INSTITUTE OF PHYSICS PUBLISHING BIOINSPIRATION & BIOMIMETICS

Bioinsp. Biomim. **1** (2006) 41–51 doi:10.1088/1748-3182/1/2/002

Using a binaural biomimetic array to identify bottom objects ensonified by echolocating dolphins

David A Helweg[1,4], Patrick W Moore[1], Stephen W Martin[2] and Lois A Dankiewicz[3]

[1] Code 235, Space and Naval Warfare Systems Center, San Diego, 53560 Hull Street, San Diego, CA 92152, USA
[2] Code 237, SSC San Diego, San Diego, CA 92152, USA
[3] Science Applications International Corporation, 3990 Old Town Ave Suite 208, San Diego, CA 92110, USA

E-mail: david_helweg@usgs.gov and patrick.moore@navy.mil

Received 23 May 2006
Accepted for publication 10 July 2006
Published 17 August 2006
Online at stacks.iop.org/BB/1/41

Abstract

The development of a unique dolphin biomimetic sonar produced data that were used to study signal processing methods for object identification. Echoes from four metallic objects proud on the bottom, and a substrate-only condition, were generated by bottlenose dolphins trained to ensonify the targets in very shallow water. Using the two-element ('binaural') receive array, object echo spectra were collected and submitted for identification to four neural network architectures. Identification accuracy was evaluated over two receive array configurations, and five signal processing schemes. The four neural networks included backpropagation, learning vector quantization, genetic learning and probabilistic network architectures. The processing schemes included four methods that capitalized on the binaural data, plus a monaural benchmark process. All the schemes resulted in above-chance identification accuracy when applied to learning vector quantization and backpropagation. Beam-forming or concatenation of spectra from both receive elements outperformed the monaural benchmark, with higher sensitivity and lower bias. Ultimately, best object identification performance was achieved by the learning vector quantization network supplied with beam-formed data. The advantages of multi-element signal processing for object identification are clearly demonstrated in this development of a first-ever dolphin biomimetic sonar.

1. Introduction

The ability of dolphins to recognize objects using echolocation has been a subject of intense study for at least three decades [1–6]. The echolocation signals used by bottlenose dolphins (*Tursiops* spp.) are broadband impulsive clicks, with time–bandwidth products near unity [7] and peak frequencies and 3 dB bandwidths generally ranging from 20 to 100 kHz [8, 9]. The combination of bandwidth and adaptive control over the amplitude and spectral content of their biosonar signals [10] has allowed bottlenose dolphins to exploit niches ranging from estuaries and the surf zone out into pelagic waters. An understanding of the acoustic and motor behaviors used by echolocating dolphins may provide insight and guidance toward the development of more capable sonar systems.

In 2000, the SSC-SD Biosonar Program Office began construction of a dolphin biomimetic sonar (DBS) for application to littoral navigation and object detection. The U.S. Navy's mine-hunting dolphin systems have proven competence in shallow and very shallow water mine countermeasures [11]. In contrast, man-made sonar systems do not reliably detect or classify mines in the highly reverberant shallow water channel. One component of the DBS is

[4] Present address: USGS-BRD-PIERC, PO Box 44, Hawaii National Park, HI 96718, USA.

a biomimetic binaural sonar receiver—a paired array that matches the bandwidth and directivity that has been measured in *Tursiops* [12]. The construction of a binaural receiver made possible the exploration of binaural models of dolphin echolocation—an idea motivated by comparative research in humans and other species [13–15].

Study of dolphin echolocating includes development of biomimetic algorithms that attempt to instantiate one or more signal processes putative to the dolphin's biosonar signal processing. In 1988, Gorman and Sejnowski demonstrated that a backpropagation network (BPN) could discriminate a target object from a clutter object using sonar backscatter [16]. This study motivated application of artificial neural networks (ANNs) to biosonar models and proof-of-concept biomimetic classifiers. Roitblat and colleagues [17] demonstrated that BPN and counterpropagation networks could identify aspect-independent objects suspended in the water column when ensonified with a synthetic dolphin sonar click or an echolocating dolphin. Substantial improvement in identification (ID) performance was realized using multi-echo summation in the novel Integrator Gateway Network, designed to model the dolphin's decision making processes [18, 19]. The Integrator Gateway could identify aspect-independent targets [18, 19] and a set of aspect-dependent geometric solids ensonified in the water column by an echolocating dolphin [20]. Other networks could use biomimetic representations [21] or constant-binwidth and constant-Q input representations to identify objects in the water column [22] or buried in sediment [23].

Sonar systems may utilize one of several available ANNs to aid in the identification of objects. These ANN architectures include backpropagation, genetic reinforcement learning, learning vector quantization, and probabilistic networks (for details on neural network theory and architecture, see [24–26]). BPNs form linear combinations of input vectors and, like linear discriminant analysis, use parametric distributions of prior information about the correct target category to minimize RMS categorization error. Genetic reinforcement learning networks (GLN) use genetic algorithms (e.g. [27]) to modify network connection weights in the process of optimizing an objective function, in this case mine identification accuracy. Learning vector quantization (LVQ) and probabilistic (PNN) architectures are implementations of the Bayes decision rule, which uses Parzen windows to estimate the class-dependent nonparametric probability density functions for each mine category type (e.g., [28]). It is possible that one or more of these networks may be specifically suited for capitalizing on those particular echolocation behaviors of the dolphin that contribute to superior object identification performance.

Data were collected using the DBS receiver to study signal processing methods for object identification (ID). Acoustic backscatter was collected from four complex metallic objects (PDM2, MANTA, ROCKAN and MK82 DST) placed proud on a sand/mud surface, and a substrate-only (NOSHAPE) condition. An object is termed 'proud' when less than 50% of the object is buried in substrate. Echoes were generated by two bottlenose dolphins that were trained to echolocate on and report the presence or absence of a target that was aligned with the objects. This study describes a systematic evaluation of several two-element ('binaural') signal processing methods and ANN architectures that were applied to the unique dolphin biomimetic sonar (DBS). Object identification accuracy was evaluated over two receive array configurations, four neural network architectures and five signal processing schemes. The target echoes analysed were reverberation-limited. Identification of ensonified targets independent of aspect using reverberation-limited echoes is a substantial leap in what has been expected of generic algorithms. Results of these analyses and summarization of the major findings are reported. A set of recommendations that guide further development of broadband object search strategies in the littoral zone are given.

2. Methods

2.1. Dolphins

The objects were ensonified with the help of two dolphins, HEP (male, 36 years) and TOD (female, 42 years), who were housed in a floating 3-pen complex located in San Diego Bay, Space and Naval Warfare Systems Center. One to two companion dolphins resided with them, but were not present in the experimental pen while sessions were being conducted. A recent audiogram on HEP showed a deficit in the higher frequencies only in the left ear, and an overall hearing deficit in the right ear [29]. Although no free-field audiogram was conducted on TOD, she did not have a medical history with ototoxic drugs. Therefore, normal hearing for her age and gender is expected, which includes the expectation of a high-frequency hearing loss [29].

