SONAR SIGNAL PROCESSING

Sonar is an example of remote sensing. Although sonar systems are used for fish-finding, acoustic imaging through turbid water for remote underwater operations, and exploration of geophysics, they are most commonly identified with detecting ships and submarines.

In principle, sonar and radar are similar because both use wave energy to detect distant targets. Yet, in practical implementation, they are vastly different. Most notable is the difference in media: sonar relies on acoustical waves, whereas radar relies on electromagnetic waves. Furthermore, the sonar medium is much more variable: channel effects are more severe, propagation rates are 200,000 times slower (1500 m/s rather than 3×10^8 m/s), frequencies are much lower (10 kHz to 100 kHz rather than 0.1 GHz to 100 GHz), and the signal bandwidths as a percentage of the carrier frequency, in general, are much larger than those in radar. There is also more noise and reverberation. Although the speeds of ships and submarines are considerably lower than those of aircraft and missiles, the much greater difference in propagation speed yields greater Mach numbers (v/c) for sonar (typically 10^{-3}) than for radar (typically 10^{-6}). As discussed later, the higher

Mach numbers achieved in sonar imply that echoes from moving targets have to be processed differently.

The differences in the parameter values imply that radar and sonar systems collect data about targets at different rates and with different resolutions. For example, several seconds or minutes can pass between each sonar transmission. In radar, hundreds or thousands of pulses are transmitted, received, and integrated within one second.

A BRIEF HISTORY OF SONAR SIGNAL PROCESSING

Sonar and sonar signal processing possess a history rich in development and implementation. Unlike radar, which has a number of civilian uses, sonar is primarily used for military purposes. Thus, most research and development of sonar technology has been sponsored by the world's navies.

Hundreds of years ago, it was recognized that sound travels in water. Leonardo da Vinci observed that sound from distant ships could be heard by placing one end of a tube in the water and the other to the ear. This system offered no gain and no directivity. Sound had to be sufficiently strong to overcome the noise induced by the motion of the boat and nearby breaking waves.

Prior to World War I, little was done beyond da Vinci's work. Any kind of signal processing would require the development of electronic technology, something that did not occur at any significant level until the twentieth century.

During World War I, most sonars were "passive" acoustic systems. One system of this era resembled a stethoscope and was composed of two air-filled rubber bulbs mounted on the end of a tube connected to earpieces. An operator listened for sounds that indicated a ship or submarine. Because it was a binaural system, the operator could estimate the bearing to the detected vessels. Later versions of this system had a similar in-water configuration, but with several bulbs attached to each earpiece. Such an arrangement offered directivity, so it had to be manually steered to detect a vessel and estimate its bearing. This is perhaps the earliest example of beam forming, a topic covered later.

Later in World War I, electric underwater transducers called hydrophones were developed using electromechanical materials that deform with the application of an electric or magnetic field (piezoelectrics and magnetostrictives). The use of these materials, which allowed the efficient coupling of electric power with underwater acoustic power, was crucial to the development of sonar because it made possible more general arrangements of sensors (arrays). Consequently, towed, horizontal line arrays were developed that offered more gain and directivity than previous passive systems. A single horizontal line array cannot be used to distinguish signals arriving from both sides of the array but approaching from the same angle. Therefore, a pair of line arrays was towed, because it was possible to resolve the "left-right ambiguity" of the target bearing. This system was the forerunner of the modern military towed-array sonar system.

After World War I, reliable, high-power electronic amplification allowed development of "active" sonars. In this type of sonar, an acoustic pulse is transmitted that generates echoes which are detected aurally, electronically, or visually (cathode ray tube). Active sonar systems were employed by ships and submarines during World War II. Such systems did not employ much signal processing because the equipment required to implement complex algorithms did not exist or was too large to install on vessels. Only simple vacuum tube electronic equipment was available. It was bulky and consumed much electrical power. Reliable, high-speed, silicon-based electronics was decades away.

Today's sonar systems employ large towed or hull-mounted arrays composed of many hydrophones. The signals from these arrays are processed by small, high-speed computers. Thus, it is possible to implement many computationally intensive, multiple input signal processing algorithms to detect, classify, and track ships and underwater targets.

The operating frequency for a modern sonar system depends on its application, which determines the required operating range and resolution. The higher the frequency, the more attenuation a signal experiences per unit distance of propagation. As shown later, for a fixed array size, the ability to resolve and locate a target increases as the frequency and signal bandwidth increase.

Modern military sonar systems generally fall into one of three categories: weapons (torpedoes), tactical systems, and surveillance systems. These three categories roughly correspond to three operating frequency ranges: high-frequency (above 10 kHz), midfrequency (1 kHz to 10 kHz), and low-frequency (below 1 kHz). High frequencies attenuate greatly per unit distance of propagation, but as explained later, offer the highest angular resolution of a target for a fixed array size. Active and passive torpedoes operate in this frequency range, because they use two-dimensional arrays that must fit within the torpedo housing and still achieve sufficient angular resolution over distances that are not too great. Active mine-hunting sonars also operate at high frequency, because high-frequency arrays yield high-resolution images of the terrain and mines that are used for identification or classification. Passive tactical sonar systems, which typically operate in the midfrequency range, are used by surface ships or submarines to avoid being successfully targeted by an attacker. They must be small and not impede maneuvering. Active tactical sonar systems are also used for searching moderately wide areas defined by the stand-off distance of particular offensive weapons, such as torpedoes or cruise missiles. Active and passive surveillance sonar systems are often large and possibly covert (therefore passive sonar) and are used to detect and track targets over a wide area. These sonars use low frequencies that propagate over great distances underwater.

SOUND IN THE OCEAN

The oceanic environment is broadly categorized as either deep water or shallow water (1). In deep water, the water channel is sufficiently deep that propagating sound is well approximated as rays. Deep water supports sound propagation with a depth-dependent sound speed, $c(d)$ (d denotes depth), which differs in regions of the ocean and times of the day and year. The channel response is approximated as a finite sum of weighted time-delayed impulse responses, each of which corresponds to the arrival of a nondispersive ray. There are several computer programs for estimating this channel response (2). In shallow water, the boundaries of the water channel (the surface, water-sediment interface, and the sediment-basement interface) are separated by a few wavelengths, and propagating sound is best approximated as a sum of modes (traveling standing waves). In general, the sound speed is depth-dependent, and the modes are dispersive or frequency-dependent. There are also computer programs to simulate this behavior (2).

The propagation effects just described imply that sound traveling through the ocean exhibits time-spreading (multipath distortion) at long ranges. Sound also spreads in angle because of horizontal inhomogeneities, and spreads in frequency because of time variations in acoustic parameters, such as the depth-dependent sound speed and surface motion. When all three forms of spreading occur, it is termed FAT (frequency, angle, time) spreading.

