SONAR DISCRIMINATION OF CYLINDERS FROM DIFFERENT ANGLES USING NEURAL NETWORKS

Lars Nonboe Andersen, Whitlow Au (*), Jan Larsen and Lars Kai Hansen

CONNECT, Department of Mathematical Modeling, Building 321
Technical University of Denmark, DK-2800 Lyngby, Denmark
email's: lna,jl,lkhansen@imm.dtu.dk
www: http://eivind.imm.dtu.dk

(*) Marine Mammal Research Program, Hawaii Institute of Marine Biology
University of Hawaii, P.O. 1106, Kailua, Hawaii, 96734, USA

ABSTRACT

This paper describes an underwater object discrimination system applied to recognize cylinders of various compositions from different angles. The system is based on a new combination of simulated dolphin clicks, simulated auditory filters and artificial neural networks. The model demonstrates its potential on real data collected from four different cylinders in an environment where the angles were controlled in order to evaluate the models capabilities to recognize cylinders independent of angles.

1. INTRODUCTION

Dolphins possess an excellent sonar system for solving underwater target discrimination and recognition tasks in shallow water (see e.g., [2]). This has inspired research in new sonar systems based on biological knowledge, i.e. modeling the dolphins discrimination capabilities (see e.g., [4] and [5]). The fact that the inner ear of the dolphin has many similarities with the human inner ear makes it tempting to use knowledge from simulations of the human auditory system when trying to model the dolphin sonar system. Neural networks have proven to be very useful for pattern recognition tasks which makes it interesting for the present application.

Based on earlier work [3], [11], [15] we will describe and present results from an experiment using a simulated dolphin signal, preprocessing based on auditory modeling and classification using a neural network to classify echoes from different cylinders and different angles.

2. DATA COLLECTION

Echoes from four cylinders made of different material compositions were measured from different angles with an echo collection system using a planar broadband calibration transducer to project and receive the acoustic signals. All data were collected, in a cylindrical tank 2.4 m in diameter, 1.8 m high and filled with sea water located at Marine Mammal Research Program, University of Hawaii.

All four cylinders were 3.8 cm in outer diameter and 10 cm high. Cylinder nr. 1 was made of degassed epoxy containing coral rock pebbles. Cylinder nr. 2 was made of aluminum and hollow with an inner diameter of 2.54 cm and open in both ends making it possible for water to fill the inside. Cylinder nr. 3 was made of aluminum and hollow with an inner diameter of 2.54 cm but closed in both ends creating a cylindrical air space inside. Cylinder nr. 4 was made of stainless steel and hollow with an inner diameter of 2.54 cm and open in both ends making it possible for water to fill the inside. The transducer was mounted on a pole and placed in the tank with the backside of the transducer approximately 12 cm from the wall and with equal distances to the bottom and to the water surface. The transducer was pointing towards the center of the tank. Each of the cylinders were mounted on a rotor with two monofilament lines attached to the two ends of the cylinder and suspended in the tank approximately 1.65 m in front of the transducer at the same depth as the center of the transducer. The cylinder was positioned in such a manner that its longitudinal axis was horizontally aligned. For each of the cylinders the rotor was initially acoustically calibrated to 0 degrees when the longitudinal axis of the cylinder was perpendicular to the normal of the transducer surface. This was accomplished by first assuming the cylinder will produce the strongest echo when placed in the 0 degrees position and then turning the cylinder to the position in which the visually strongest echo was observed on an oscilloscope. A simulated dolphin echolocation signal was projected at the cylinders and a Gage 1012 Data Acquisition Board operating at a sampling frequency of 1 MHz was used to digitize the echoes with a resolution of 16 bits. Each echo consisted of 1024 points with the time window placed so major reflections from the water surface and tank walls and bottom were left out. For each cylinder echoes were collected from angles from 0 degrees (the longitudinal axis of the cylinder perpendicular to the normal of the transducer surface) to 90 degrees (the longitudinal axis of the cylinder parallel to the normal of the transducer surface) at increments of 1 degree. For each angle 10 echoes were collected at a rate of 2 echoes per second) resulting in total: $10 * 91 * 4 = 3640$ echoes.