2.2. Apparatus/design

2.2.1. Configuration. The geometry of the experimental apparatus and targets is illustrated in figure 1(A). A stationing hoop was positioned 1 m underwater, with a secondary detection object, a target object, and three recording hydrophones placed directly on the axis of the biosonar beam. The dolphins were trained to detect the presence or absence of the detection object using echolocation, so thus also fully ensonified the target object. By asking the dolphins to echolocate on a secondary detection object in the water column, this procedure allowed backscatter to be collected from the bottom substrate in the absence of a target object, called the NOSHAPE condition. A sheet-PVC shield coated with closed-cell neoprene blocked echolocation between trials. The hoop station was attached to a floating pier, which caused the dolphin's position to vary in relation to the stationary bottom position of the target object as a function of tide height. Average water depth below the dolphins was 5.71 ± 0.49 m, corresponding to grazing angles of $46.25 \pm 2.33^{\circ}$ and range to the substrate and object of 7.9 ± 0.36 m. Across the study, tide change during sessions resulted in an average range change of 0.03 ± 0.15 m, which was negligible compared to the actual range.

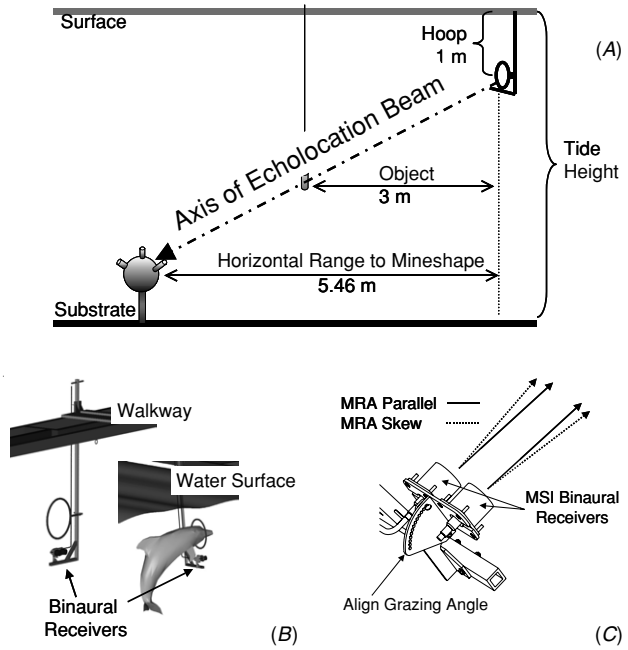

Figure 1. Geometry of the binaural data collection. The top panel (A) illustrates the overall geometry, including the relative position of the dolphins' stationing hoop, secondary detection object and notional object on the bottom. The bottom left panel (B) shows placement of the hoop station and binaural receiving hydrophones relative to the walkway and water, and to the stationed dolphin. The bottom right panel (C) illustrates the attachment of the binaural receiving hydrophones and the two horizontal orientations of the MRAs, parallel and skew (shown parallel).

2.2.2. Electronics. The outgoing click was collected on a Reson model TC-4013 omni-directional hydrophone (receive sensitivity of −211 dB re: 1 V μPa^{-1}) placed 1 m in front of the stationed dolphin, on axis. A gain of +20 dB and bandpass filtration from 12–145 kHz was provided by a DL Electronics 4302 filter/amplifier. Echoic data were collected on a biomimetic binaural receiver consisting of two channels of directional information (figure 1(B)—and (C)). Two circular 1–3 composite hydrophones (d = 50 cm) were specially manufactured for this effort by Material Systems Incorporated (MSI). The MSI binaural receiver was modeled after the dolphin's hearing capabilities in the echolocation band, with both hydrophones having a vertical and horizontal beamwidth of 15° at 120 kHz and 30° at 60 kHz [12]. A gain of +40 dB was added and channels were bandpass filtered from 13–135 kHz. A spatial separation of 12.5 cm between the two MSI receivers mimicked the typical horizontal separation of *Tursiops*' auditory bullae. During data collection, the horizontal receive axes of the MSI hydrophones were either parallel (MSI-P) or skewed (MSI-S). In the skewed condition, 5° plastic shims diverged the receive beams outward by a total of 10°. Placement below the station hoop put the hydrophones in close proximity to the dolphin's lower jaw, and they were angled, based on water depth, toward the substrate/target object. All three channels of acoustic data (click and echo) were simultaneously sampled at 313.94 kHz to 16 bits of resolution using an Interactive Circuits and Systems Ltd Model

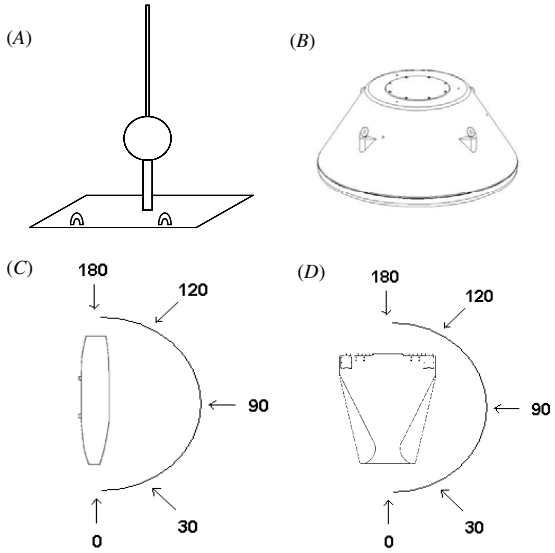

Figure 2. Object shapes, including orientation angles for the aspect-dependent shapes. Panels A and B provide side views of the PDM2 and MANTA, respectively. These are aspect-independent shapes and thus were ensonified at one orientation only. Panels C and D show overhead views of the DST and ROCKAN, respectively. These shapes are aspect dependent, and the arcs indicate the orientations at which they were ensonified.

ICS-130, 32 channel simultaneous sampling analog-to-digital converter VME board. Data collection was hosted on a seven-slot portable VME chassis Unix-based computer (Force 2CE single board computer) and stored to hard drive.

2.2.3. The Binaural'00 echo database. Acoustic backscatter was collected from four metallic objects (PDM2, MANTA, ROCKAN and MK82 DST; figure 2), and substrate only (NOSHAPE). Dimensions ($l \times w \times h$, or $h \times d$, in cm) were as follows: PDM2 = 77 × 77 × 120.5 (sphere diameter = 24), MANTA = 38 × 95, ROCKAN = 100 × 79 × 8–46, DST = 113 × 12–28 × 28. The objects were placed proud on an A1 bottom type (sand/mud, smooth surface, low clutter). Data were collected at various orientations for each object. The PDM2 (figure 2(A)) and MANTA (figure 2(B)) are aspect-independent shapes, therefore only one orientation was necessary for each. Five orientations of the DST (figure 2(C)) and ROCKAN (figure 2(D)) relative to the dolphin's echolocation beam were used; labeled 0°, 30°, 90°, 120° and 180°.

2.2.4. Neural networks. Backpropagation (BPN), genetic reinforcement learning (GLN), learning vector quantization (LVQ) and probabilistic (PNN) artificial neural networks were utilized in this experiment. The networks, collectively referred to as ANNs, were used to produce single-ping identifications, so information from successive pings in echo trains were not combined as a means to improve identification performance [16–23]. Networks were constructed using NeuralWare Professional II/Plus software. To identify object echoes, each ANN architecture was provided with

echo magnitude spectra, using bins ranging from 19.5625–156.5 kHz, with a frequency resolution of 1.22 kHz/bin. Initial testing demonstrated that linear amplitude scaling outperformed logarithmic and power scaling, and that constant-binwidth (FFT type) representations outperformed constant-Q representations (third-octave and wavelet types). Therefore, all reported results have utilized linear amplitude scaling and constant-binwidth representations. All ANN input vector values were normalized between ± 1, as is required [14–26]. This eliminated absolute amplitude differences, thus, object echo identification was based on differences in spectral peaks and notches across object types. In each ANN type, echo input vectors were processed in the architecture-specific hidden layers, after which a set of activation values is computed for the output layer. The output layer had five elements, one each for NOSHAPE, MANTA, PDM2, DST and ROCKAN. The output unit with highest activation value was taken to represent the network's object identification decision.

