENT

The Southern Ocean is among the regions least disturbed by anthropogenic noise. However, PALAOA records reveal a high degree of abiotic and biotic acoustic activity in the Southern Ocean. During austral summer in particular, the Antarctic underwater soundscape is dominated by the vocalizations of Weddell seals (*Leptonychotes weddellii*), Ross seals (*Ommatophoca rossii*), crabeater seals (*Lobodon carcinophaga*), leopard seals (*Hydrurga leptonyx*) and various baleen (Mysticeti) and toothed (Odontoceti) whale species.

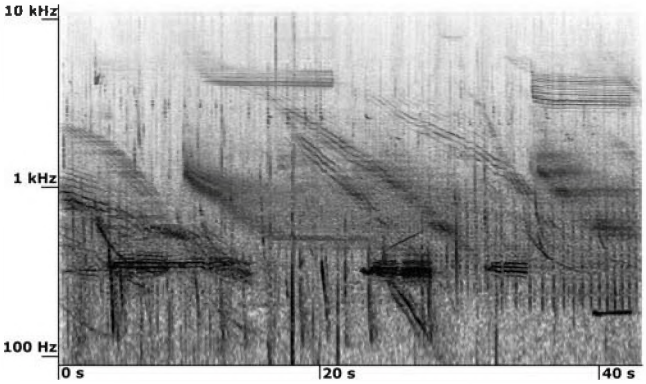

Figure 2: Spectrogram of a PALAOA sound file.

Figure 2 shows a spectrogram of a typical sound file as recorded at the PALAOA Station in austral summer. Overlapping vocalizations from different animals/species significantly complicate the detection of specific

vocalization "targets." However, human listeners can easily distinguish leopard seal vocalizations from those of other marine mammals because of their distinctive sound. It is believed that this distinctive sound is a result of the pulsed structure of the leopard seal vocalizations. To develop a detection algorithm for leopard seal calls, this publication exploits in detail the temporal structure of the pulse repetition rate (PRR) throughout the calls. The PRR feature is exclusively linked to leopard seal vocalizations (at least in the vicinity of the PALAOA Station) and seems to be a robust feature for the detection while other marine mammal vocalizations are present.

3. LEOPARD SEAL VOCALIZATIONS

3.1 State of knowledge

A first (partial) spectrogram of a leopard seal vocalization was published by Ray (1970) while Stirling and Siniff (1979) described four different leopard seal call types quantitatively. A comprehensive description of the vocal repertoire of leopard seals was published by Rogers et al. (1995). Rogers identified twelve different call types by analyzing recordings of captive and free living animals (Prydz Bay, Antarctica). The frequency span of the analyzed call types ranges between 65 Hz and 4800 Hz. Thomas et al. (1983) recorded ultrasonic vocalizations with frequencies up to 164 kHz of leopard seals in captivity during hunting activity. However, ultrasonic vocalizations have so far not been reported from field studies.

By far the most frequent vocalizations of leopard seals are the so called high double trill (HDT – see Figure 3) and the low double trill (LDT). In the PALAOA recordings, the HDT (frequency range: 2.5 - 4.5 kHz) and the LDT call type (frequency range: 230 - 470 Hz) represent more than 70 % of all leopard seal vocalizations while Rogers et al. (1995) reported 79 % of such calls for their data set.

Due to its distinct Signal to Noise ratio (SNR) this paper focuses on the analysis and detection of HDTs.

3.2 The high double trill (HDT)

Figure 3 (top) depicts the waveform and spectrogram of a high double trill. The waveform clearly shows that the call is separated into two segments which consist of a series of short pulses. These pulses cause an amplitude modulation of the main signal. This modulation generates so called sidebands, which are revealed in the spectrogram in Figure 3 (bottom). The frequency difference between the sidebands equals the frequency of the PRR. This implies that an increasing (decreasing) PRR is causing increasing (decreasing) frequency differences between the sidebands. In general the number of sidebands is determined by the type of amplitude modulation. For sinusoidal modulation, only two sidebands are generated while the primary frequency is rendered invisible in the spectrogram. By contrast, triangular or rectangular modulations cause multiple (> 2) sidebands. The type of amplitude modulation of the HDT is in-between a sinusoidal and triangular modulation (see Figure 5).