Any sound deliberately transmitted in the ocean, upon reception, is contaminated by noise and echoes from the ocean boundaries and inhomogeneities called reverberation. First, consider the simplest passive sonar system configuration: a nondirectional radiating point-target (source) with a nondirectional point-hydrophone (receiver) in a time-invariant, homogeneous (space-invariant), infinite medium. Here, a transmitted signal $s(t)$ travels directly from the source to the receiver. At the receiver, the pressure field is given by

$$p_s(t) = \frac{s(t - R_{sr}/c)}{R_{sr}} \tag{1}$$

where R_{sr} is the range of the receiver with respect to the source. The signal $p_s(t)$ is corrupted by additive noise $n(t)$, which is white, Gaussian, and isotropic in the most restricted case. This noise originates from sources in the ocean (radiated noise from ships and breaking waves on the ocean surface) and from noise introduced by the system electronics. Because noise is the primary source of interference in passive sonar systems, they are often called "noise-limited."

Next, consider the simplest active sonar system configuration: a nondirectional point-projector (source), a nondirectional point-hydrophone (receiver), and a point-target in a time-invariant, homogeneous (space-invariant), infinite medium. When the source signal is scattered from the point-target, the pressure field at the receiver is given by

$$p_t(t) = \left[\frac{a}{R_{st}}\right]\left[\frac{s(t - R_{st}/c - R_{tr}/c)}{R_{tr}}\right] \tag{2}$$

where a is proportional to the fraction of sound scattered by the point-target, R_{st} is the range of the point-target with respect to the source, and R_{tr} is the range of the receiver with respect to the point-target. Thus, the effect of propagation is a time delay and a decay in amplitude. The source signal is also scattered from the surface, bottom, and volume inhomogeneities (fish) to produce reverberation. At the receiver, the reverberation pressure field is given by

$$p_{rev}(t) = \sum_i \left[\frac{b(i)}{R_{sb}(i)}\right]\left\{\frac{s[t - R_{sb}(i)]/c - R_{br}/c)}{R_{br}(i)}\right\} \tag{3}$$

where $b(i)$ is proportional to fraction of sound scattered by the ith scatterer, $R_{sb}(i)$ is the range of the ith scatterer with respect to the source, and $R_{br}(i)$ is the range of the receiver with respect to the ith scatterer. In a realistic ocean environment, $b(i)$ may be proportional to a surface area for surface reverberation (surface roughness), or it may be proportional to a volume (volume reverberation). Because reverberation is the primary source of interference in active sonar systems, they are often called "reverberation-limited."

The purpose of sonar signal processing is to enhance the detectability of a particular type of signal from noise, reverberation, or any source of deliberate interference. Generally speaking, a sonar operator's ability to detect and track a target improves if a signal processing system increases the signal-to-reverberation ratio (SRR), the signal-to-noise ratio (SNR), or the signal-to-interference (SIR), defined as the ratios of the expected received signal power to the expected powers of the reverberation, noise, or reverberation and noise plus any deliberate interference. Accordingly, SRR, SNR, and SIR are measures of system performance.

It has become customary to express the SNR and SRR in terms of the sonar equations, which are written as a sum of logarithms of the power or energy:

$$\begin{aligned} \mathrm{EL} &= \mathrm{SL} - \mathrm{TL} + \mathrm{TS} \\ \mathrm{SNR} &= \mathrm{EL} - \mathrm{NL} \\ \mathrm{SRR} &= \mathrm{EL} - \mathrm{RL} \end{aligned} \tag{4}$$

where EL is the echo level, TL is the transmission loss from the projector to hydrophone, NL is the ambient noise level, and RL is the reverberation level. These and other terms commonly used in variations of the sonar equations that account for other factors affecting signal excess are given in Table 1. The accepted units for the sonar equations are μPa for pressure and meters for length. A real ocean environment is time-varying and inhomogeneous, and the noise field is anisotropic. Therefore, expressing the performance of a sonar system with a sonar equation is only approximate because there are convolutional, rather than multiplicative, relationships between the source array, receiver array, target and medium in time and space.

The transmission loss, noise level, and reverberation level depend on how acoustic energy spreads (propagates) away from a projector. Two types of spreading are commonly considered: spherical spreading (deep water, short range, all frequencies), and cylindrical spreading (shallow water, medium and low frequencies, long range). Consider transmission loss. If spherical spreading occurs, then TL = 20 log $r + \alpha_L r$, where r is the range from the projector to hydrophone, and α_L is called the absorption loss coefficient. If cylindrical spreading occurs, then the change in range at long range is approxi-

Table 1. Sonar Equation Terms

Term	Name	Description
AN	Ambient noise	Power of ambient noise at hydrophone
DI	Directivity index	Measure of projector or hydrophone directivity
DT	Detection threshold	Signal power required for detection
EL	Echo level	Echo power
SE	Signal excess	Excess of signal over detection threshold
SL	Source level	Power level of projector
TL	Transmission loss	Power drop due to spreading and absorption
TS	Target strength	Measure of target reflectivity

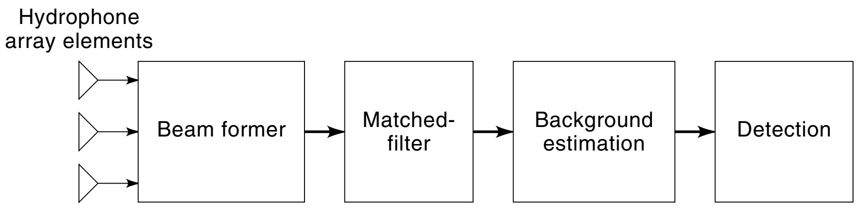

Figure 1. System architecture of a passive or active sonar receiver.

mately given by TL = 10 log $r + \alpha_L r$. The reverberation and ambient noise levels are also affected by the propagation. Consider the case of volume reverberation at medium and high frequencies where scattering occurs at every point in the ocean. If spherical spreading occurs, then RL changes in range by -20 log r, where r is now the range from a colocated projector and hydrophone to a point in space. If cylindrical spreading occurs, then the change in range at long range is approximately given by -30 log r. Unlike volume reverberation, surface reverberation is independent of the type of spreading and changes in range by -30 log r. With a colocated projector and hydrophone, the time (range delay) is related to range by

$$t = \frac{2r}{c} \tag{5}$$

Thus, formulas for reverberation yield the time-dependence of the expected power of the reverberation component of a received signal.

Although the sonar equation is a simple tool generally for "back-of-the-envelope" calculations, it is useful for quantifying the improvement gained through signal processing. A more detailed description of sonar equation terms is given in Ref. 3.

Conceptually, improvement of SNR or SRR is achieved in two separate ways because signals can be described as functions of both time (or frequency) and space (position). Filtering a received signal in the time domain or frequency domain exploits the coherence of signal and eliminates noise or reverberation that does not occupy the intervals of time or frequency occupied by the signal. Filtering in the spatial domain allows directing sound toward or received from a particular direction and is accomplished by combining the signals from projectors or hydrophones distributed in the water. Filtering is a principal function of sonar signal processing described in detail in the following section.