2.2.5. Signal processing comparisons. Five digital signal processing models were devised to test various methods for exploiting the acoustic backscatter information provided by the MSI binaural receiver. DSP was done on 256-point time domain waveforms, extracting spectral magnitude vectors $\mathbf{M}(f)$ from FFTs prior to ANN object echo identification processing. All the models were tested with data collected under both MSI-P and MSI-S horizontal receive beam conditions. Each model was tested using artificial neural networks to classify the object echoes.

In one model, time series from the left and right binaural channels were summed in place, a notional beam-forming process that emphasizes reflections on-axis and de-emphasizes off-axis backscatter. The summed time series was converted to a simple spectral amplitude representation (BEAM model; see equation (1)),

$$\mathbf{M}(f) = \mathrm{FFT}(\mathbf{X}_\mathrm{L}(t) + \mathbf{X}_\mathrm{R}(t)) \qquad \text{for} \quad 0 \leqslant t \leqslant (\mathrm{npts} - 1), \tag{1}$$

where npts = 256, the spectral magnitude vector was computed from the FFT, and the resulting vector is used as the input to the neural network classifiers as described in section 2.2.4 above.

In the second model, a difference product of the two channels was derived by computing the linear amplitude spectra of each channel and then subtracting the right from the left channel (DELTA model; see equation (2)),

$$\mathbf{I} = \mathbf{M}_\mathrm{L}(f) - \mathbf{M}_\mathrm{R}(f) \qquad \text{for} \quad 19.5 < f \leqslant 156.5\ \mathrm{kHz}, \tag{2}$$

where $\mathbf{M}_\mathrm{L}(f)$ was the spectral magnitude vector for the left channel, and $\mathbf{M}_\mathrm{R}(f)$ was the spectral magnitude vector for the right channel, and $\mathbf{I}$ was the resulting vector used as the input to the neural network classifiers.

For the third model, linear amplitude spectra were computed for each channel, and the two channels concatenated (BOTH model; see equation (3)),

$$\mathbf{I} = \mathbf{M}_\mathrm{L}(f) \cup \mathbf{M}_\mathrm{R}(f) \qquad \text{for} \quad 19.5 < f \leqslant 156.5\ \mathrm{kHz}, \tag{3}$$

where the ANN input vector $\mathbf{I}$ was twice as long as all other DSP processes.

In the fourth model, separate neural networks were trained to classify spectra from the left and right channels, and the object identity derived by linear consensus of the neural network outputs (VOTES model; see equation (4)),

$$V(\mathrm{trial}) = \mathbf{O}_\mathrm{L} \wedge \mathbf{O}_\mathrm{R}, \tag{4}$$

where V(trial) is the vote on any given trial, computed by ANDing the output vector $\mathbf{O}_\mathrm{L}$ from the ANN using $\mathbf{M}_\mathrm{L}$ as the input with the output vector $\mathbf{O}_\mathrm{R}$ from the ANN using $\mathbf{M}_\mathrm{R}$ as the input. If the logical sum of the vectors was 0, then the trial was discarded. If the logical sum of the vectors was 1, then the two networks agreed, the trial was retained and used to create the confusion matrix and to compute the network identification accuracy. Thus, the VOTES scheme differs from the others in that the computation was logical and occurred by comparing the classification pattern of two independent ANNs. All other schemes were implemented prior to ANN processing.

Finally, recall that previous neural network models of dolphin echolocation performance have been monaural [17–23]. Thus, benchmark monaural performance was tested using spectral representation of echoes from the left channel only (MONO model), thus $\mathbf{I} = \mathbf{M}_\mathrm{L}(f)$ for $19.5 < f \leqslant 156.5$ kHz.

2.2.6. Network identification accuracy. Each network was trained with standardized training data, and the networks' accuracy with novel data was tested with standardized generalization data. Ten training and generalization sets were generated from TOD's data contained in the Binaural'00 Echo database. Using random draw without replacement, 250 echoes for each object type (NOSHAPE, MANTA, PDM2, DST, ROCKAN) were selected as training data, and the remaining echoes formed the generalization set. For the aspect-dependent targets (DST, ROCKAN), the 250 echoes were comprised of 50 echoes from each of the five orientations. Thus, the prior probabilities of object type membership were equal. This process was repeated ten times, thereby creating ten standardized data sets that were used to train and test each identification model. In addition to the generalization sets from TOD's data, network accuracy was evaluated using the echoes generated by HEP.

Identification metrics were computed for each combination of horizontal receive beam condition (MSI-P/MSI-S), signal processing scheme and neural network type. Per cent correct identification accuracy is equivalent to the probability of a hit for any given target. Probability of hit (p(H)) and probability of false alarm (p(FA)) were used to estimate the object identification sensitivity index (d') and bias (ln(β), henceforth β) [31]. Because object identification here is a multiple-alternative model, the constant-ratio rule was applied [32]. If the number of hits or number of false alarms was zero, a log-linear correction was used to permit parameter estimation [33]. All ROC statistics were calculated using Information Structuring Systems, Ltd mABX Calculator software, in which d' and β were estimated from p(H) and p(FA) using a jackknifing process. d' can be computed

Table 1. Number of click–echo samples for each object orientation and binaural receive axis position (MSI-P/MSI-S) collected from both dolphins.

		Aspect (°)				
Object	Dolphin	0	30	90	120	180
NOSHAPE	HEP	717/498	–	–	–	–
	TOD	3630/2993	–	–	–	–
PDM2	HEP	9/158	–	–	–	–
	TOD	1136/2287	–	–	–	–
MANTA	HEP	72 / 24	–	–	–	–
	TOD	714/450	–	–	–	–
DST	HEP	260/128	11/355	478/275	430/180	237/29
	TOD	1425/681	599/1749	692/1662	1405/635	741/450
ROCKAN	HEP	185/232	231/230	149/189	258/73	181/155
	TOD	778/941	1174/764	1195/1533	1038/663	801/939

Note: '–' indicates that aspect was irrelevant for these objects.

from p(H) and p(FA) after normalizing or assuming homoscedasticity, which allows the simple relationship $d' = (Z_H - Z_{FA})$ where Z_H is the Z score computed from p(H) and Z_{FA} is the Z score computed from p(FA). β was computed using

$$\beta = \exp\left(-\left(\frac{Z_H}{2}\right)^2 - \left(\frac{Z_{FA}}{2}\right)^2\right) \Big/ \sqrt{2\pi}. \qquad (5)$$

2.2.7. Data sessions. Every object was ensonified by both HEP and TOD in separate 30-trial sessions, with one session per animal conducted in a day. Each animal completed two data sessions for every object at all orientations. In one of the sessions, the horizontal receive axes were parallel (MSI-P) and in the other the axes were skewed (MSI-S). Objects or their orientations were changed across sessions, with baseline substrate-only data (NOSHAPE) collected during the first ten trials after a change (five trials with MSI-P and five with MSI-S). Trials were arranged in blocks of ten, with an equal number of GO (object present) and NO-GO (object absent) trials presented per block. Trial type sequencing was determined by a modified series [34] such that successive trials within a block were controlled (0.5 first-order conditional probability of a GO trial following a NO-GO, or vice versa).

