A Study on the Classification of Underwater Acoustic Signal Using an Artificial Neural Network

신경회로망을 이용한 수중음향 신호의 식별에 관한 연구

Young-Nam Na*, Taebo Shim*, Jeong-Woo Han*, Chun-Duck Kim**
나 영 남*, 심 태 보*, 한 정 우*, 김 천 덕**

ABSTRACT

In this study, we examine the applicability of the classifier based on an artificial neural network (ANN) for the low-frequency acoustic signal in shallow water environment. The estimations of the Doppler shift and frequency spreading effect at 220 Hz reveal the frequency variation of less than 2 Hz with time. This small variation enables the ANN-based classifier to identify signals using only tonal frequency information. The ANN consists of 4 layers, and has 60 input processing elements (PEs) and 4 output PEs, respectively. When measured tonal signals in the frequency 200-250 Hz are applied to the ANN-based classifier, the classifier can identify more than 67% of the signals for instantaneous frame and more than 91% for averaged one over 5 frames.

요 약

본 연구에서는 천해환경에서 저주파 음향신호의 신경회로망에 기초한 식별시스템 적용 가능성을 살펴 본다. 220 Hz 주파수에서 도플러 변이와 주파수 확산 효과를 추정한 결과 시간에 따라서 2 Hz 이하의 변화를 보인다. 이러한 주파수의 작은 변화는 신경회로망에 기초한 식별시스템이 단지 토널 주파수 정보만으로도 신호의 식별을 가능하게 한다. 신경회로망은 모두 4개의 층으로 이뤄져 있으며, 입력과 출력 처리요소는 각각 60개와 4개로 구성되어 있다. 주파수 200-250 Hz 대역에서 실측한 토널 신호를 신경회로망에 기초한 식별시스템에 입력시킨 결과 순간적인 프레임의 경우에 대해서는 67% 이상, 그리고 연속되는 5개의 프레임을 평균한 경우에 대해서는 91% 이상의 신호를 식별할 수 있다.

I. Introduction

Classification of underwater contacts from sonar was mainly conducted by trained human operators by listening the acoustic signal. But with the advent of high speed computers, visual displays have become available to increase the auditory information. This in trun has enabled sonar operators to process visual and auditory information simultaneously, yielding faster reaction times and improving their interpretations.[1-3] The displays are very complex and require that potential sonar operators go through extensive training as well as memory tests for visual displays and auditory pitch. Moreover, as sensor technology develops and the threat becomes increasingly sophisticated, this task is becoming more and more difficult due to the increasing volume and complexity of data available for the processing. This gives rise to a substantial increase in operator workload although many techniques concerning digital signal processing guarantee the successful contacts.

In an attempt to solve these problems, an artificial neural network (ANN) is one of the new, basically different signal processing approaches that is currently receiving much attention for weapon applications.[4-6] An ANN technology attempts to mathmatically and/or electrically model neurons and synapses,

* ADD, Chinhae
진해 국방과학연구소
** National Fisheries University of Pusan
부산 수산대학교
접수일자 : 1995년 5월 11일

and then interconnect these models in architectures suitabile for signal processing tasks.

In shallow water, however, an acoustic signal may be distorted by the environments through which it propagates. That is, it inevitably undergoes the energy loss and scattering in the time domain due to multipath propagation (or Doppler scattering in the frequency domain). The spreads in frequency and angle of an acoustic signal reflecting from a fixed ocean bottom may exist in the presence of source and receiver motion. Moreover, an acoustic signal may be corrupted by background noise.[6]

In this paper, we examine the possibility of applying an ANN in classifying underwater acoustic signal in shallow water environment. We employ the ANN which is based on the backpropagation learning scheme to identify four different tones in the frequency of 200-250 Hz. Although many other factors may be considered in classifying signals by passive sonar, tonal frequency information from narrow band processing is still the most important one. We attempt to estimate the number of frequency bins on which most of the acoustic energy exists. The background noise is normalized using the Order Truncate Average (OTA) scheme.[7] The ANN is trained with tonal frequency information and then tested with the measured data in shallow water. Finally, the ANN results are compared with that of the so called "eye integration" by humankind.