FUNCTIONS OF SONAR SIGNAL PROCESSING

Sonar signal processing systems vary in their complexity and capability, depending on their application and the number of signals they process. Yet, almost all systems must do beam forming, matched filtering, detection, and background estimation. These functions are interrelated. In reception, they are performed sequentially as shown in Fig. 1. In transmission, only beam forming is done.

Beam Forming

Many sonar systems, particularly those for military use, do not employ a single projector or hydrophone. Many sensors are used and arranged in a regular pattern. Such an arrangement, called an "array," allows projecting acoustic energy to or receiving energy from a given direction. Thus, the sonar operator, or autonomous weapon, can interrogate a particular volume of the ocean and avoid a large echo from an interfering target (a sea mount, the surface, a fish school) or reduce the interference from an acoustic noise source (distant shipping or a noise-source countermeasure).

Beam forming is the combining of projector or hydrophone signals to direct or receive acoustic energy to or from a given direction in the ocean. The degree of precision with which this is accomplished depends on the spatial distribution and number of projectors or hydrophones and the operating frequency. Consider a monochromatic pressure plane wave of the form

$$p(t, x, y) = e^{j(\omega t - k_x x - k_y y)} \tag{6}$$

where $k = \sqrt{k_x^2 + k_y^2} = \omega/c$ is the called the wave number, ω is the radian frequency, and c is the propagative speed. Also consider a horizontal linear array of uniformly spaced hydrophones as shown in Fig. 2. If we use the signal at the first hydrophone as a reference signal and realize that monochromatic signals are presented by each hydrophone, then the signal from each hydrophone is given by

$$r_i(t) = e^{j\omega(t-(d/c)\cos\theta)} \quad \text{for } i = 1, \ldots, n \tag{7}$$

where θ is the plane-wave arrival angle. Suppose that the hydrophone signals are added together in the form of the weighted sum

$$y(t, \theta) = e^{j\omega t} \sum_{i=1}^{n} w_i e^{-jdk\cos\theta} \tag{8}$$

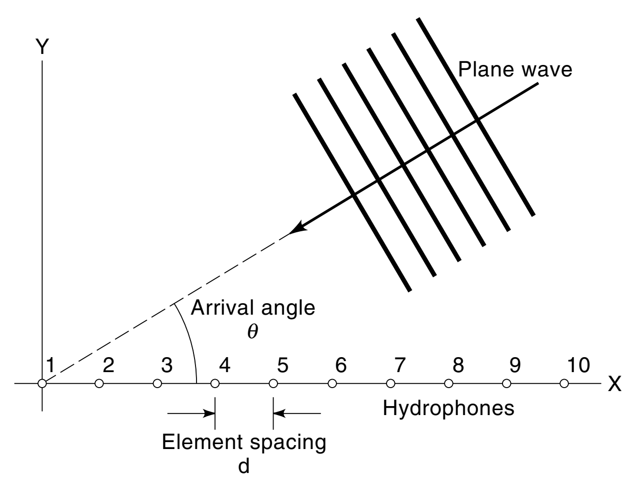

Figure 2. A horizontal line array with uniformly spaced hydrophones.

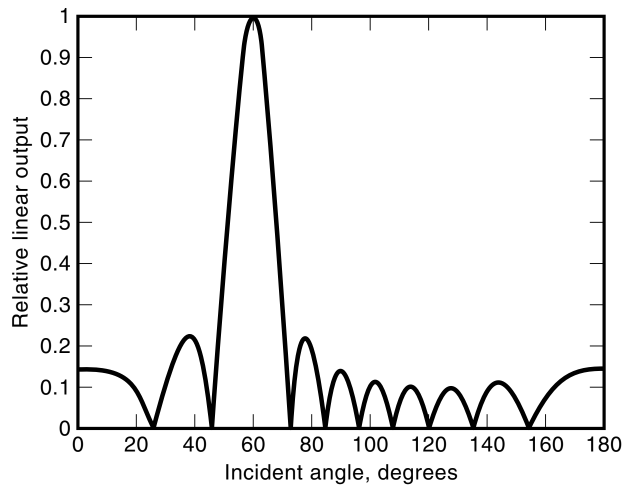

Figure 3. A beam pattern for a horizontal linear array with ten hydrophones uniformly spaced by one-half wavelength. The beam pattern is steered 60° from boresight.

where each weight w_i is a complex number. The sum is also a monochromatic signal, and if we constrain the weights with magnitudes no greater than 1, then the amplitude of $y(t, \theta)$ is maximized if we choose the weights as

$$w_i = e^{jdk\cos\theta} \quad \text{for } i = 1, \ldots, n \tag{9}$$

With this choice of weights, plane waves arriving from other directions do not produce an output signal with as large an amplitude as the signal arriving from angle (azimuth) θ. Thus, the choice of weights "steers" the array in the direction of the incoming plane wave.

Figure 3 displays the magnitude of the response of the array previously described to plane waves arriving at all angles between 0° and 180° of azimuth. The plot in the figure is called a "beam pattern" with several features common to all beam patterns. First there is a "main lobe" which points in the direction the beam is steered. The width of the main lobe reflects how tightly the acoustic energy is directed or received. The remainder of the beam pattern is composed of "sidelobes" and "nulls." It is desirable to have a beam pattern with a main lobe that is as narrow as possible and sidelobes as small as possible. The width of the main lobe and the maximum level of the sidelobes are changed by adjusting the magnitude of the weights (called "shading") or by increasing the length of the array. For an array of fixed length and fixed number of projectors or hydrophones, shading reduces the sidelobe level but at the expense of a wider main lobe. Lengthening the array with more elements reduces both the main lobe width and sidelobe level.

The linear array of uniformly spaced sensors is the simplest beam former to analyze. However, beam forming is done for any array configuration. In general, for n projectors or hydrophones arranged in a three-dimensional pattern, the beam former output is given by

$$y(t, \theta, \phi) = e^{j\omega t} \sum_{i=1}^{n} w_i e^{-j\omega\tau_1(\theta,\xi)} \tag{10}$$

where $\tau_i(\theta)$ is the time delay between the first and ith sensor for a plane wave arriving at an azimuth of θ and elevation ξ.

Generally speaking, the beam pattern is a function of the array size in any one dimension and also of the operational frequency. Therefore, what really counts is the size of the array array in wavelengths: the greater the number of wavelengths across an array, the narrower the beam width. Radar systems typically operate at frequencies in the GHz region where the wavelengths are measured in centimeters or fractions of centimeters. The wavelengths for sonar systems are generally much larger. Hence, radar systems are generally capable of higher angular resolution for a fixed array size.

There are several common array configurations used in military sonar systems, some of which are shown in Fig. 4. Tactical sonar systems, which typically operate at frequencies from 1 kHz to 10 kHz, often employ towed-line arrays hundreds of meters long. They also use spherical arrays mounted inside an acoustically transparent, water-filled housing installed on the hull of a ship or submarine. Figure 5 shows a spherical array mounted on the bow of a cruiser. Surveillance sonars, which typically operate at frequencies below 1 kHz, use large line or planar arrays mounted on the sea bottom or suspended in the water. These low-frequency arrays can also be hundreds or thousands of meters long. Torpedo sonars operate at frequencies above 10 kHz and employ planar arrays mounted on the torpedo's flat nose or on the side of the torpedo body.