2.3. Procedure

Before every session, water temperature, salinity, general weather conditions, tide height and object type and orientation were recorded. Prior to the start of each trial, the detection object was placed at the appropriate depth if the trial was an 'object present' type. The dolphin, upon receiving a hand gesture, submerged and placed its head in the stationing hoop. The data acquisition program was initiated and the PVC shield raised, allowing the dolphin to echolocate *ad libitum*. The data collection program stayed active for 5 s—an ample time to collect all the outgoing echolocation clicks and acoustic backscatter. Nearby dolphins were prevented from echolocating during this time by having them hold their heads out of the water in a trainer's hand station to avoid potential confounds in the collected click and echo data. The trial was bounded by the same 5 s time frame, whereby the animal would either touch a paddle at the water surface to indicate an 'object present' decision (the 'GO' response) or remain in the hoop for the trial duration to indicate an 'object absent' decision (the 'NO-GO' response). Correct responding was reinforced with a tone and fish reward. Incorrect responses occurred if the dolphin responded 'NO-GO' when the object was present (miss) or 'GO' when it was absent (false alarm). The behavioral paradigm was used to ensure that the dolphins actively searched for and ensonified the target objects shapes.

3. Results

3.1. Click and object echo characteristics

Detailed analysis of the dolphins' echolocation signal structures was conducted using 36 819 click–echo pairs, 5744 from HEP and 31 075 from TOD (table 1). TOD produced about six times more echolocation clicks than HEP, probably due to her poor visual capabilities. Differences in the acoustical properties of the two dolphins' echolocation clicks and corresponding acoustic backscatter were examined. Six scalar measurements were made for each click and echo: the peak–peak amplitude (p–p, in dB re: 1 μPa), energy flux density (EFD, in dB re: 1 μPa^2 s Hz^{-1}), frequency at maximum power (kHz), half-power bandwidth (kHz), center frequency (kHz), and RMS bandwidth (kHz). Click amplitudes and energy flux densities were source levels (SL), while echo amplitudes were received levels (RL). All spectral measures were made across the full 156.5 kHz Nyquist bandwidth. Separate two-sided t-tests for unequal variances were run on the outgoing clicks and returning echoes using Statistica software. Sample sizes were large, thus we employed a conservative alpha ($\alpha = 0.01$) to test for significance. The t-tests revealed differences in the echolocation clicks produced by the two dolphins, all significant at the $\alpha = 0.01$ level. HEP's clicks had substantially lower source levels and were higher in frequency compared to TOD's clicks; but bandwidth was comparable (although statistically different). The differences are summarized in table 2, and typical clicks for HEP and TOD are illustrated in figure 3.

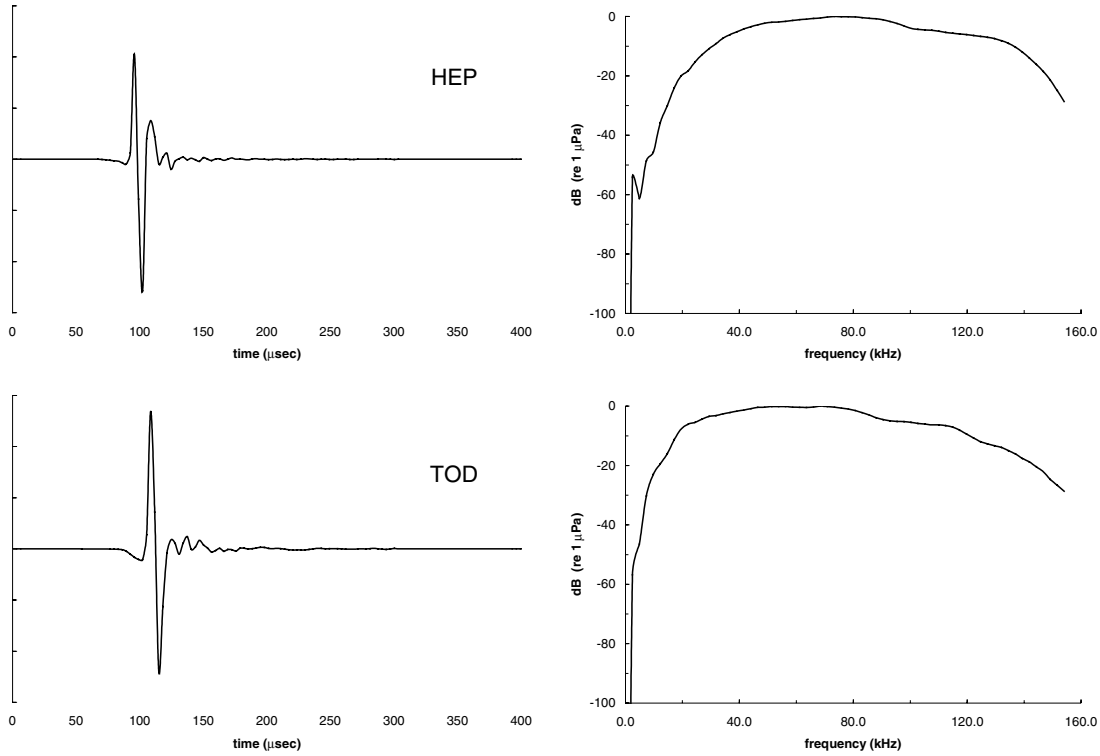

Figure 3. Sample echolocation clicks for each dolphin. HEP's sample is in the top row, TOD's in the lower row. The left panels illustrate the time waveforms, and the right panels show the power spectra. Time is in μs, frequency in kHz, and spectral amplitude in relative dB (re: 1 μPa).

Table 2. Summary statistics for the dolphins' clicks. All *t*-tests were significant at $\alpha = 0.01$.

Click property	*t* value (df)	Mean (±sd) HEP	Mean (±sd) TOD
SL (p–p)	50.89 (10347.75)	182.97 (4.99)	186.88 (6.95)
SL (EFD)	45.60 (10731.98)	126.58 (4.73)	129.93 (6.83)
Peak *F*	−28.37 (9198.81)	75.99 (23.67)	65.99 (28.84)
3 dB BW	8.78 (10878.69)	53.69 (15.57)	55.82 (22.79)
Center *F*	−28.08 (10166.50)	74.14 (11.27)	69.29 (15.39)
RMS BW	45.57 (9313.53)	23.83 (3.09)	25.93 (3.82)

The animal's outgoing click provided the synchronization signal for extracting the object echo data at the correct range gate. The speed of sound in water, computed from the session parameter information, was also used to extract the echoes from the data streams. Table 3 presents mean values (±1 sd) of several key acoustic characteristics of the object echoes by dolphin and target type. Because the objects were proud on the bottom, the backscatter was convolved with bottom reverberation and it was not possible to estimate target strength. Thus, echo amplitudes were described using simple received levels (RL) at the calibrated hydrophone. Because the

Table 3. Descriptive statistics for the echoes collected from HEP (top) and TOD (bottom).