II. Data Measurement and Analysis

Acoustic data to be applied to the ANN are spectrum levels with frequency. The data have been collected in the Southeast Sea of Korea, where the water depth is around 150 m and the bottom sediment consists of sand, silt and clay. The sound source emitting tones in the frequencies of 204.7, 216.0, 229.1, and 240.4 Hz, was towed at around 10 kts and at 20 m depth. The receiver was a horizontal line array having 11 sensor elements and was installed on the sea bottom. Signal arriving at a RF receiver is decoded and recorded for further processing (Fig.1). Data sampling rate in analog-to-digital conversion is 24.4 kHz and the frequency resolution of power spectrum is 0.1875 Hz.

Figure 2 shows the procedure to identify the signal using an ANN. The digitized data are transformed into frequency domain by the discrete Fourier trans-

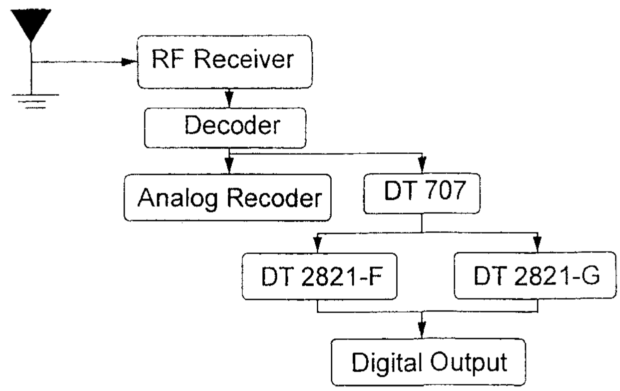

Fig.1 Signal flow from a RF receiver to digital output.

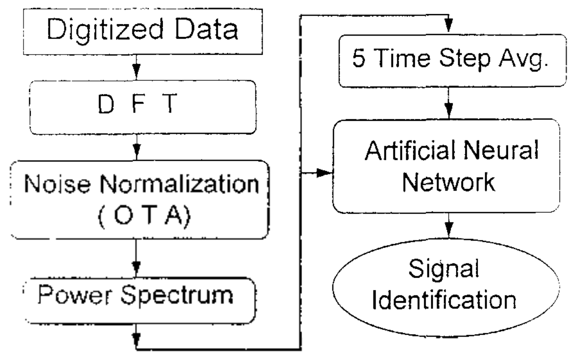

Fig.2 Procedure to classify tonal signals using an ANN.

form (DFT) and then normalized by the OTA scheme. Normalization is the process of whitening the background noise spectrum. The OTA normalizer was developed in 1978 by Wolcin.[7] The following steps describe the OTA technique: (1) The K bin values in the k-th set of bins Ω_k are ordered to form a new sequence (Y_1, Y_2, Y_3, ..., Y_K), where Y_1 is the smallest and Y_K is the largest. (2) The sample median Y_M is identified, and all bins having values greater than rY_M are excluded. A scale factor r is a function of K, and it is approximately 1.4 for $K=11$. Assume L bins are left after the exclusion process. (3) The noise mean estimate μ_k is then obtained using L remaining bins:

$$\mu_k = \sum_{i=1}^{L} \frac{Y_i}{L} \tag{1}$$

The OTA is proven to be appropriate when multiple tones are present.[8,9] The normalized power spectrum levels are averaged over five frame steps to lower the false alarm rate of the classifier. The spectrum level information is applied to the ANN to identify the four signals.

In shallow water acoustic signals experience the energy loss and the Doppler shift in the frequency domain. Doppler shift can occur, in particular, due to the effects of inner waves and currents.[6] Along with these effects, random positional variations of source and receiver are often of fundamental and dominant importance in the high frequency. Hence, proper number of frequency bins or bandwidths should be estimated so that most of the energy can be included within the selected range of bins.