Although beam forming is done with analog circuitry, digital processing is more convenient and, hence, the principal form of implementation today. Analog circuitry is bulky, comparatively inflexible, and allows for only a small number of fixed beam patterns. In contrast, digital processing allows for almost any value of beam-forming weight, which can be de-

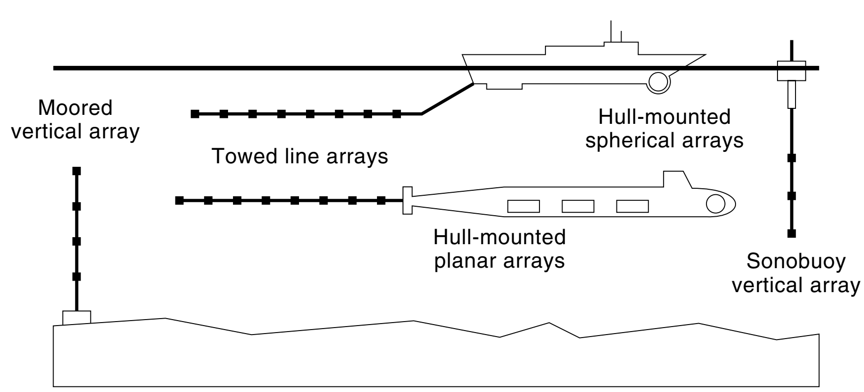

Figure 4. Common sonar array configurations on ships, submarines, and deployed systems.

Figure 5. A spherical, midfrequency sonar array on the bow of a cruiser in drydock.

rived adaptively in situ. For reception, beam forming is done on a computer using samples of the hydrophone outputs. On transmission, the signals for the projectors, each with its own unique time delay and amplitude, are generated by a computer, sampled, delivered to a digital-to-analog converter, and amplified to drive a projector.

As stated earlier, beam forming allows an operator to reduce the receiving sensitivity of a sonar to sources of noise or reverberation. In principle, this is accomplished by placing nulls in the beam pattern coincident with the angular position of these sources. In the case of a linear array with uniformly spaced hydrophones, the beam pattern in Eq. (8) is a polynomial in $e^{-jkd\cos\theta}$. Therefore, placement of the nulls is equivalent to determining the roots of a polynomial. If a null is required at some $\theta = \theta_0$, then the polynomial in Eq. (8) must have a zero at $e^{-jkd\cos\theta_0}$. Because the polynomial is of degree n, it can have as many as n unique zeros, and so as many n nulls may be steered against interference sources. Placement of the zeros is accomplished by selecting appropriate values for the weights $w_1, \ldots, w_n$.

The previous formulation assumed that direction of the interference sources is known, which allows direct calculation of the weights. In practice, calculation of the weights is done indirectly. One method for determining the weights begins with finding an estimate of the hydrophone data correlative matrix given by $\boldsymbol{R} = E\{\boldsymbol{r}\boldsymbol{r}^H\}$, where $\boldsymbol{r}^T = \{r_1(t), \ldots, r_n(t)\}$ is a vector of monochromatic signals. The weights are determined by solving the minimization problem:

$$min_{\mathrm{w}}\boldsymbol{w}^H\boldsymbol{R}\boldsymbol{w} \quad \text{subject to} \quad \boldsymbol{w}^H\boldsymbol{\eta}(\theta_{\mathrm{d}}) = 1 \tag{11}$$

where $\boldsymbol{\eta}^T(\theta_{\mathrm{d}}) = \{1, e^{-jkd\cos\theta_d}, \ldots, e^{-j(n-1)kd\cos\theta_d}\}$ and θ_{d} is the desired direction of maximum signal response, typically a "look direction" where a target exists. The solution is given by

$$\boldsymbol{w} = \frac{\boldsymbol{R}^{-1}\boldsymbol{\eta}(\theta_{\mathrm{d}})}{\boldsymbol{\eta}^H(\theta_{\mathrm{d}})\boldsymbol{R}^{-1}\boldsymbol{\eta}(\theta_{\mathrm{d}})} \tag{12}$$

This method works well if the echo or radiated signal in the hydrophone data from the target is dominated by noise and reverberation. This is usually true if noise-generating countermeasures are dropped by an evading target.

The beam-forming problem for reducing noise and reverberation becomes more complicated if the sonar platform (ship, submarine, torpedo) or the sources of interference are moving. In this case, the angular positions of the sources move with respect to the sonar platform, and beam forming becomes a time-varying problem. This dictates modifying any algorithm for beam steering and null placement to use only timely data to derive an estimate of the correlative matrix $\boldsymbol{R}$. One method, called the recursive least square (RLS) estimation algorithm, does this by exponentially weighting the contribution of each measured time series used to estimate $\boldsymbol{R}$, weighing heavily the most recently measured time series (4).

To this point, beam forming has been presented in terms of receiving (directing) acoustic energy from (to) a remote point in space. Certain assumptions were made in deriving the results presented thus far. In particular, it was assumed that the array and point of interest are far enough apart to assume that an acoustic field is approximated as a plane wave. A more general view of receiving acoustic energy, called matched-field processing, recognizes that the acoustic field received is a complex function of the hydrophone and projector locations and the way sound propagates in the ocean.

Suppose that a single source (projector) is placed in the ocean and the output signals are available from hydrophones placed nearby in some general configuration. If the oceanic environment and the positions of the projector and hydrophones were exactly known, then the output signals from the hydrophones could be exactly predicted. Of course, in practice, only the hydrophone positions and output signals are measured, whereas the projector location and environment are usually not well known. It is possible, however, to assume values for the projector location and environmental parameters, calculate the resulting hydrophone output signals based on those assumptions, and compare them with the measured outputs. If the difference is small, then the assumed project location and environmental parameters are close to the real values. This is the fundamental principle of matched-field processing (5).