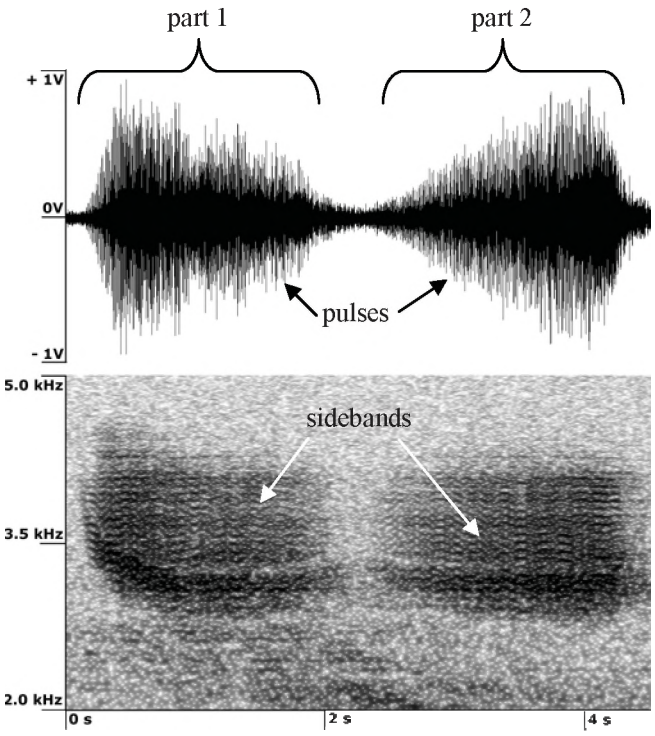

Figure 3: Waveform and spectrogram of a HDT call.

A total of 150 HDTs were analyzed to gain information about the temporal structure and the frequency characteristics of this call type (Table 1). The HDTs feature a high variability of the call duration, ranging from 1.9 s - 9.0 s. The calls cover a frequency range between 2500 Hz and 4450 Hz.

Table 1: Acoustic features of the HDTs recorded at PALAOA.

	Min	Max	Mean	Stdv
Call duration	1.9 s	9.0 s	4.5 s	1.6 s
Frequency	2500 Hz	4450 Hz	---	---

4. THE ENVELOPE-SPECTROGRAM TECHNIQUE (tEST)

So far, HDT and LDT descriptions regarded the PRR as constant for the duration of the call (Rogers, 2007; Rogers et al., 1995). By contrast, spectrograms of calls recorded by PALAOA reveal varying side-band distances over the duration of HDT and LDT calls, suggesting a variation of the PRR in the course of the call.

To accurately analyse the temporal structure of the PRR, a Matlab based algorithm was developed. The respective sound snippet was first band pass filtered with the frequency range of the target signal (HDT: 2.5 - 3.5 kHz). The envelope of the absolute values of a band passed waveform was calculated by detecting all maxima values (peaks) in the waveform and interpolating (1-D) the detected points. The resulting waveform was then down-sampled to a sampling rate of 1000 Hz and transformed into the frequency domain by means of a Fast Fourier Transformation (FFT-Parameters: Hamming window 256 points; 50 % overlap).

This algorithm, named tEST (the Envelope-Spectrogram Technique) hereinafter, provides the spectrogram of the envelope, i.e. the frequency-evolution of the PRR. If the SNR over the frequency range of the target signal is low (< 6 dB), it can be helpful to use a narrower filter which covers only the frequency range of the signal of the highest energy.

4.1 Applying tEST on HDT calls

Figure 4 shows the result of tEST applied on a HDT. The signal was processed as described in the former paragraph.