This research was supported by the Danish Natural Science and Technical Research Councils through the Computational Neural Network Center (CONNECT). JL furthermore acknowledge the Radio Parts Foundation for financial support.

3. PREPROCESSING

Features from the echoes were extracted using a combination of a matched filter, envelope detection, a gammatone filterbank, time integration and PCA.
To find the beginning of each echo a matched filter, which was implemented as the time reversed version of the transmitted simulated dolphin click, was applied. The start of each echo was chosen as the peak of the envelope of the output from the matched filter (see e.g. [6]). The first 512 data points following the start was then used as the input signal to the gammatone filterbank.

The gammatone filter is defined by its impulse response [12]

$$g(t) = at^{(n-1)} \exp(-2\pi bt) \cos(2\pi f_c t + \varphi), \ t > 0 \quad (1)$$

where b largely determines the duration of the impulse response and thus, the bandwidth of the filter; n is the order of the filter and it largely determines the slope of the skirts. When the order of the filter is in the range 3-5, the shape of the magnitude characteristic of the gammatone filter is very similar to that of the roex(p) filter commonly used to represent the magnitude characteristic of the human auditory filter [9]. The equivalent rectangular bandwidth (ERB) of the auditory filter is given by [14]

$$ERB = 24.75(4.37 f_c/1000 + 1Hz) \quad (2)$$

where f_c is center frequency of the filter. The center frequencies of the filters has been determined using Fay's modification of Greenwood's equation [13] for estimating cochlea frequency distributions along the basilar membrane

$$f_c(x) = 0.008 f_{\max}(10^{2.1x} - 1.0) \quad (3)$$

where $f_{\max}$ is the maximal frequency perceived by the animal and x is the position of the filter on the basilar membrane expressed as the proportion between the distance from the basal end and the full length of the basilar membrane ($x = 0$ at the basal end and $x = 1$ at the apical end). A high-frequency boundary of 150 kHz for $f_{\max}$ was used to coincide with the bottlenose dolphin upper frequency limit of hearing [8].
If we choose $n = 4$ and f_c/b is large, which is the case here, then b and the 3-dB bandwidth, BW, of the filter are given by [10]

$$b = 1.019 ERB \quad (4)$$

$$BW = 0.887 ERB \quad (5)$$

The Q-value of a filter is defined as $Q = f_c/BW$ and by using the described function for center frequencies all filters have approximately a constant Q-value of 10 which is between the Q-values measured for the bottlenose dolphin using the two different techniques: Critical Bandwidth (CB) ($Q = 2.2$) [1] and Critical Ratio (CR) ($Q = 12.3$) [7].
The remaining constants a and φ in the gammatone filter are chosen as $a = 1$ as the amplification variable and $\varphi = 0$ as the phase variable.

The filterbank consists of N_{filt} such gammatone filters and N_{filt} is limited to 15 for computational reasons. The output of the filterbank when filtering the input signal with the bank of filters consists of N_{filt} new signals, $y(n_{\text{filt}}, t)$, where n_{filt} is the filter number. Each of these signals was split up in N_{bin} time bins of a chosen length of $\delta_{\text{bin}} = 25\ \mu$s and the energy was calculated in each time bin, n_{bin} using

$$E(n_{\text{filt}}, n_{\text{bin}}) = \int_{(n_{\text{bin}}-1)*\delta_{\text{bin}}}^{n_{\text{bin}}*\delta_{\text{bin}}} y^2(n_{\text{filt}}, t) dt \quad (6)$$

This *energy image* was then log-transformed using $E_{\log} = \log_{10}(E)$.

$E_{\log}$ consists of $N_{\text{filt}} \cdot N_{\text{bin}}$ values which is reduced by first finding the principal components of a training set and then use the N_{eig} largest eigenvalues to project all the $E_{\log}$'s to a lower dimensional space (see e.g., [17]). N_{eig} was chosen to 10 for computational reasons, which means 10 features were extracted from each echo.