To illustrate matched-field processing, consider a shallow-water oceanic environment usually defined as any area where the depth is 300 m or less. In such an environment, it is known that the pressure field as a function of depth d due to a monochromatic omnidirectional source (projector) with amplitude A at range r_s and depth d_s is expressed by

$$
\begin{aligned}
p(d) &= \sum_{n=1}^{N} a_n \psi_n(d) \\
a_n &= \frac{A}{\sqrt{k_n r}} \psi_n(d_s) e^{-jk_n r_s}
\end{aligned} \tag{13}
$$

where k_n is the horizontal wave number, and $\psi_1(d), \ldots, \psi_N(d)$ are orthogonal functions called modes. The exact forms of the modes depend on the velocity of sound as a function of depth $c(d)$. If $c(d)$ and d_s are known, then the hydrophone outputs can be predicted exactly or at least to the limit of the accuracy of the mode propagative model used. In practice, only the outputs from hydrophones are available. Thus, if pressure measurements are available from a vertical array of M hydrophones, a measurement vector is formed with pressure measurements from different depths, written as $\boldsymbol{p}^T = \{p(d_1), \ldots, p(d_M)\}$. An hypothesized pressure field vector is given by $\hat{\boldsymbol{p}}^T = \{p(\hat{d}_1), \ldots, p(\hat{d}_M)\}$, where

$$
\begin{aligned}
\hat{p}(d) &= \sum_{n=1}^{N} \hat{a}_n \psi_n(d) \\
\hat{a}_n &= \frac{B}{\sqrt{k_n \hat{r}}} \psi_n(\hat{d}_s) e^{-jk_n \hat{r}}
\end{aligned} \tag{14}
$$

where $\hat{r}_s$ is the hypothesized source range, $\hat{d}_s$ is the hypothesized source depth, and B is chosen so that $\hat{\boldsymbol{p}}^H \hat{\boldsymbol{p}} = 1$. Assuming that the modes are known, the matched-field processor output is given by the inner product of the measured field and normalized hypothesized field:

$$
\mathscr{P}(\hat{r}_s, \hat{d}_s) = |\hat{\boldsymbol{p}}^H \boldsymbol{p}|^2 \tag{15}
$$

If M is sufficiently large that it may be assumed that

$$
\sum_k \phi_i(d_k) \phi_j^*(d) \approx 0 \quad \text{for } i \neq j \tag{16}
$$

it follows that

$$
\mathscr{P}(\hat{r}_s, \hat{d}_s) = \left| \sum_n \hat{a}_n^* a_n \right|^2 \tag{17}
$$

Maximizing this sum with respect to $\hat{r}_s$ and $\hat{d}_s$ yields the best estimate of the source range and depth. Because it is assumed that the modes are known, the procedure described here is one of determining the correct weighted sum of modes that match the measured pressure field. Hence, it is referred to a matched-mode processing.

Matched-field processing is computationally intensive because it requires an exhaustive search over a multivariable acoustic parametric space. Significant computational benefits result from matched-mode processing because of the assumed structure of the pressure field (modes). However, the modal representation of an acoustic field is not appropriate in deep-water or range-dependent, shallow-water environments. Matched-field processing has been extended to include the estimation of more sonar system parameters, such as noise level and ocean acoustic properties, to achieve greater robustness.

Detection and Matched Filtering

Detection is the process of deciding whether a particular portion of the beam-former output contains a target echo. In its most simple form, it is merely deciding if there is enough energy to declare that a target is present. This is typically accomplished by comparing the value of the beam-former output at a particular time with a threshold γ whose value is some multiple of the estimated background level. The decision is made using the recorded echo from a single transmission (single-ping detection) or several echoes (multiple-ping or sequential detection). The same detection algorithms used in radar systems are also employed in sonar systems.

There is considerable processing of the raw hydrophone data before detection. First beam forming is done to steer the sensitivity of the hydrophone array in several directions, allowing the operator to observe the entire environment. The beam-former outputs are then bandpass filtered to contain only the frequency band of interest and to eliminate out-of-band noise and reverberation. This is followed by windowing which divides the beam-former output into several overlapping pieces. Finally, each portion of the windowed output is Fourier transformed and displayed. At this point, detection is done.

The output of a passive sonar signal processing system is displayed to an operator in several different ways. Typically the square magnitude of the Fourier transforms of the windowed data are displayed as either a color contour (planar) plot, or waterfall plot. For a fixed beam, successive transforms are displayed, thus providing a two-dimensional display with frequency as one axis and time as the other. Alternatively, a fixed time is chosen (a single data window), and a two-dimensional display of frequency versus beam angle is displayed.

Passive systems identify the presence of target sources emitting signals of fixed frequency. Such targets appear as fixed lines, or “tonals,” in the frequency versus time display previously described. An operator looks for such lines in the display, which, over time, drift in frequency because the target moves (motion-induced Doppler). In the frequency versus beam display, the target appears as a peak, which shifts from beam to beam because of its motion. Both displays show the signatures of short, transient signals for the target as well. These signals appear as short lines or frequency sweeps. In any case, the tonals and transients are observed by an operator, who can thus track the target.

In its simplest form, detection in active sonar systems is essentially deciding between two mutually exclusive events: (1) only noise and reverberation are the active echo (hypothesis H_0) or (2) a target echo, noise, and reverberation are in the active echo (hypothesis H_1). Detection in active sonar systems lends itself to automation, as in torpedoes, but can still involve an operator, as with many tactical and surveillance systems.

After beam forming and filtering, an active echo $r(t)$ is commonly processed by a matched-filter receiver:

$$m(\alpha_1, \ldots, \alpha_n) = \left| \int r(t) g^*(t|\alpha_1, \ldots, \alpha_n)\, dt \right|^2 \tag{18}$$

where $g(t|\alpha_1, \ldots, \alpha_n)$ is the unity energy filter function, which models the expected form of the target echo subject to the parameters $\alpha_1, \ldots, \alpha_n$, such as speed and range. In the case of a stationary point-target, the target echo is nothing more than a time-delayed version of the transmitted signal $f(t)$. Thus,

$$g(t|\tau) = f(t - \tau) \tag{19}$$

More generally, if a point-target is moving, then the transmitted pulse compresses or expands on reflection. Thus,

$$g(t|\tau, s) = f[s(t - \tau)] \tag{20}$$

where $0 < s$ is the Doppler variable given by

$$s = \frac{c \pm v}{c \mp v} \approx 1 \pm 2v/c \tag{21}$$

where v is the range rate or velocity of the target along the line of sight. More often, the Doppler effect is modeled as a simple spectral shift of the signal. In this case, if f_c is the signal carrier frequency, then

$$g(t|\tau, \phi) = f(t - \tau) \exp(j2\pi\phi t) \tag{22}$$

where

$$\phi = (s - 1) f_c = \Delta s\, f_c \tag{23}$$

called the "carrier frequency Doppler shift." The matched-filter function in Eq. (20) is called the wideband, point-target reflection model, and the function in Eq. (22) is called the narrowband, point-target reflection model. As discussed at the end of this article, the wideband model is used when the signal bandwidth is a significant fraction of the signal carrier frequency. Without loss of generality, the narrowband model is used throughout the remaining discussion on detection.

The point-target models described above do not model the echoes from real-world targets. However, they are used in practice for several reasons. First, they are simple. Second, no general model for a target echo may be available, especially if the type of target is unknown. Finally, if the target is composed of many highlights, the matched filter produces a large response to each of the target highlights.

If we consider the case of searching for a moving target in a fixed direction, then we must perform matched filtering over a range of time delays and Dopplers. This yields a two-dimensional surface called a "range Doppler map," which contains peaks that are responses to one or more targets. The remainder of the surface is the response of the matched filter to noise and reverberation. Detection is accomplished by comparing the matched-filter output threshold, which is some fixed value higher than the average matched-filter response, with the noise and reverberation. If the value of the surface exceeds the threshold, then a target is declared, and the bin is tagged as a target response. Otherwise, the bin is tagged as containing no target energy. The result is a simplified, range Doppler map that contains the target responses and a few noise and reverberation responses that happened to exceed the detection threshold (false alarms).