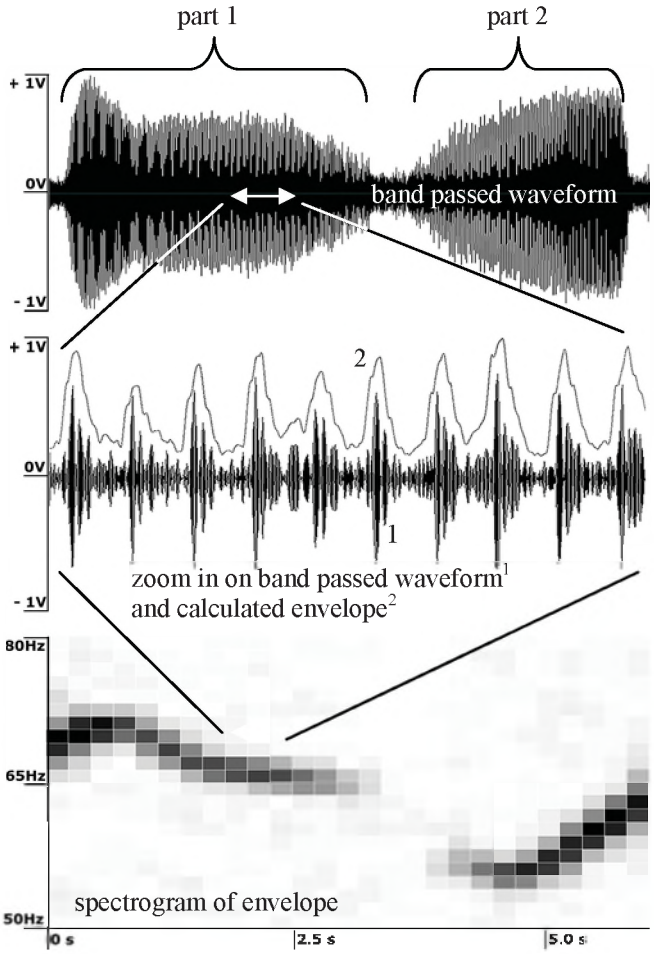

Figure 4: Results of tEST applied on a HDT.

The result of the FFT of the envelope signal is displayed in the lower part of Figure 4. For the selected sample, the pulse repetition rate of the envelope varies between 52 Hz and 72 Hz (20 Hz bandwidth @ 2 Hz resolution). The first part of the call is characterized by descending rates. In the second part the pulse repetition rate is ascending.

The same 150 HDTs as used for the spectral description were analysed with tEST. All vocalizations showed descending repetition rates in the first part of the call and ascending rates in the second part. The observed frequencies ranged between 45 Hz and 75 Hz.

5. DETECTION OF LEOPARD SEAL VOCALIZATIONS

5.1 Introduction

In summary, the previous sections showed:

(a) The analysis of the acoustic environment (Section 2) confirmed that vocalization of various whale and seal species occur simultaneously within the frequency bands of the target vocalizations. Thus the likelihood of false positive detections will be high using detection methods such as energy summation or comparing energies in different frequency bands.

(b) The call durations of the HDT calls vary widely (Section 3) which renders detection algorithms based on matched filter/spectrogram correlation difficult. Further more the detection performance of these methods is linked to the representativeness of available examples of the target vocalization. 150 calls are probably not enough samples to create an effective filter.

(c) Leopard seal calls exhibit temporal modulation of the PRR throughout HDTs providing a unique feature of this leopard seal call type (Section 4). Other marine mammal vocalizations in frequency bands overlapping with those of the target vocalizations are likely not to pass as "false positives" if the detection algorithm is to exploit this rather unique feature.

For the detection of the HDTs based on the PRR feature a Hidden Markov Model (HMM) was applied. The next paragraphs will give a short introduction to HMMs and how they are used for the detection.

5.2 Hidden Markov Models (HMM)

Hidden Markov Models (HMM) are statistical models for the detection and classification of transient patterns, representing state of the art tools in human speech

recognition (Rabiner, 1993). HMMs are particularly well adapted to this call type of variable duration, as they allow detection of temporally changing structures. For leopard seal vocalizations this implies that the detection probability is high irrespectively of the calls duration, as long as the envelope follows the specific temporal evolution (see Figure 6). A short introduction and description on how to build a Hidden Markov Model are given in the following paragraphs. Unfortunately, the scope of this paper only allows an abbreviated, qualitative description of Hidden Markov Models. For detailed information see Rabiner, 1993 and Deller et al. 2000. All model parameters used in this study are available on request.

(a) Feature extraction: To extract the call type's typical features, it is recommended to choose the best available samples. Consequently the best 100 samples (high SNR) were selected out of the available 150 samples for this process. The features are extracted by means of a time-frame based analysis (frame length: 256 ms) of the envelope signal of the 100 sample files of variable duration (2 - 9 s, depending on call duration). For each time frame, a feature vector is calculated (see Figure 5), representing the respective energy distribution as a function of frequency.

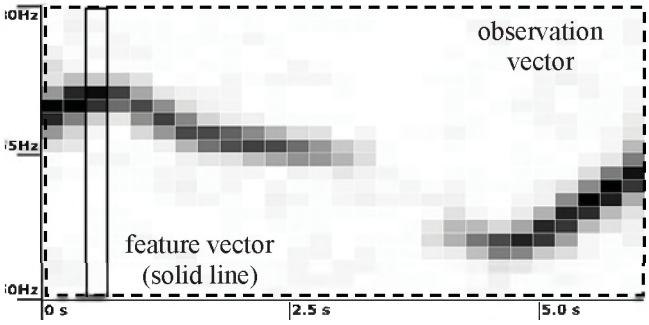

Figure 5: Feature vector (enclosed by solid box) and observation vector (enclosed by dashed box) of a HDT.