4. CLASSIFICATION

The extracted features were classified using a feed-forward net with a modified SoftMax [18] normalization as presented in [15] (see also, [17], [16]). The 2-layer feed-forward network with n_I inputs, n_H hidden neurons and $c-1$ outputs, where c is the number of classes, is defined by:

$$h_j(\boldsymbol{x}) = \tanh\left(\sum_{\ell=1}^{n_I} w_{j\ell}^I x_\ell + w_{j0}^I\right), \quad (7)$$

$$\phi_i(\boldsymbol{x}) = \sum_{j=1}^{n_H} w_{ij}^H h_j(\boldsymbol{x}) + w_{i0}^H \quad (8)$$

where $w_{j\ell}^I$, w_{ij}^H are the input-to-hidden and hidden-to-output weights, respectively. All weights are assembled in the weight vector $\boldsymbol{w} = \{w_{j\ell}^I, w_{ij}^H\}$. In order to interpret the network outputs as probabilities we used a *modified* normalized exponential transformation [15] similar to SoftMax [18],

$$\widehat{z}_i = \frac{\exp(\phi_i)}{\sum_{j=1}^{c-1} \exp(\phi_j) + 1}, \quad 1 \le i \le c-1, \quad (9)$$

and $\widehat{z}_c = 1 - \sum_{i=1}^{c-1} \widehat{z}_i$.

The network was optimized using the maximum a posteriori technique, i.e., the cost function is the sum of the log-likelihood and a regularization term (prior).

$$C(\boldsymbol{w}) = S_{\mathcal{T}}(\boldsymbol{w}) + R(\boldsymbol{w}, \boldsymbol{\kappa}) \quad (10)$$

where $R(\boldsymbol{w}, \boldsymbol{\kappa})$ is a weight decay parameterized by a set of regularization parameters $\boldsymbol{\kappa}$.
The full scheme for optimizing the network was presented at ICASSP 98 [15] but in this work only the weights and the regularization parameters were optimized using a second-order Gauss-Newton scheme based

on the training set for the weights and a gradient descend scheme based on the validation set for the regularization parameters as described in [15]. For a more detailed description and use of outlier detection, see [15].

The number of inputs in the network correspond to the number of extracted features ($n_i = 10$), the number of hidden units, n_H was chosen to 5 and the number of classes c was chosen as the number of cylinders ($c = 4$). Five hidden units was chosen to give an acceptable size of the network which then consists of $11 \cdot 5 + 6 \cdot 3 = 73$ free weights.

5. EXPERIMENT

To investigate the capability of the system to discriminate between the cylinders independent of angle the echoes were divided in a training set and a test set with the training set consisting of the echoes from some angles and the test set consisting of the echoes from the remaining angles. Experiments were made for training angles separated by 2, 5, 10 and 15 degrees, resulting in training sets of 1840, 760, 400 and 280 data respectively (0 and 90 degrees was chosen to be included in the training set) and test sets of 1800, 2880, 3240 and 3360 data respectively. It should be mentioned for optimizing the neural network the training set was again split into two equally sized sets denoted training set and validation set in [15] used for optimizing different parameters in the network but as described in [15] weights in the network were finally retrained on the combined set of training and validation.

Table 1 reports the performance of the system for the different training separation angles. The system is capable of discriminating between the different cylinders independent of angle with a probability of misclassification *pmc* of 0.059 if the spacing between the training angles is 2 degrees. It is noted that closer spacing and thereby more training angles as expected decreases the probability of misclassification.

	2 Deg.	**5 Deg.**	**10 Deg.**
Train.	0.016	0.000	0.005
Val.	0.064	0.164	0.260
Test	0.059	0.186	0.186
Test after retrain	0.059	0.138	0.195

Table 1: Probability of misclassification for different separation of training angles

6. DISCUSSION AND CONCLUSION

This paper presented a new system for underwater object discrimination particularly focusing on eliminating the influence of the objecs angles. An experiment involving cylinders made of various material compositions and recordings from different angles demonstrated the potential of the system provided enough angles is used when training the system.

In nature the dolphin is possibly also capable of extracting angle independent information since the dolphin can possibly not obtain information from all angles before it recognizes the object. It would thus be interesting to setup an experiment with a dolphin in which the angles available for the dolphin were strictly controlled as in the experiment described in this paper. Such an experiment is of course not easy to carry out due to the water motions in a penn with a dolphin and the fact that it is difficult to control the position of the dolphin to an accuracy of only a few degrees.