The value of the detection threshold depends on the statistical nature of the target and clutter. Consider examining a range Doppler map at the point (τ_0, ϕ_0) where a target response exists. Let the value of the matched filter at this point $[m(\tau_0, \phi_0)]$ be described by the random variable z. If the probability density functions of the two detection hypotheses, $f_Z(z|H_0)$ and $f_Z(z|H_1)$, are known, then the probability of detection is given by

$$P_\mathrm{d} = \int_\gamma^\infty f_Z(z|H_1)\, dz \tag{24}$$

where z is the matched-filter output. The probability of a false alarm is given by

$$P_\mathrm{fa} = \int_\gamma^\infty f_Z(z|H_0)\, dz \tag{25}$$

The density functions depend on the statistical nature of the noise, reverberation, and target. The simplest model is a nonfluctuating point-target in white Gaussian noise. In this case, if the return contains a target echo, the probability density function of the matched-filter output is given by

$$f_Z(z|H_1) = \frac{1}{2\sigma^2} \exp\left[-\frac{(z + A^2)}{2\sigma^2}\right] I_0\left(\frac{A\sqrt{z}}{\sigma^2}\right) \quad \text{for } z \geq 0 \tag{26}$$

where

$$\sigma^2 = E\{m(\tau, \phi)\}_{\text{noise and reverb}} \tag{27}$$

and A is the amplitude of the return signal. This is known as the Rician density function, which is used to model the matched-filter response to a stationary point-target. If the return does not contain an echo, but only noise and reverberation, then the probability density function of the matched-filter output is given by

$$f_Z(z|H_1) = \frac{1}{\sigma^2} \exp\left(-\frac{z}{\sigma^2}\right) \quad \text{for } z \geq 0 \tag{28}$$

Equation (26) must be integrated numerically, but the values have been tabulated and are available in almost any text on detection theory. The false alarm probability is determined in closed form given by

$$P_\mathrm{fa} = \exp\left(-\frac{\gamma}{\sigma^2}\right) \tag{29}$$

If the point-target fluctuates, and its amplitude is modeled as a complex Gaussian random variable, then the probability density function of the matched-filter output is given by

$$f_Z(z|H_1) = \frac{1}{\sigma_{\mathrm{T}}^2 + \sigma^2} \exp\left(-\frac{z}{\sigma_{\mathrm{T}}^2 + \sigma^2}\right) \quad \text{for } z \geq 0 \tag{30}$$

where

$$\sigma_{\mathrm{T}}^2 = E\{m(\tau, \phi)\}_{\text{target}} \tag{31}$$

In this case, the probability of detection is given by

$$P_{\mathrm{d}} = P_{\mathrm{fa}}^{1/(1+SNR)} \tag{32}$$

where the false alarm probability is given by Eq. (29), and the signal-to-noise ratio is given by

$$\mathrm{SNR} = \frac{E\{m(\tau, \phi)\}_{\text{target}}}{E\{m(\tau, \phi)\}_{\text{noise and reverb}}} \tag{33}$$

The previous equations reveal the dependence of the detection process on the detection threshold γ. There are a number of ways to choose a detection threshold, but the most common approach is to choose the false alarm rate first and then determine (and simply live with) the resulting probability of detection. This approach, known as the Neyman–Person detection method, is popular because setting the false alarm rate at an acceptable level avoids task-loading an operator with tracking too many false targets.

The probabilistic models described above are commonly used in detection analysis for sonar systems. They are used for a "first cut" analysis if no other information about the target or environment is available. However, sonar systems are routinely deployed in environments where the statistical fluctuations of the noise and reverberation cannot be modeled by a complex Gaussian process. The most common attribute of an environment that deviates from the simple models described above is that the tails of the probability density functions $f_Z(z|H_0)$ and $f_Z(z|H_1)$ contain more area than would be present if Gaussian statistics were valid. In such cases, using a threshold derived for a fixed false alarm rate given Gaussian noise and reverberation yields a true false alarm rate higher than predicted.

In instances where non-Gaussian noise and reverberation prevails, extensive measurements must be performed to gather enough data to estimate its probability density function and, if possible, the probability density function of matched-filter response to the target. It is possible to estimate the power of the noise and reverberation and to use the estimate to derived a detection threshold. This is known as background estimation.

Background Estimation

Background estimation is the process of estimating the power and frequency distribution of the noise or reverberation in the beam-former output during reception. It is performed by examining a portion of the beam-former output time signal that is assumed to contain no target echo. It typically uses the discrete values of the beam-former output as inputs to a statistical estimation algorithm. The estimated background level

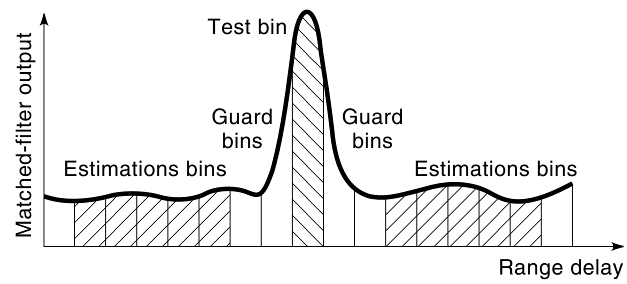

Figure 6. The test bin, guard bins, and estimation bins used for estimating the background level for constant false alarm rate detection.

is then used to determine the detection threshold for a given false alarm probability.

Consider Fig. 6 which shows a target response in a matched-filter output. The output is divided into bins, which reflect the digitization of the analog data received from the beam former. It is assumed that the test bin contains the matched-filter target response and that the values in the estimation bins are used to estimate the expected value of the background. The guard bins are not used directly, but provide a "buffer space" between the test bin and the estimation bins, so that no target energy "spills" into the estimation bins and biases the estimate.

The simplest way to estimate the background level is to average the values of all of the matched-filter values in the estimation bins. The estimated background level is given by

$$\hat{\sigma}^2 = \frac{1}{M} \sum_i z_i, \tag{34}$$

where z_i is a sample of the matched-filter output in the ith bin and the summation is taken over M estimation bins. Assuming that the noise and reverberation is Gaussian, the probability of false alarm is given by Eq. (29). Therefore, substituting $\hat{\sigma}^2$ for σ^2 in this equation and solving for γ yields the detection threshold used in the detection bin:

$$\gamma = -\hat{\sigma}^2 \ln P_{\mathrm{fa}} \tag{35}$$

The arrangement of estimation bins, guard bins, and test bin is shifted to the right a fixed number of bins, usually commensurate with the resolution of the matched filter. The estimation and detection process is then repeated.