Object	*N*	RL (p–p)	RL (EFD)	Peak *F*	3 dB BW	Center *F*	RMS BW
	Mean values (±sd) for TOD						
NOSHAPE	1215	120.26 (5.46)	74.44 (4.06)	23.10 (23.46)	4.00 (1.92)	44.48 (13.99)	27.99 (4.79)
PDM2	167	118.12 (4.34)	74.53 (3.72)	20.36 (17.85)	3.85 (1.38)	44.79 (15.41)	29.50 (4.44)
MANTA	96	120.76 (3.56)	75.60 (3.00)	22.71 (19.62)	3.79 (1.35)	43.39 (10.11)	28.69 (3.81)
DST	2383	120.96 (6.05)	75.22 (3.81)	29.03 (26.55)	3.94 (1.76)	46.36 (16.31)	27.48 (4.63)
ROCKAN	1883	120.39 (4.93)	74.00 (3.27)	25.29 (27.21)	4.13 (1.92)	47.52 (15.82)	29.69 (4.42)
	Mean values (±sd) for TOD						
NOSHAPE	6623	125.26 (6.13)	80.10 (4.30)	33.30 (23.47)	3.81 (1.74)	48.26 (13.28)	27.18 (4.28)
PDM2	3423	125.50 (4.28)	80.76 (3.12)	29.07 (19.98)	3.50 (1.38)	43.94 (9.79)	25.02 (2.99)
MANTA	1164	130.58 (5.15)	83.69 (4.15)	40.30 (16.79)	3.76 (1.54)	47.69 (9.39)	22.94 (2.74)
DST	10039	127.50 (6.85)	79.90 (4.56)	38.43 (26.91)	3.71 (1.78)	52.20 (14.07)	27.83 (4.50)
ROCKAN	9826	130.80 (6.63)	82.79 (4.98)	47.24 (23.42)	4.08 (1.96)	56.54 (12.33)	26.46 (3.73)

Table 4. Generalization sensitivity (d') and bias (β) for a selection of networks, broken out by object type.

	Network type			
Object type	BPN	GLN	LVQ	PNN
NOSHAPE	1.736 (0.865)	0.177 (0.145)	2.226 (0.678)	2.715 (−4.176)
MANTA	3.635 (0.550)	1.980 (0.492)	4.288 (1.333)	5.451 (6.745)
PDM2	3.173 (0.451)	1.462 (0.157)	3.662 (0.915)	4.163 (5.016)
DST	2.158 (0.687)	0.520 (0.349)	2.316 (0.650)	2.829 (5.117)
ROCKAN	2.121 (0.643)	0.485 (0.361)	2.355 (0.699)	3.093 (5.680)

configuration of the calibrated hydrophone was constant, the RL provided a metric for comparing amplitudes.

Note the differences between the acoustic properties of the echo + reverberation in table 3 and the dolphin's ensonification signals in table 2. The peak–peak received levels are about 36 dB down after returning from the target object and bottom and traversing approximately 16 m. Peak frequencies have dropped almost an octave, and bandwidth is significantly reduced. At 313.94 kHz A/D rate and average sound speed of about 1512 m s^{-1}, the range window represented by the 1024-pt echo samples was approximately 4.95 m. This range is large compared to the object shape dimensions, thus bottom reverberation seems to dominate the backscatter. 'Object present' conditions had variable signal excess over the 'target absent' (NOSHAPE) condition, suggesting variable signal-to-reverberation levels that would contribute to variation in object identification accuracy. Moreover, variations among the frequency and bandwidth of the echoes were complex, suggesting that identification accuracy would be lower for the aspect-dependent objects (DST, ROCKAN) than for the aspect-independent shapes (PDM2, MANTA).

3.2. Evaluation of ANNs

To gauge the relative accuracy of the four networks (BPN, GLN, LVQ, PNN), a set of networks were trained using TOD's MONO, MSI-P signal processing scheme, and the generalization accuracy measured across the ten data sets. RMS error in the BPN reached asymptote at about 25% to 30%, thus all BPN networks were trained until RMS error reached 27% or 500 000 epochs had elapsed. Each GLN was run for 10 000 generations, 50 parents per generation, using identification accuracy as the metric (which required about 12 h per network). The LVQ were trained until RMS error reached 5% or 56 250 epochs had elapsed. The PNN recall parameter sigma was optimized numerically using holdout training [35]; the values converged to 0.4 for all training sets. For each network type, the confusion matrices were pooled and estimates of d' and β were computed.

The results are presented in table 4 and inspection reveals several conclusions. One, the GLN was characterized by poor sensitivity across all object types (NOSHAPE, MANTA, PDM2, DST, ROCKAN). Two, the BPN and the Parzen classifiers (LVQ and PNN) were characterized by good sensitivity. However, the PNN was plagued by biased categorization, with many DST and ROCKAN echoes mistakenly identified as the NOSHAPE type. Thus, we eliminated the GLN and PNN architectures from further consideration.

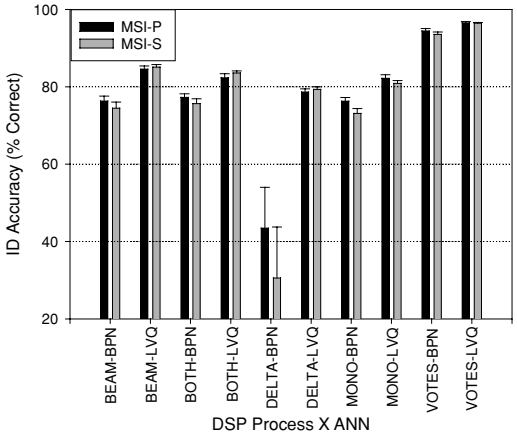

Figure 4. Average identification accuracy for the set of signal processing and neural network combinations when the binaural receiver axis alignment was parallel (MSI-P) and skewed (MSI-S). Error bars are ±1 sd.

3.3. Evaluation of receive axis alignment and signal processing schemes

The relative utility of the five signal processing schemes (MONO, BEAM, BOTH, DELTA, VOTES) and binaural receiver axis alignments (MSI-P, MSI-S) were tested using only the BPN and LVQ architectures. Separate networks were created for each combination of signal processing and receiver configuration. The networks were trained using the ten standardized learning sets and accuracy was tested across the ten standardized generalization sets.

For the VOTES scheme, two networks were trained, one for the left channel ('left ear') and one for the right ('right ear'). For each echo in the generalization set, the object category decisions for the left and right were compared. The confusion matrix was compiled using only those echoes for which agreement was observed for both channels, and the 'per cent binaural agreement' also was computed. For the MSI-P receiver configuration, binaural agreement for the BPN was 61.3%. Thus, 61.3% of the echoes were included in the accuracy, sensitivity and bias calculations. Binaural agreement was 75.1% for the LVQ networks. For the MSI-S configuration, binaural agreement for the BPN was 58.2% and 79.9% for the LVQ networks.