All feature vectors (n = duration of sample file / 256 ms) of one sample file comprise the so called observation vector (which actually is a matrix - see Figure 7). Hence, 100 sets of spectral vectors from the spectrograms of the envelopes are extracted.

(b) Information reduction: To condense the numerous ensuing feature vectors to a set of 'most significant' feature vectors, a "k-mean (squared Euclidean distance)" cluster algorithm (Deller et al., 2000) was applied to the training set. This creates the so called codebook of 10 (number empirically chosen) codebook vectors representing the target vocalization's most significant (sub-)set of 10 feature vectors. For each feature vector of an observation vector the best matching (minimal distant) codebook vector is determined, which results in an observation sequence. Each of these consists of a series of integers, representing the sequence of IDs of the best fitting codebook vectors. Thus, each set of the 100 spectral vectors is quantized to a one dimensional array of integers. The resulting set of 100 quantized vectors represents the quantized training set.

(c) Generate the HMM: Evaluating model parameters describing the quantized training set best. A Hidden Markov Model is a quintuple, comprising (a) the number of (hidden) states S; (b) the state transition matrix A (transition probabilities between the states); (c) the observation probability matrix B; (d) the state probability vector at time t=1, $\pi(1)$; and (e) the number of observable outputs Y (number of codebook vectors), or in short:

HMM={S, $\pi(1)$, A, B, Y}.

The number of states (S) was assigned to 5. The number of 10 observable outputs (Y) is given by the size of the codebook. The state transition matrix (A), the observation probability matrix (B) and the state probability vector at time=1 ($\pi(1)$) were determined by applying the forward/backward algorithm (Deller et al., 2000) on the quantized data set.

In the first step, the model parameters A, B and $\pi(1)$ are initialized (see below) and the algorithm calculates the match between the model and the training set. In the second step the algorithm starts to modify the model parameters. The algorithm guarantees that every iteration has a matching likelihood that is >= to the previous one. Once the matching likelihood converges, training is done.

Critical to this process is the initial guess of the model parameters. In the case of the described target signals, meaningful parameters were unknown. If the initial guess is too far away from the optimal parameters, then the algorithm will only find a local maximum but not the global one. For this reason the initial parameters were randomized and the resulting HMMs used to analyse one sample file including a known number of target signals repeatedly. The best fitting model parameters (giving the highest detection probability for the target signals) were then chosen for the further process.

5.3 Detection of HDTs using the HMM

To detect HDTs with the optimized HMM (5 states), a 6-s window is continuously slid in steps of 1.0 s (~ 83 % overlap) over the data stream: the respective window content is first band pass filtered (2500 Hz - 3500 Hz). The resulting waveform is then used to calculate the envelope, which is used to derive the observation sequence as described in section 5.2. In a final step the probability of the observation sequence of each window under the assumption of the model $P_{(window|model)}$ is calculated. In Figure 6 an

example for the output of the HMM detection algorithm is given. The spectrogram shows four signals - two HDTs (c) of different duration, an artificial signal (b) and a Weddell seal call (a). The detector output is shown in the lower part of Figure 6. The detection threshold is set automatically by the algorithm depending on the overall SNR. If the detector output reaches the detection threshold, a call is "detected". In this sample the HDT calls are clearly detected by the system but not the Weddell seal call or the artificial signal present in the same frequency band.