The detection process described is called bin-average or cell-average constant false alarm rate (CFAR) processing because the probability of a false alarm has a fixed value. It works well as long as all of the estimation bins contain only noise and reverberation. If other target returns occupy the estimation cells, then the power of the background estimate is high (biased), and the detection threshold is too high. Thus, if the test bin contains a target response, it might not exceed the threshold, and the target is not detected. More robust estimation algorithms have been developed to circumvent this and other nonuniformities in the background. For example, a trimmed-mean estimate is performed where the highest value acquired from the estimation cells is discarded before averaging. Alternatively, the mode of the values in the estimation cells is used as the background estimate. This is known as order-statistic CFAR processing.

SCATTERING AND SIGNAL MODELING

Some knowledge of the scattering properties of the environment and target are essential for evaluating the performance of a sonar system. Because the matched-filter is the principal processing algorithm in the detection state of a sonar signal processing system, it is essential to understand how the matched-filter responds to a return containing echoes from the target and the environment.

Signal Scattering and the Ambiguity Function

Consider the case of narrowband scattering where it is sufficient to model a Doppler shift by a spectral shift. Under the assumption of wide-sense stationary scattering, it can be shown that the expected value of the matched filter to a scatterer is given by

$$E\{m(\tau,\phi)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S(\hat{\tau},\hat{\phi})|\chi(\hat{\tau}-\tau,\hat{\phi}-\phi)|^2 d\hat{\tau}\,d\hat{\phi} \tag{36}$$

where $S(\tau, \phi)$ is the scattering function of the scatter,

$$\chi(\tau,\phi) = \int_{-\infty}^{\infty} x(t)x^*(t-\tau)e^{-j2\pi\phi t}\,dt \tag{37}$$

is the narrowband uncertainty function, and $|\chi(\tau, \phi)|^2$ is called the ambiguity function (9). The scattering function is estimated from measured data or derived if the geometry of the scatters is simple. The integral in Eq. (36) is a linear convolution between the signal ambiguity function and the target scattering function.

The scattering function of several simple scatterers is known. A simple point-target at range τ_0 and a range rate inducing a Doppler frequency shift of ϕ_0 has a scattering function that is a two-dimensional delta (Dirac) function:

$$S(\tau,\phi) = \delta(\tau-\tau_0,\phi-\phi_0) \tag{38}$$

The scattering function of a line-target with the same range and Doppler and length L is given by

$$S(\tau,\phi) = G_{2L/c}(\tau-\tau_0)\delta(\phi-\phi_0) \tag{39}$$

where

$$G_W(t) = \begin{cases} 1 & \text{if } 0 < t < W \\ 0 & \text{otherwise} \end{cases} \tag{40}$$

and is called the "rectangular-pulse function." The scattering function of simple volume reverberation, as seen by high-frequency sonar systems, straddles the $\phi = 0$ line as shown in Fig. 7. The overall amplitude of the scattering function dies off according to the way energy spreads in the environment. For example, if acoustic energy propagates by spherical spreading, then the amplitude decays in range delay as $1/\tau^2$. The profile of the scattering function along the ϕ axis for a fixed range is usually modeled by a simple unimodal function (such as a Gaussian pulse), but for simple analysis it is modeled as a rectangular-pulse function.

Scattering function analysis lends itself to quick and simple analysis of system performance if simple models for the target (point or line) and the environment are used. It is used to estimate the relative expected values of the responses of the matched filter to target and reverberation, which are expressed as signal-to-noise ratios.

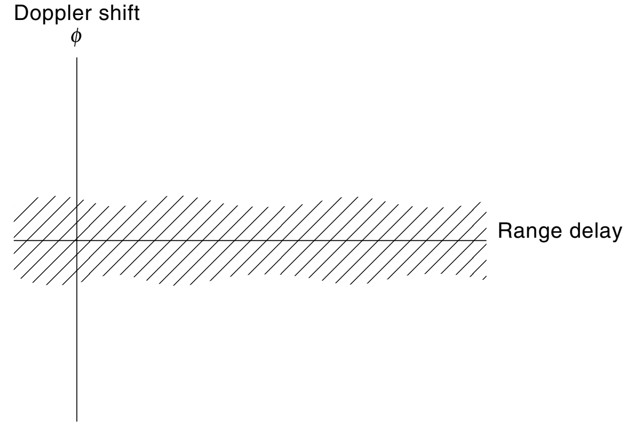

Figure 7. A scattering function for volume reverberation.

Equation (36) also reveals that sonar system performance depends on the shape of the ambiguity function, which is controlled by modulating the sonar signal. Thus, the ambiguity function is another "parameter" that is adjusted by the system designer. A great deal of technical literature has been written about the signal design problem, which couches the problem in terms of the volume distribution of the ambiguity function. A few examples demonstrate this important point.

Consider a simple continuous-wave (CW) signal, which is nothing more than a gated tone given by

$$x(t) = \frac{1}{\sqrt{T}}G_T(t) \tag{41}$$

The narrowband ambiguity function for this signal is given by

$$|\chi(\tau,\phi)|^2 = G_{2T}(t-T)\left|\left(1-\frac{|\tau|}{T}\right)\frac{\sin[\pi(T-|\tau|)\phi]}{\pi(T-|\tau|)\phi}\right|^2 \tag{42}$$

This ambiguity function is shown in Fig. 8. It is a simple "lump" whose width in range delay is T and width in Doppler

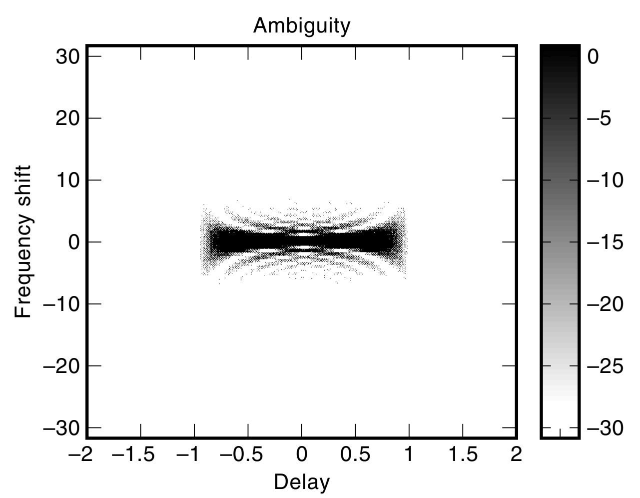

Figure 8. A narrowband ambiguity function of a continuous wave (CW) signal.

is approximately $1/T$. These values determine the resolution of the signal. Point-targets separated in range and Doppler by more than these values are separate responses in a range Doppler map. Now consider the case of a linear frequency modulated (LFM) signal given by

$$x(t) = \frac{1}{\sqrt{T}} G_T(t) \exp\left(\frac{j\pi B t^2}{T}\right) \tag{43}$$

The narrowband ambiguity function for this signal is given by

$$|\chi(\tau,\phi)|^2 = G_{2T}(t-T)\left|\left(1-\frac{|\tau|}{T}\right)\frac{\sin[\pi(T-|\tau|)(\phi-B\tau/T)]}{\pi(T-|\tau|)(\phi-B\tau/T)}\right|^2 \tag{44}$$

This ambiguity function is shown in Fig. 9. The resolution of this signal is approximately $1/B$ in range and approximately $1/T$ in Doppler. Although these values are quite high and demonstrate the "pulse compression" property of the LFM, the signal cannot discriminate between point-targets separated in range and Doppler cells aligned with the time-frequency slope of the signal. Thus, the signal is used to over-resolve (image) stationary targets of large range. It also offers some processing gain (SRR improvement due to matched filtering) over a CW against point-targets in volume reverberation.