To obtain an overview of the relative accuracy of the networks, the overall per cent correct was computed for each. The per cent corrects were averaged over the ten data sets, and are presented in figure 4. All signal processing combinations resulted in above-chance identification accuracy (chance for

five target categories is 20%). Manipulation of binaural receiver axis alignment had a negligible effect on identification (ID) accuracy, with mean accuracy of 79.2% correct for MSI-P and 77.2% for MSI-S. However, the most obvious effect is the interaction between signal processing schemes and neural network type, most notably the highly variable, poor performance of the BPN with DELTA processing. A two-way ANOVA (5 DSP schemes × 2 ANN types + interaction) supports these observations, $F(9190) = 276.28$, $p < 0.01$, with interaction and both main effects significant. After excluding the DELTA signal processing scheme, the LVQ slightly outperformed the BPN (mean% correct = 86.4 versus 80.1, respectively). The variability in ID accuracy introduced by the different signal processing schemes was more significant, with the MONO scheme having the lowest mean% correct and VOTES the highest. Duncan's multiple range post hoc test ($\alpha = 0.01$) revealed that the identification accuracy for the VOTES and MONO schemes were significantly different, but that BEAM and BOTH were not. The VOTES scheme, however, benefited from error reduction by elimination of samples (40% in the case of the BPN × MSI-S combination) when left and right channel IDs did not match. Given the results, we did not further analyse the MSI-S receiver alignment data, nor the DELTA and VOTES signal processing schemes.

3.4. Final evaluations

3.4.1. Comparison of ID accuracy across object type. We retained the MSI-P data, and conducted a final evaluation of the BEAM, BOTH and the MONO benchmark signal processing schemes. For each combination of signal processing and neural network, the confusion matrices for the ten standardized data sets were pooled, from which estimates of identification sensitivity (d') and bias (β) were computed (presented in table 5). Differences between object types are evident, but only subtle differences were revealed between signal processing schemes (DSP) and neural network architectures (ANN). LVQ sensitivity always was slightly greater than that of the BPN. When the results were collapsed across target types, a two-way ANOVA (3 DSP schemes × 2 network types) revealed no interaction and no effect of DSP scheme. After further collapsing across DSP scheme, a *t*-test confirmed the significant effect of ANN on sensitivity ($t(298) = 20.54$, $p < 0.01$), with LVQ more sensitive than BPN. Qualitatively, the lowest mean sensitivity ($d' = 2.584$) was obtained for the combination of monaural DSP (MONO) and backpropagation (BPN), and the highest sensitivity ($d' = 3.249$) was obtained using binaural beam-formed DSP (BEAM) and learning vector quantization (LVQ).

3.4.2. Accuracy across orientations for the aspect-dependent objects. Identification of the two aspect-dependent objects (DST and ROCKAN) was extremely accurate in spite of the disparity in the echoes across 180° of aspect change and large jumps in aspect angle (see figure 2). Identification accuracy for each aspect angle is presented as a function of signal processing and neural network in figure 5. The BEAM and BOTH schemes outperformed MONO for both targets. Additionally, LVQ outperformed BPN, especially for orientations of 30°, 90° and 120°, consistent with the results above where all object types were pooled.

Table 5. Generalization accuracy (p(H), p(FA)), sensitivity (d') and bias (β) for each object type. Signal processing (DSP) and network architecture (ANN) are contrasted.

Object	DSP	ANN	p(H)	p(FA)	d'	β
NOSHAPE	MONO	BPN	0.644	0.085	1.743	0.874
NOSHAPE	MONO	LVQ	0.791	0.077	2.234	0.690
NOSHAPE	BEAM	BPN	0.647	0.099	1.665	0.758
NOSHAPE	BEAM	LVQ	0.825	0.066	2.442	0.701
NOSHAPE	BOTH	BPN	0.672	0.090	1.784	0.800
NOSHAPE	BOTH	LVQ	0.788	0.076	2.232	0.707
MANTA	MONO	BPN	0.952	0.024	3.637	0.553
MANTA	MONO	LVQ	0.967	0.007	4.288	1.333
MANTA	BEAM	BPN	0.969	0.020	3.298	0.360
MANTA	BEAM	LVQ	0.981	0.004	4.758	1.441
MANTA	BOTH	BPN	0.959	0.023	3.735	0.485
MANTA	BOTH	LVQ	0.965	0.003	4.525	2.018
PDM2	MONO	BPN	0.926	0.041	3.185	0.470
PDM2	MONO	LVQ	0.943	0.018	3.673	0.935
PDM2	BEAM	BPN	0.937	0.033	3.365	0.512
PDM2	BEAM	LVQ	0.956	0.013	3.947	1.040
PDM2	BOTH	BPN	0.938	0.039	3.302	0.359
PDM2	BOTH	LVQ	0.945	0.015	3.776	1.089
DST	MONO	BPN	0.779	0.079	2.176	0.698
DST	MONO	LVQ	0.811	0.073	2.336	0.666
DST	BEAM	BPN	0.767	0.068	2.225	0.850
DST	BEAM	LVQ	0.821	0.065	2.434	0.730
DST	BOTH	BPN	0.775	0.068	2.251	0.831
DST	BOTH	LVQ	0.811	0.083	2.270	0.572
ROCKAN	MONO	BPN	0.783	0.086	2.146	0.624
ROCKAN	MONO	LVQ	0.817	0.070	2.381	0.679
ROCKAN	BEAM	BPN	0.789	0.099	2.088	0.505
ROCKAN	BEAM	LVQ	0.851	0.068	2.530	0.566
ROCKAN	BOTH	BPN	0.793	0.082	2.208	0.634
ROCKAN	BOTH	LVQ	0.823	0.070	2.404	0.663

3.4.3. Generalization to novel dolphin data. Using the most accurate signal processing combination (BEAM-LVQ), a second test of generalization accuracy was conducted that contrasted TOD's results with ROC estimates that summarize ID accuracy of HEP's echo data set (table 6). Recall that the BEAM-ANN system was trained with TOD's data, thus here we have a first test of generalization across 'sonar sources' (the two dolphins). The data clearly demonstrate that the BEAM-LVQ processing scheme could classify MANTA and PDM2 echoes generated by HEP. However, sensitivity to the NOSHAPE condition and aspect-dependent DST and ROCKAN echoes was low, and bias suggests a great proportion of false alarms to the aspect-independent MANTA and PDM2 categories.

4. Discussion

The development of a binaural dolphin biomimetic sonar receiver produced data that were used to study signal processing methods for object identification. Echo spectra from four metallic objects proud on the ocean floor and a