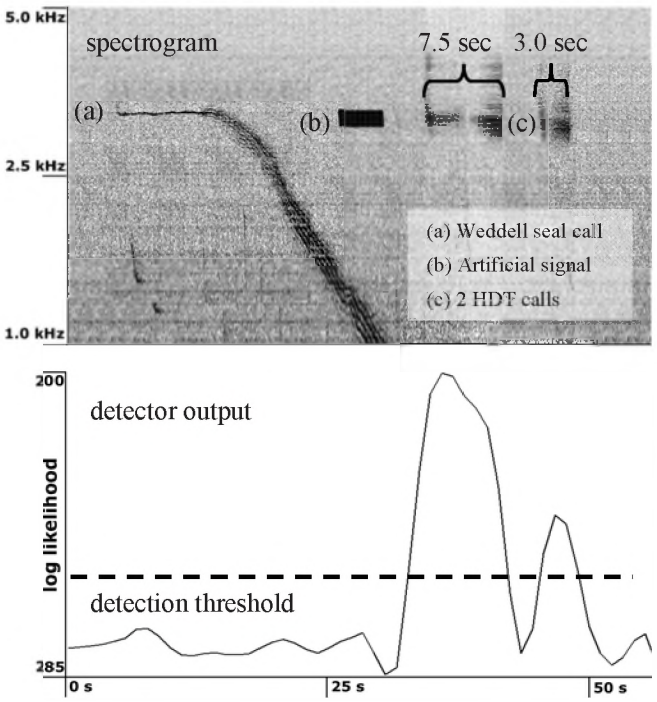

Figure 6: Detector output for a Weddell seal call, an artificial signal and two HDT calls of different duration.

5.4 Results of the detection system running over a test data set

To test the detection algorithm, independent data (i.e. not including the 100 calls used to develop the HMM) from a 4 day period of variable SNR was selected and analysed. The period started out with a good bandwidth related SNR of 10 dB (between 2.5 kHz and 3.5 kHz), which deteriorated to SNR < 0 dB during the last day of the period when R/V Polarstern approached the recording station. To deal with the low SNR the spectrogram of the envelope was manipulated using a wavelet based denoising technique (Kovesi, 2000 and Kovesi, 1999). Also an anisotropic diffusion was performed on the spectrogram to enhance the contrast at sharp intensity gradients (Kovesi, 2000). The use of the denoising and the anisotropic diffusion algorithm increased significantly the detection performance. Thus, the algorithms were directly integrated into the system.

A subset of data was selected (2 min of data every 10 min) to create a reference data set, which was used to evaluate the detection algorithm. The result of the test run is presented in Figure 7. The light line represents the manually detected calls; the dark line represents the automatically detected calls. The total number of manually detected calls in the selected files was 1527 and the detection rate of the system 72 %. The overall temporal evolution of the two curves shows a high degree of similarity. Analysing all files over the period the HMM based algorithm detected 7548 HDTs.

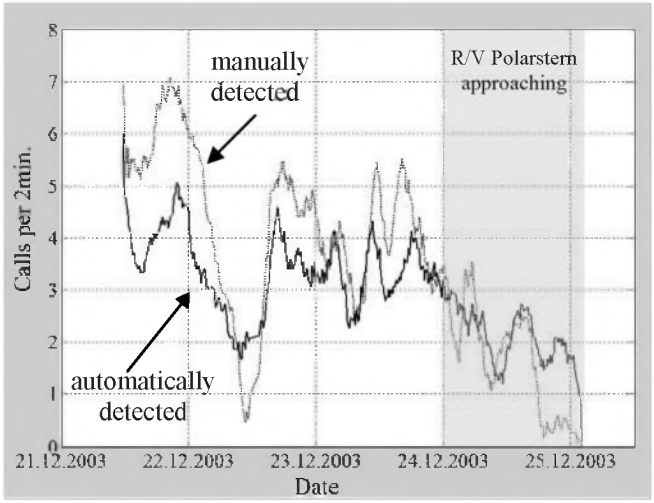

Figure 7: Results of a test run over a four day period.

Most of the false positives (in total 12 %) occurred during the period when R/V Polarstern was approaching - especially when the vessel was close to the recording station (second half of day 24.12.2003). A detailed error analysis will be included in upcoming studies, to determine the exact cause of these false detections.

The detection algorithm is rather fast. Analysing a sound file of 2 minutes duration (48 kHz - 16 bit) takes about 40 seconds (Desktop PC with single Intel Pentium IV 3.4 kHz processor and 2GB RAM). Thus the HMM based detection system is suitable for real-time applications.

6. DISCUSSION AND OUTLOOK

The tEST algorithm which was developed in the course of this study is a useful tool for analysing the temporal evolution of pulse repetition rates in animal calls. The analysis of the high double trill (HDT) of the leopard seal revealed for the first time a temporal variation in the repetition rate of the pulses.

First results of the analysis on low double trills (LDT) of the leopard seal showed also a temporal variation in the pulse repetition rate throughout this call type. The frequency range of the PRR is compared to the HDT around 4 times

lower (bandwidth between 13 Hz and 20 Hz). Future work will concentrate on the detection of LDTs using the PRR feature in combination with a Hidden Markov Model.