The Doppler-shifted frequency f_1 of a transmitted CW signal depends on the frequency f_o emitted by the source, the path geometry, as well as on the source and receiver velocities.[6] For one-way transmission, the shifted frequency can be given as follows :

$$\frac{f_1}{f_o} = \frac{1+(V_r/c)\cos(\varphi-2\theta)}{1-(V_s/c)\cos\varphi} \tag{2}$$

where, V_s = source velocity,
V_r = receiver velocity,
θ = bottom slope,
φ = acoustic ray angle to the horizontal surface,
c = sound speed.

The notations in the equation are shown in Fig.3. If the velocities of a source and a receiver are very small compared with the sound velocity, that is, $V_s/c \ll 1$ and $V_r/c \ll 1$, the Eq.(2) may be approximated as

$$\frac{(f_1-f_o)}{f_o} = \frac{\Delta f_1}{f_o} \approx (V_s/c)\cos\varphi + (V_r/c)\cos(\varphi-2\theta)$$
$$= (V_s/c)\left[\cos\varphi + (V_r/V_s)\cos(\varphi-2\theta)\right] \tag{3}$$

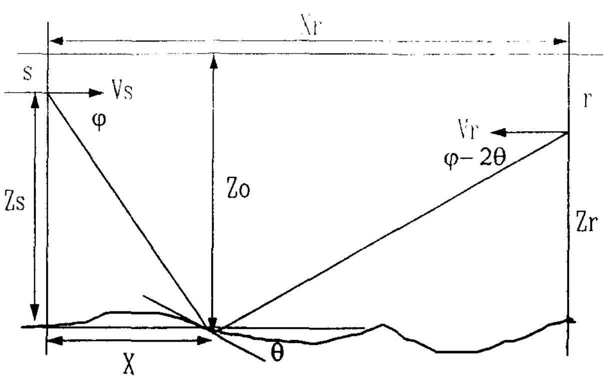

Fig.3 One-way bottom-bounce propagation geometry.
s : source, r : receiver, Vs : source speed, Vr : receiver speed, θ : bottom slope, φ : acoustic ray angle to horizontal.

For $\varphi = 60°$, $V_s = 10$ kts, $V_r = 0$ kts, $c = 1500$ m/s, and $f_o = 220$ Hz, the frequency shift results in 0.377 Hz.

On the other hand, the expected energy return by frequency spreading, $I(Q)$ can be expressed as[10]

$$I(Q) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} f(x, y)\, \delta\{Q-Q'(x, y)\}\, dy \tag{4}$$

where, $$Q' = \frac{V_{Sx}x + V_{Sy}y}{c[x^2+y^2+D^2]^{1/2}} - \frac{V_{Rx}(R-x) - V_{Ry}y}{c[(R-x)^2+y^2+D_1^2]^{1/2}}$$

$Q = (f-f_o)/f_o$, relative Doppler scaling,
$f(x, y)$ = expected received energy per unit area,
$\delta(Q-Q') = 1$ for $Q = Q'$, and 0 otherwise,
V_{Sx}, V_{Sy} = source velocities,
V_{Rx}, V_{Ry} = receiver velocities,
D = source-to-bottom depth,
D_1 = receiver-to-bottom depth,
R = horizontal range,
x = down-range, horizontal from source to receiver,
y = cross-range.

The frequency spreads, measured as the width 1/e value of each frequency spread curve, increase with rms slope and are symmetric with Q for source and receiver both moving in the same direction. However, the symmetry with respect to the Doppler peak is lost at larger rms slope for source or receiver being fixed. For an extreme case of bottom slope, rms 15°, $D = 4$ km, and $R = 22$ km, the ΔQ width is around 0,0046 for $V_{Rx} = V_{Sx} = 15$ m/s.[10] But it is even smaller for source being fixed. Even when we adopt the extreme width, the resulting frequency spreading may come to around 1.012 Hz at 220 Hz.

The source frequencies themselves have deviated about 1 Hz from the tonal frequencies during the measurement. After all, to identify one signal using an ANN, like the manner in spectrogram, a maximum bandwidth of 2.8 Hz is required. Since the frequency resolution of power spectrum is 0.1875 Hz, the number of frequency bins is 2.8/0.1875 or 15.