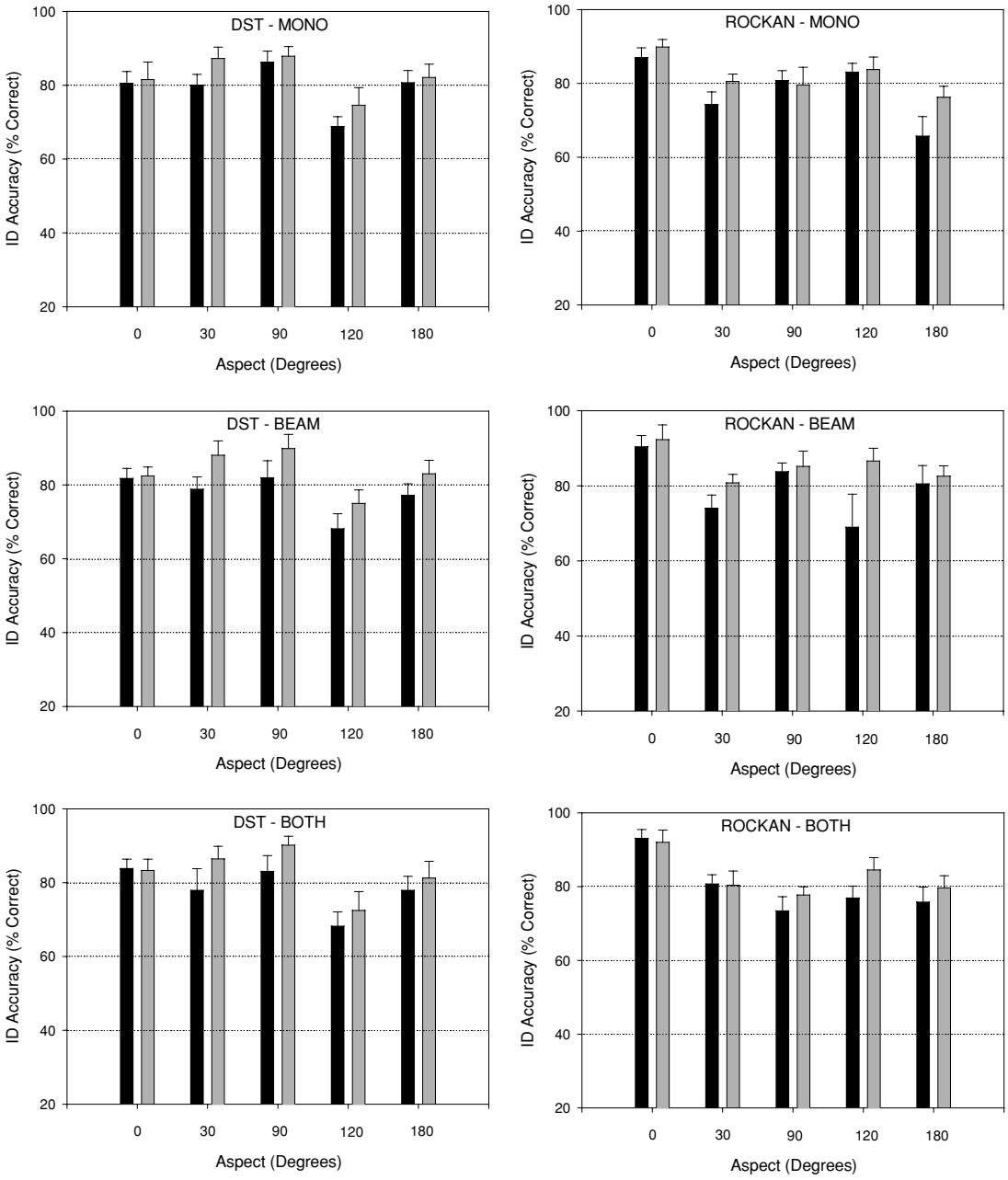

Figure 5. Average identification accuracy for the DST (left panels) and ROCKAN (right panels) at five aspect angles (+1 sd). Performance of the BPN and LVQ networks is contrasted across the MONO, BEAM and BOTH signal processing schemes. Results from BPN are in black and LVQ in gray.

Table 6. Generalization sensitivity (d') and bias (β) for the BEAM processing and LVQ network, contrasting dolphin sources and broken out by object type.

	Dolphin	
Object type	TOD (d' and (β))	HEP (d' and (β))
NOSHAPE	2.442 (0.701)	0.217 (0.083)
MANTA	4.758 (1.441)	2.843 (2.099)
PDM2	3.947 (1.040)	2.537 (2.048)
DST	2.434 (0.729)	0.374 (0.241)
ROCKAN	2.529 (0.566)	0.426 (0.118)

no-object, substrate-only condition were classified. Accuracy was evaluated over two receive array configurations, four neural network architectures and five signal processing schemes. All signal processing combinations resulted in above-chance identification accuracy. The horizontal axes of the binaural receive array beams were either parallel or diverged 10°. This manipulation did not significantly affect object identification accuracy. The four neural networks included standard parametric backpropagation (BPN) and nonparametric learning vector quantization (LVQ) architectures, along with genetic learning (GLN) and

probabilistic (PNN) network architectures. The GLN was inaccurate and inefficient. The PNN was about as accurate on average as LVQ, but was characterized by high bias. LVQ was consistently the most accurate network across all signal processing schemes (86% correct across all objects and signal processing schemes), but BPN was also accurate (80% correct).

The processing schemes included four methods that capitalized on the binaural data, plus a monaural benchmark process (MONO). Of these schemes, a differencing process (subtraction of spectra, DELTA) resulted in lowest accuracy, and a voting process that compared identification at each 'ear' independently (VOTES) resulted in the highest accuracy. However, the voting process also resulted in disagreement on, and thus exclusion of, a substantial portion of the backscatter samples. If emission of sonar pings does not have a cost (i.e., can occur *ad libitum*), then the VOTES processing scheme is optimal. The remaining binaural processes of simple beamforming (BEAM) or concatenation of spectra from both 'ears' (BOTH) outperformed the monaural benchmark, with higher sensitivity and lower bias. Identification of the two aspect-dependent objects (DST and ROCKAN) was accurate, given the disparity in the echoes across 180° of aspect change and large jumps in aspect angle. The binaural processing schemes (BEAM and BOTH) outperformed the monaural benchmark for both objects. Additionally, the LVQ network outperformed the BPN, especially for orientations of 30°, 90° and 120°.

For the present data, the nonparametric LVQ network and binaural BEAM signal processing scheme resulted in highest object identification accuracy without exclusion of backscatter samples. Moreover, generalization tests revealed that the BEAM-LVQ processing scheme could classify MANTA and PDM2 echoes generated by the second dolphin, HEP. However, sensitivity to the NOSHAPE condition and aspect-dependent DST and ROCKAN echoes was low, and bias suggests a great proportion of false alarms to the MANTA and PDM2 categories. Thus, the echo spectral features that were suitable for good identification of objects ensonified by the dolphin TOD did not appear to be readily available in the echoes generated by the dolphin HEP. This could have been caused by spectral differences in the dolphins' biosonar signals, or by the lower overall source level of HEP's signals.

The application of a binaural receive array presents a significant advancement in bio-inspired signal processing research for object identification. Additionally, although the application of neural networks to sonar pattern recognition is not new, the questions posed in this study were addressed with complex, real-world data. For example, many biosonar classification studies focused on objects ensonified in the free-field [17–22]. The current study used backscatter generated from objects placed on the bottom, which substantially degrades the acoustic backscatter by convolution of target echo with substrate reverberation. The nonstationary nature of free-field ambient noise in earlier studies had acoustic characteristics that were substantially separable from the coherent echoes generated by the objects. The object echoes analysed here were reverberation-limited, thus the acoustic characteristics of the echoes are very similar to the characteristics of the interfering backscatter. The ANN architectures used here were generic commercial off-the-shelf architectures designed for generalized pattern classification, rather than algorithms engineered specifically for identification of ensonified bottom objects using specifically engineered sonar signals. In a nutshell, identification of ensonified targets independent of aspect using acoustic properties of reverberation-limited echoes is a substantial leap in what has been expected of generic algorithms in previous studies (e.g., [17–22]).

The advantages of multi-element signal processing are demonstrated in this evaluation. The binaural BEAM processing scheme is computationally inexpensive and resulted in improvement of identification accuracy over the monaural scheme. Previous work [36] and the current data suggest that the nonparametric LVQ classifier will outperform the parametric BPN, and identification accuracy likely will improve above the 86% observed here as the number of orientations that are represented are increased. However, very-high-resolution short-range sonars may generate highly aspect-dependent backscatter, with acoustic characteristics that may be indicative of non-diagnostic features of the objects (such as a seam or lip). True integration of echoes from multiple orientations does not occur within the LVQ processing scheme, and thus we remain confident that object identification will be optimized through application of sonar imaging techniques [37, 38].