However, the specific PRR modulation is exclusively linked to leopard seal vocalizations (at least in the vicinity of PALAOA Station), facilitating the development of a HMM based detection system to detect the target vocalizations in a data set which was entirely overlaid by vocalizations of other marine mammals.

A problem of detection algorithms is often their validation - especially when working with huge data sets. HMM based detection systems provide the "matching probability" between the signal and the used model for each call detected. Analysing this probability over time can help to identify regions where the "matching probability" is low and a validation is necessary in particular.

In summary, it is noted that the Hidden Markov Model worked rather reliably. The detection rate over the 4 day test period was high (72 %) although the SNR was unfavourable (< 10 dB). The number of false positive detections (12 %) was tolerable, because most of the false positives occurred during the period when R/V Polarstern was approaching the recording station when the SNR was temporarily < 0 dB. The detector worked 3 times faster than real-time and is therefore suitable for time critical applications.

ACKNOWLEDGEMENTS

Many thanks to Cornelia Kreiß for pooling the data manually and creating the necessary reference data set for deriving the efficiency of the detection system. The authors especially want to thank Marie A. Roch for her helpful remarks and suggestions. Internal and external reviewers provided useful comments on previous drafts of this manuscript. Delphine Dissard, Catherine Audebert and an anonymous reviewer provided the French translation of the abstract. Setting up the PALAOA observatory would not have been possible without the extensive support of the AWI logistic department.

REFERENCES

Boebel, O., Kindermann, L., Klinck, H., Bornemann, H., Plötz, J., Steinhage, D., Riedel, S. and Burkhardt, E. (2006): Acoustic Observatory Provides Real-Time Underwater Sounds from the Antarctic Ocean. In: EOS, 87, pp. 361 and 366.

Deller, J. R., Hansen, F. H. L. and Proakis, J. G. (2000): Discrete-Time Processing of Speech signals. In: Wiley-IEEE Press, Chapter 12, pp. 677-744.

Kovesi, M. (2000): MATLAB and Octave Functions for Computer Vision and Image Processing. School of Computer Science & Software Engineering, University of Western Australia. Available from: <http://www.csse.uwa.edu.au/~pk/research/matlabfns/>.

Kovesi, M. (1999): Phase Preserving Denoising of Images. In: The Australian Pattern Recognition Society Conference (DICTA '99), December 1999, Perth, pp. 212-217.

Muir, S. F., Barnes, D, K.A. and Reid, K. (2006): Interactions between humans and leopard seals. In: Antarctic Science, 18(1), pp. 61-74.

Rabiner, L. and Juang, B. H. (1993): Fundamentals of Speech Recognition, Prentice Hall PTR, New York, 496 pp.

Ray, G. C. (1970): Population ecology of Antarctic seals, Volume 1. In: Antarctic Ecology (Ed. Holdgate, M. W.), Academic Press, New York, pp. 398-414.

Reeves, R. R., Stewart, B. S., Clapham, P. J. and Powell, J. A. (2002): National Audubon Society: Guide to Marine Mammals of the World. Alfred A. Knopf, New York, 531 pp.

Rogers, T. L. (2007): Age-related differences in the acoustic characteristics of male leopard seals, *Hydrurga leptonyx*. In: Journal of the Acoustic Society of America, 122(1), pp. 596-605.

Rogers, T. L., Cato, D. H. and Bryden, M. M. (1996): Behavioural significance of underwater vocalizations of captive leopard seals, *Hydrurga leptonyx*. In: Marine Mammal Science, 12(3), pp. 414-427.

Rogers, T. L., Cato, D. H. and Bryden, M. M. (1995): Underwater vocal repertoire of leopard seals (*Hydrurga leptonyx*) in Prydz Bay, Antarctica. In: Sensory Systems of Aquatic Mammals (Ed. Kastelein, R. A., Thomas, J. A. and Nachtigall, K.), De Spil Publishers, Woerden, pp. 223-236.

Stirling, I. and Siniff, D. B. (1979): Underwater vocalizations of leopard seals (*Hydrurga leptonyx*) and crabeater seals (*Lobodon carcinophagus*) near South Shetland Islands, Antarctica. In: Canadian Journal of Zoology, 57, pp. 1244-1248.

Thomas, J. A., Fisher, S. R. and Evans, W. E. (1983): Ultrasonic vocalizations of leopard seals (*Hydrurga leptonyx*). In: Antarctic Journal of the US, 17, page 186.