Ⅲ. Application of an ANN

An ANN consists of many interconnected processing elements (PEs) which model the behavior of neurons. In its simplest form, a PE is a threshold unit which accepts a number of inputs and produces an output only if the sum of the inputs is greater than an internal threshold. One of the most interesting aspects of an ANN is its learning capability. An ANN learns by adaptively changing the interconnection strengths between the PEs. In this way, a classifier can be built not by programming the network but by presenting it with a number of training examples and allowing it to build up the discriminant function automatically.

The updating of network weights with the backpropagation algorithm is analogous to the process used by the adaptive least mean square (LMS) filter.[11] The algorithm computes an estimate of the gradient of the mean square error with respect to the weights of the system, and this information is used to move the weights so that the error approaches a minimum. It is important to note that a neural network is a nonlinear device, typically with many local minima. The existence of local minima often makes it difficult for the network to reach its global minimum with unsupervised training. A derivation of the backpropagation algorithm is now briefly presented.

Given x_k as the input vector and w_k as the weight vector at time k, the output of the linear summation of a single PE is given by :

$$s_k = x_k^T w_k \tag{5}$$

and the output of the PE is :

$$y_k = \tanh\,(s_k) \tag{6}$$

The error here is defined as $e_k = d_k - y_k$ where d_k is the desired signal at time k.

The gradient of the mean square error with respect to the weights is given by :

$$\frac{\partial E(e_k^2)}{\partial w_k} \tag{7}$$

where E denotes the expected value. By eliminating the expected value, we obtain a stochastic estimate of the gradient :

$$\frac{\partial e_k^2}{\partial w_k}$$

resulting in the following weight update equation :

$$w_{k+1} = w_k - \mu \frac{\partial e_k^2}{\partial w_k} \tag{8}$$

where μ is the learning gain that controls the speed of convergence. Taking the partial derivative we obtain the following :

$$\begin{aligned} \frac{\partial e_k^2}{\partial w_k} &= 2\ e_k\ \frac{\partial e_k}{\partial w_k} \\ &= 2\ e_k\ \frac{\partial (d_k - y_k)}{\partial w_k} \\ &= -2\ e_k\ y'_k\ \frac{\partial s_k}{\partial w_k} \\ &= -2\ e_k\ y'_k\ x_k \end{aligned} \tag{9}$$

where $y'_k = \dfrac{dy_k}{ds_k}$.

Here, we have

$$\begin{aligned} y'_k &= \frac{d\ \tanh(s_k)}{d\ s_k} \\ &= 1 - \tanh^2(s_k) \\ &= 1 - y_k^2 \end{aligned} \tag{10}$$

and thus the weight update equation for a single PE becomes :

$$w_{k+1} = w_k + 2\,\mu\, e_k (1 - y_k^2)\ x_k \tag{11}$$

In a multilayer neural network, desired signals are typically available only for the PEs in the output layers. A general weight update equation for any PE is given by Widrow and Lehr[12] as :

$$w_{k+1} = w_k + 2\,\mu\,\delta_k\, x_k \tag{12}$$

where δ_k is called the square error derivative for the particular PE, and x_k is the input vector for that PE. For the output layer PEs, where desired signals are available, δ_k is obtained from Eq.(10) and is given by $\delta_k = e_k(1 - y_k^2)$. For a PE not in the output layer δ_k is given by[11] :

$$\delta_{mk}^l = e_{mk}^l \tanh'(s_{mk}^l) \tag{13}$$

where δ_{mk}^l is the square error derivative at time k for the m^{th} PE in the l^{th} layer, s_{mk}^l is the output of the same PE at time k, and e_{mk}^l is the backpropagated