Acknowledgments

The acoustic backscatter data analysed in this paper were collected through the persistence and patience of Jennifer Cull, Justine Zafran, and Jenny Briar, who trained HEP and TOD to perform the behavioral task and ran all data collection trials. We thank Mike Phillips (SSC-SD Code 2374), who developed the digital data collection software. Project ALTER was possible thanks to funding from Ms Nancy Harned, Dr John Tague, and Mr Dave Armstrong in ONR321US.

References

[1] Altes R A 1995 Signal processing for target recognition in biosonar *Neural Netw.* **8** 1275–95
[2] Au W W L 1993 *The Sonar of Dolphins* (New York: Springer) 277 pp
[3] R-G Busnel and J F Fish (ed) 1980 *Animal Sonar Systems* (New York: Plenum) 1135 pp
[4] Evans W E 1973 Echolocation by marine delphinids and one species of freshwater dolphin *J. Acoust. Soc. Am.* **54** 191–9
[5] P E Nachtigall and P W B Moore (ed) 1988 *Animal Sonar: Processes and Performance* (New York: Plenum) 862 pp
[6] Thomas J A and Kastelein R A (ed) 1990 *Sensory Abilities of Cetaceans: Laboratory and Field Evidence* (New York: Plenum) 710 pp
[7] Au W W L 1980 Echolocation signals of the Atlantic bottlenose dolphin (*Tursiops truncatus*) in open waters *Animal Sonar Systems* ed R-G Busnel and J F Fish (New York: Plenum) pp 251–82
[8] Au W W L 1974 Measurement of echolocation signals of the Atlantic bottlenose dolphin, *Tursiops truncatus* Montagu, in open waters *J. Acoust. Soc. Am.* **56** 1280–90

[9] Houser D S, Helweg D A and Moore P W 1999 Classification of dolphin echolocation clicks by energy and frequency distributions *J. Acoust. Soc. Am.* **106** 1579–85
[10] Moore P W B and Pawloski D 1990 Investigations on the control of echolocation pulses in the dolphin (*Tursiops truncatus*) *Sensory Abilities of Cetaceans* ed J Thomas and R Kastelein (New York: Plenum) pp 305–16
[11] Moore P W B 1997 Mine hunting dolphins of the Navy *Proc. SPIE* **3079** 1–6
[12] Au W W L and Moore P W B 1984 Receiving beam patterns and directivity indices of the Atlantic bottlenose dolphin *Tursiops truncatus J. Acoust. Soc. Am.* **75** 255–62
[13] Blauert J 1997 *Spatial Hearing: The Psychophysics of Human Sound Localization* (Cambridge, MA: MIT Press) 494 pp
[14] Gilkey R H and Anderson T R 1997 *Binaural Hearing in Real and Virtual Environments* (Mahwah, NJ: Lawrence Erlbaum) 795 pp
[15] H L Hawkins, T A McMullen, A N Popper and R R Fay (ed) 1996 *Auditory Computation* (New York: Springer) 517 pp
[16] Gorman R P and Sejnowski T J 1988 Analysis of hidden units in a layered network trained to classify sonar targets *Neural Netw.* **1** 75–89
[17] Roitblat H L, Moore P W B, Nachtigall P E, Penner R H and Au W W L 1989 Natural echolocation with an artificial neural network *Int. J. Neural Netw.* **1** 239–48
[18] Moore P W B, Roitblat H L, Penner R H and Nachtigall P E 1991 Recognizing successive dolphin echoes with an Integrator Gateway Network *Neural Netw.* **4** 701–9
[19] Roitblat H L, Moore P W B, Nachtigall P E and Penner R H 1991 Natural dolphin echo recognition using an Integrator Gateway Network *Adv. Neural Inf. Process. Syst.* **3** 273–81
[20] Helweg D A, Au W W L, Roitblat H L and Nachtigall P E 1996 Acoustic basis of recognition of aspect-dependent objects by an echolocating bottlenose dolphin *J. Acoust. Soc. Am.* **99** 2409–20
[21] Roitblat H L, Moore P W B, Helweg D A and Nachtigall P E 1993 Representation and processing of acoustic information in a biomimetic neural network *From Animals to Animals 2: Simulation of Adaptive Behavior* ed J-A Meyer, S W Wilson and H L Roitblat (Cambridge, MA: MIT Press) pp 90–9
[22] Au W W L 1994 Comparison of sonar discrimination: dolphin and an artificial neural network *J. Acoust. Soc. Am.* **95** 1–8
[23] Roitblat H L, Au W W L, Nachtigall P E, Shizumura R and Moons G 1995 Sonar recognition of targets embedded in sediment *Neural Netw.* **8** 1263–73
[24] Kohonen T 2001 *Self-Organization Maps* 3rd edn (New York: Springer) 501 pp
[25] Rojas R 1996 *Neural Networks: A Systematic Introduction* (New York: Springer) 502 pp
[26] Wasserman P D 1989 *Neural Computing: Theory and Practice* (New York: Van Nostrand-Reinhold) 230 pp
[27] Goldberg D E 1989 *Genetic Algorithms: In Search, Optimization and Machine Learning* (Reading, MA: Addison-Wesley) 412 pp
[28] Fukunaga K and Hummels D M 1987 Bayes error estimation using Parzen and K-NN procedures *IEEE Trans. Pattern Anal. Mach. Intell.* **9** 634–43
[29] Brill R L, Moore P W B and Dankiewicz L A 2001 Assessment of dolphin (*Tursiops truncatus*) auditory sensitivity and hearing loss using jawphones *J. Acoust. Soc. Am.* **109** 1717–22
[30] Floyd R W 1980 Models of cetacean signal processing *Animal Sonar Systems* ed R-G Busnel and J F Fish (New York: Plenum) pp 615–23
[31] Green D M and Swets J A 1988 *Signal Detection Theory and Psychophysics* (Los Altos, CA: Peninsula) 505 pp
[32] Clark F R 1957 Constant-ratio rule for confusion matrices in speech communication *J. Acoust. Soc. Am.* **29** 715–20
[33] Hautus M J 1995 Corrections for extreme proportions and their biasing effects on estimated values of d' *Behav. Res. Methods Instrum. Comput.* **27** 46–51
[34] Gellermann L W 1933 Chance orders of alternating stimuli in visual discrimination experiments *J. Gen. Psychol.* **42** 206–8
[35] Specht D F 1990 Probabilistic neural networks and the polynomial Adaline as complementary techniques for classification *Neural Netw.* **1** 111–21
[36] Helweg D A, Houser D S and Moore P W 2006 Evaluation of algorithms for detection, classification, and identification of proud VSW mine simulators ensonified by an echolocating bottlenose dolphin *U.S. Navy J. Underw. Acoust.* **53** 63–88
[37] Altes R A, Moore P W B and Helweg D A 1998 Tomographic image reconstruction of MCM targets using synthetic dolphin signals *SSC San Diego Technical Document* 2993, DTIS
[38] Altes R A, Helweg D A and Moore P W 2001 Biologically inspired synthetic aperture sonar *SSC San Diego Technical Report* 1848, DTIS