7. REFERENCES

[1] W.W.L. Au & P.W.B. Moore: "Critical Ratio and Critical Bandwidth for the Atlantic Bottlenose Dolphin," *J. Acoust. Soc. Am. 88*, pp. 1635–1638, 1990.

[2] W.L.L. Au: *The Sonar of Dolphins*, New Jersey, USA: Springer Verlag, 1993.

[3] W.L.L. Au, L.N. Andersen, A.R. Rasmussen, H.L. Roitblat & P.E. Nachtigall: "Neural Network Modeling of a Dolphin's Sonar Discrimination Capabilities," *J. Acoust. Soc. Am. 98*, pp. 43–50, 1995.

[4] H.L. Roitblat, P.W.B. Moore, P.E. Nachtigal, R.H. Penner & W.W.L. Au: "Natural Echolocation with an Artificial Neural Network," *Neural Networks 1,* pp. 239–248, 1989.

[5] P.W.B. Moore, H.L. Roitblat, R.H. Penner & P.E. Nachtigall: "Recognizing Successive Dolphin Echoes with an Integrator Gateway Network," *Neural Networks 4,* pp. 701–709, 1991.

[6] W.S. Burdic: *Underwater Acoustic System Analysis*, New Jersey, USA: Prentice Hall, 1991.

[7] C.S. Johnson: "Masked Tonal Threshold in the Bottlenose Porpoise," *J. Acoust. Soc. Am. 44*, pp. 965–967, 1968.

[8] C.S. Johnson: "Auditory Masking on One Pure Tone by Another in the Bottlenose Porpoise," *J. Acoust. Soc. Am. 49*, pp. 1318–1318, 1971.

[9] R.D. Patterson & B.C.J. Moore: "Auditory Filters and Excitation Patterns as Representations of frequency resolution," In: B.C.J. Moore (eds.),*Frequency Selectivity in Hearing*, London, Academic Press Limited, pp. 123–177, 1986.

[10] R.D. Patterson, J. Holdsworth, I. Nimmo-Smith & P. Rice: "SVOS Final Report: The Auditory Filterbank," *APU report 2341*, 1988.

[11] R.D. Patterson, K. Robinson, J. Holdsworth, D. McKeown, C. Zhang & M. Allerhand: "Complex Sounds and Auditory Images," In *Advances in the Biosciences Vol. 83*, UK, Pergamon Press, pp. 429–446, 1992.

[12] A.M.J.H. Aertsen & P.I.M. Johannesma: "Spectro-temporal Receptive Fields of Auditory Neurons in the Grassfrog. I. Characterisation of Tonal and Natural Stimuli," *Biol. Cybern., 38,*, pp. 115–135, 1980.

[13] R.R. Fay: "Structure and Function in Sound Discrimination among Vertebrates," In: D.B. Webster, R.R. Fay & A.N. Popper (eds.),*The Evolu-*

tionary Biology of Hearing, New York: Springer-Verlag, pp. 229–263, 1992.

[14] B.R. Glasberg, B.C.J. Moore: "Derivation of Auditory Filter Shapes from Notched-Noise data," *Hearing Research,* 47, pp. 103–138, 1990..

[15] J. Larsen, L.N. Andersen, M. Hintz-Madsen & L.K. Hansen : "Design of Robust Neural Network Classifiers," in J. Principe *et al.* (eds.) *Proc. ICASSP'98*, Seattle, USA, vol. 2, pp. 1205–1208, 1998

[16] L.N. Andersen, J. Larsen, L.K. Hansen & M Hintz-Madsen: "Adaptive Regularization of Neural Classifiers," in J. Principe *et al.* (eds.) *Proc. IEEE Workshop on Neural Networks for Signal Processing VII*, Piscataway, NJ: IEEE, pp. 24–33, 1997.

[17] C.M. Bishop: *Neural Networks for Pattern Recognition*, Oxford, UK: Oxford University Press, 1995.

[18] J.S. Bridle: "Probabilistic Interpretation of Feedforward Classification Network Outputs with Relationships to Statistical Pattern Recognition," *Neurocomp. - Algor. Arch. and Appl.*, Berlin, Germany: Springer-Verlag, vol. 6, pp. 227–236, 1990.