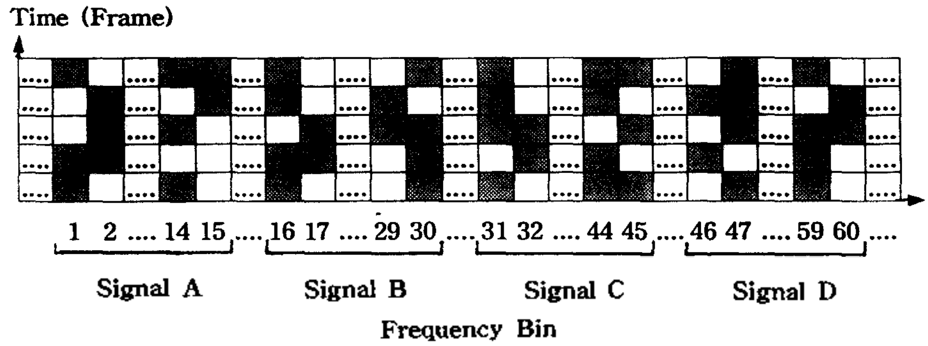

Fig.4 Input pattern for the ANN. A total of 60 frequency bins are used to identify the four different signals where each bin is 0.1875 Hz wide. The shaded bin denotes the case when the SNR is greater than 2 dB and thus has the value of 1.

error for that PE at time k. The backpropagated error is given by:

$$e^{l}_{m,k} = \sum_{i=1}^{a} \delta^{l+1}_{i,k}\; w_{(m,l)\to i,k} \tag{14}$$

where $w^{l+1}_{(m,l)\to i,k}$ is the weight that connects the m^{th} PE in the l^{th} layer to the i^{th} PE in the $(l+1)^{th}$ layer. Thus the backpropagated error for a PE that is not in the output layer is given by the sum of the squared error derivatives of the PEs in the following layer, each scaled by the PE weight that connects the PE being evaluated with the PE in the next layer.

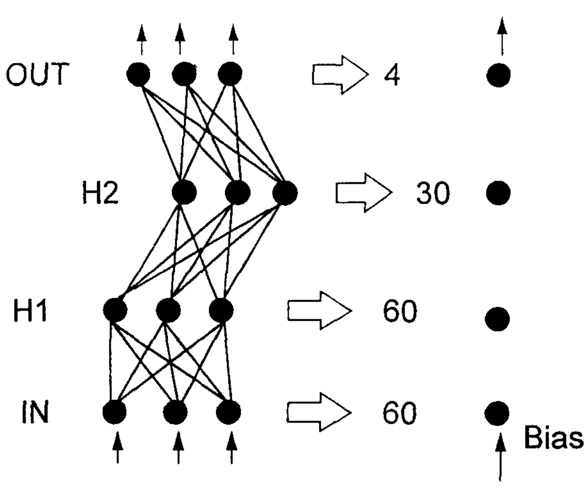

Fig.5 The ANN structure to identify the four signals. The input layer has 60 PEs, 15 PEs for each signal. The two hidden layers have 60 and 30 PEs, respectively. The output layer gives four results, each one having the value of 0-1.

In this paper, a total of 60 input PEs (15 PEs for each signal) are used as shown in Fig.4. In the figure, each frequency bin is 0.1875 Hz wide. The shaded bins denote the cases when the normalized power spectrum levels are greater than 2 dB and thus have the value of 1. Other bins have the value of 0. The ANN is trained to identify the specified signal if all of the 15 bins have the value of 1. The number of possible combinations for the training data set is $2^4 = 16$.

Figure 5 shows the architecture of the ANN employed in this study. The ANN consists of 4 layers: the input, the first hidden, the second hidden, and the output. The output layer has four PEs which generate results ranging from 0 to 1. A threshold, to decide whether the signal exists or not, is set to 0.9. A

Table 1. Recalling test results for the training data set after learning. In desired results, the specified signal is set to 1 or 0 if all 15 frequency bins have value of 1 or 0.

Desired Results				Test Results			
A	B	C	D	A	B	C	D
1	1	1	1	0.994196	0.996318	0.993251	0.994833
1	1	1	0	0.992100	0.994034	0.996587	0.009933
1	1	0	1	0.993457	0.996636	0.002019	0.994681
1	1	0	0	0.994753	0.994223	0.004292	0.008466
1	0	1	1	0.996177	0.003140	0.994079	0.993911
1	0	1	0	0.993066	0.007473	0.996523	0.007453
1	0	0	1	0.993993	0.004589	0.003616	0.993069
0	1	0	0	0.004736	0.997693	0.005597	0.003175
0	1	1	0	0.005487	0.996652	0.997323	0.004240
0	1	1	1	0.005354	0.994660	0.995776	0.996483
0	0	1	0	0.003833	0.002336	0.998484	0.003669
0	0	1	1	0.002677	0.003671	0.993401	0.993925
0	0	0	0	0.005385	0.005165	0.005620	0.007464
0	0	0	1	0.007324	0.004031	0.005109	0.994566
0	1	0	1	0.004346	0.994162	0.005039	0.993477
1	0	0	0	0.990229	0.008383	0.003581	0.005308

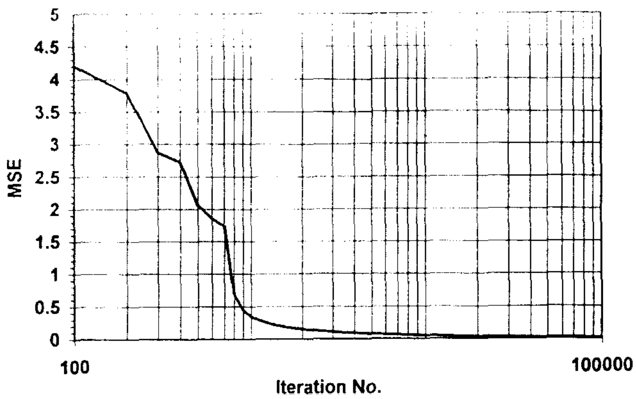

Fig.6 Mean square error with the iteration number in learning.

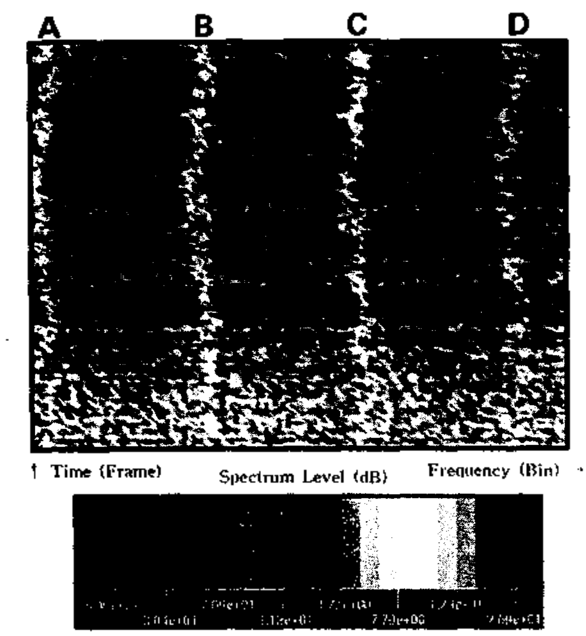

Fig.7 Spectrogram of the 7th receiving sensor. The bin numbers corresponding to the four signals A, B, C, and D are 8, 76, 142, and 206, respectively.

total of 155 PEs including the bias PE are interconnected in the ANN. The number of learning cycles is 100000 when the learning gain μ, is set to 0.45. In the learning, the ANN converges very fast and approaches mean square error 0 at the iteration number 100000 (Fig.6).

Table 1 shows the results for the training data set. In the desired results, the specified signal is set to 1 or 0 if all 15 frequency bins have value of 1 or 0. It can be shown that the ANN successfully identifies the four signals after learning.

IV. Results and Discussions

Figure 7 shows the spectrogram example of the 7th receiving sensor. The instantaneous power spectrum levels are displayed over 167 time steps and 220 frequency bins. The bins on which the four tones occur are 8th, 76th, 142nd, and 206th, respectively. In the figure, it can be shown that the frequency bin spreading is noticeable and it amounts to around 15 bins (or 2.8 Hz) as estimated previously. At the beginning of the spectrogram, it is not clear enough to identify the signals, but it becomes better as the time progresses. As a whole, the signals of A are the strongest and therefore the easiest to identify, and those of D are the weakest and therefore the most

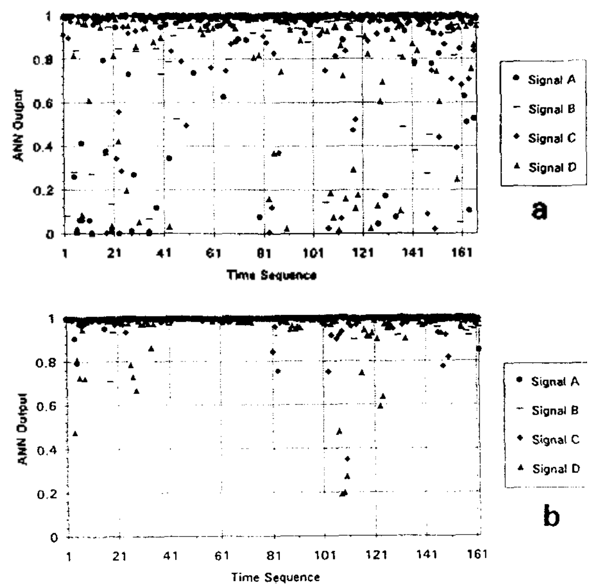

Fig.8 Classification results by the ANN for the four signals of the 7th receiving sensor : (a) instantaneous frame ; (b) five frames averaged.

Table 2. Classification results by the ANN for the four signals, A, B, C, and D.

Sensor	Instantaneous				Five Time-step Averaged			
No.	A	B	C	D	A	B	C	D
2	75.4	74.3	72.5	73.3	98.8	97.5	96.3	92.0
3	80.8	80.2	70.7	74.9	98.8	98.8	96.3	96.9
4	76.0	82.6	76.0	78.4	96.9	98.1	96.9	100.0
5	79.0	79.6	74.3	69.5	98.1	96.3	97.5	96.3
6	79.0	77.2	67.7	70.7	98.1	96.9	93.8	96.3
7	79.8	84.0	74.8	70.6	98.6	98.7	96.2	91.1
8	79.0	77.2	75.4	72.5	98.8	98.1	97.5	98.8
9	91.0	86.8	80.2	85.6	100.0	98.1	97.5	99.4
10	83.8	80.2	79.6	81.4	100.0	95.1	95.7	97.5
11	83.2	83.8	81.4	80.8	99.4	99.4	97.5	96.9
Avg.	80.7	80.6	75.3	75.8	98.8	97.7	97.5	96.2

difficult to identify.

The ANN-based results corresponding to the spectrogram are shown in Fig.8. The ANN outputs range from 0 to 1 over the 167 time steps. If a threshold to identify a signal is set to 0.9, a very large false alarm rate is expected for the instantaneous case (Fig.8). To lower the false alarm rate, an averaging over a number of time steps may be performed. Figure [illegible] gives the results when the five steps averaged information is applied to the ANN. Compared with the results for an instantaneous case, these are much more improved in a false alarm rate.

Table 2 summarizes the ANN results for 10 receiving sensors. The threshold for the decision is 0.9. For the 7th sensor, the results show that they are the best with signal B and the worst with signal D in both cases. This fact conincides with that in spectrogram. The minimum rate to identify the signal is 91.1% for the averaged case while it is 67.7% for the instantaneous case. The averaged case corresponds to the "eye integration" by humankind. At a glance, we can idenfity the four different tonals in the spectrogram.

In this paper, it is shown that the Doppler shift or frequency spreading is small enough to be able to estimate the suitable number of frequency bins on the spectrograms. Also, it is shown that an ANN-based classifier can successfully identify the four signals using only power spectrum information.

In real situations, however, there are many non-environmental factors such as propellers and generators that cause much tonal frequency variation with time. It is very common that the tonal frequency may be severely shifted according to the status of sound generating equipments. For example, the low-frequency spectra of cavitating surface ship propellers are usually dominated by tonal components at harmonics of the ship propellers which generally operate at from 60 to 350 rpm. The fundamental repetition frequencies of this type of sound vary from 4 to 18 Hz and the strongest components are generally the harmonics between 10 and 70 Hz.

In addition, the signals may have much noise and be intermittent. In this case, we can hardly classify the signals any more from the tonal frequency information. Other information, such as from broadband analysis and even from non-acoustic sensors, must be utilized to classify the signals. To make full use of data, the different sensor data processing systems must share information to a much greater extent. To achieve this, it is highly required the integration of peripheral knowledge of the current situation derived from a variety of sources at various levels (data fusion) as well as the utilization of knowledge from [illegible] experience (ANN-based classifier).

V. Summary

We attempt to apply an ANN based classifier in identifying low frequency acoustic signals in shallow water. The temporal frequency variation by the Doppler shift and spreading is estimated to be less than 2 Hz at 220 Hz, and this enables an ANN based classifier to be applicable using only tonal frequency information in the low frequency.

When the measured tonal signals in the frequency 200-250 Hz are applied to the ANN based classifier, the classifier can identify more than 67% of the signals for instantaneous case and more than 91% for averaged one over 5 time steps. This result is very comparable to that of the "eye integration" by humankind.

References

1. M. Bouvet and S. C. Schwartz, "Comparison of Adaptive and Robust Receivers for Signal Detection in [illegible]ent Underwater Noise," IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol.37[illegible], [illegible] 621-626.
2. T. B. Haley, "Applying Neural Networks to Automatic Active Sonar Classification," 1990 IEEE 10th International Conference on Pattern Recognition, Vol.2, 1990, 41-44.
3. R. P. Gorman and T. J. Sejnowski, "Learned Classification of Sonar Targets Using a Massively Parallel Network," IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol.36(7), 1988, 1135-1140.
4. W. P. Webster, "Artificial Neural Networks and Their Application to Weapons," Naval Engineers Journal, 1991, 46-59.
5. C. P. Sheppard and C. R. Gent, "A Neural Network Based Sonar Classification System," Underwater Systems Design, Nov./Dec., 1991, 28-31.
6. Y. V. Zakharov and V. P. Kodanev, "Doppler Scattering Adapted Reception in a Hydroacoustic Communication Channel," Acoustical Physics, Vol.41(2), 1995, 219-223.
7. J. Wolcin, "On the Statistical Properties of Noise Background Equation Schemes," NUSC Tech. Memo. No.

781159, Naval Underwater Systems Center, New London, CT, 1978.

8. W. A. Struzinski and E.D. Lowe, "A performance comparison of four noise background normalization schemes proposed for signal detection systems," J. Acoust. Soc. Am., Vol.76(6), 1984, 1738-1742.
9. W. A. Struzinski and E.D. Lowe, "The effect of improper normalization on the performance of an automated energy detector," J. Acoust. Soc. Am., Vol.78 (3), 1985, 936-941.
10. X. Zabal, M. H. Brill, and J. L. Collins, "Frequency and angle spreads of acoustic signals reflecting from a fixed rough boundary," J. Acoust. Soc. Am., Vol.79 (3), 1986, 673-680.
11. M. M. Nelson and W. T. Illingworth, "How Do Neural Networks Learn?," in A Practical Guide to Neural Nets, Addison-Wesley Publishing Co., 1990, 36-56.
12. B. Widrwo and M. A. Lehr, "30 years of adaptive neural networks: Perceptron, madaline, and backpropagation," IEEE Proceedings, Sept. 1990, 1415-1442.

▲Young-Nam Na
Senior Research Scientist,
Agency for Defense Development
(Vol. 14, No. 1E, 1995)

▲Taebo Shim
Chief Research Scientist,
Agency for Defense Development
(Vol. 14, No. 1E, 1995)

▲Jeong-Woo Han
Research Scientist,
Agency for Defense Development
(Vol. 13, No. 3, 1994)

▲Chun-Duck Kim
Professor, Department of Electrical Engineering
National Fisheries University of Pusan
(Vol. 12, No. 3, 1993)