A number of other signals have been derived to control the volume distribution of the ambiguity function to make a sonar system more effective in detecting or imaging certain classes of targets. Of particular note are the time-frequency, hop-coded signals. Such signals are based on Costas arrays, one of which is displayed in Fig. 10 (8). If such a pattern is shifted vertically and horizontally, it intersects the original pattern at no more than one other "pulse." If a series of CW pulses is concatenated in time, each with a different frequency allocated in the same relative fashion as the pulses in the Costas array, then the narrowband ambiguity functions looks much like a "thumbtack." An example of such an ambiguity function is shown in Fig. 11. Hop-code signals are used to image high Doppler targets composed of several point highlights.

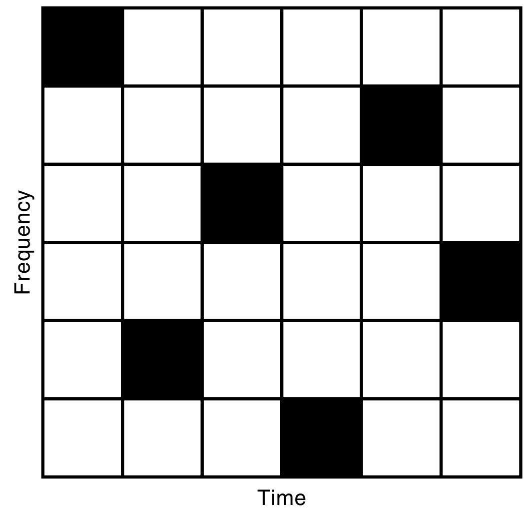

Figure 10. A Costas array for designing hop-code signals.

Wideband Versus Narrowband Processing

Thus far, it has been assumed that a Doppler shift could be modeled by a spectral shift, implying that the narrowband, point-target reflection model in Eq. (22) is valid. Use of such a model in matched-filtering is called narrowband processing. When the relative motion between the sonar projector/hydrophone and a target is sufficiently large, the effects of time dilation must be considered. If this is true, then the wideband, point-target reflection model in Eq. (20) is valid. Use of such a model in matched-filtering is called wideband processing.

Suppose that a signal of time length T and bandwidth B is transmitted from a stationary projector/hydrophone and is reflected by a target with and approaching line-of-sight velocity v. The received signal has length sT, where s is given by Eq. (21). Thus, the difference in signal duration is $(s-1)T$. The signal range resolution is approximately $1/W$. Therefore,

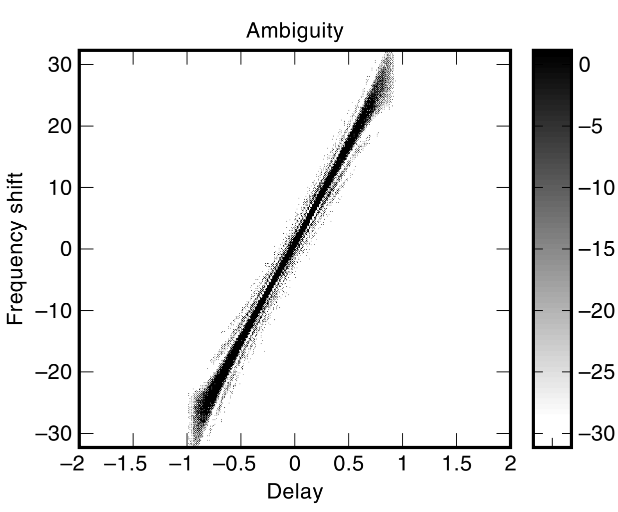

Figure 9. A narrowband ambiguity function of a linear, frequency-modulated (LFM) signal with $BT = 30$.

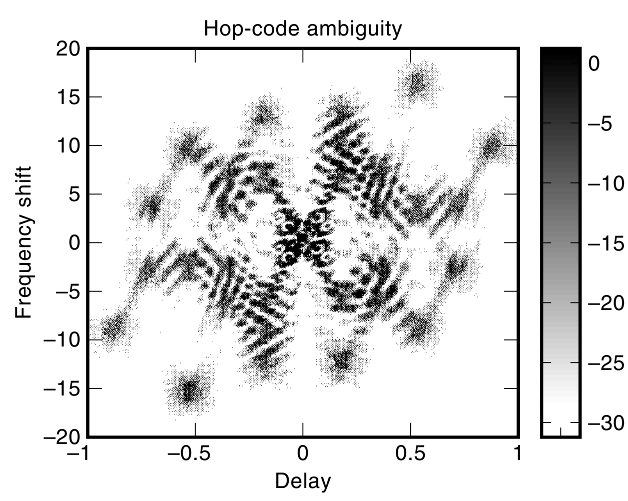

Figure 11. The narrowband ambiguity function of a hop-code signal based on the Costas array in Fig. 10.

if the change in length is equal to this narrowband signal resolution or larger, then the matched-filter output is large in two or more adjacent bins. In other words, the energy is split between the bins. This implies at least a 3 dB drop in the matched-filter response from that attained if narrowband processing is sufficient. Thus, the criterion for wideband processing is given by

$$(s-1)T > 1/W \tag{45}$$

Using the formula for the carrier frequency Doppler shift ϕ in Eq. (23), the criterion is given as

$$f_c/W > \frac{1}{T\phi} \tag{46}$$

Wideband processing implies that the scattering function and the signal ambiguity function must be defined differently. Accordingly, the expected value of the wideband matched-filter output is given by

$$E\{m(\tau,s)\} = \int_{\hat{\tau}=-\infty}^{\infty}\int_{\hat{s}=0}^{\infty} S(\tau,s)|\chi[s/\hat{s},\hat{s}(\tau-\hat{\tau})]|^2\,d\hat{\tau}\,d\hat{s}, \tag{47}$$

where $S(\tau, s)$ is the wideband scattering function,

$$\chi(\tau,s) = \int_{-\infty}^{\infty} x(t)x^*[s(t-\tau)]\,dt \tag{48}$$

is the wideband uncertainty function, and $|\chi(\tau, s)|^2$ is called the wideband ambiguity function. The integral in Eq. (47) is not a linear convolution as defined in the narrowband case. The distinction is not always important for calculating back-of-the-envelope performance predictions. For example, the narrowband assumption is used when calculating processing gains for signals used for detecting slowly moving (low-Doppler) targets.

CONCLUSION

Readers seeking a more detailed general overview of sonar system design and deployment or an understanding of the environmental parameters that affect sonar system performance should consult references such as Urick (3). Readers seeking a knowledge of the basic theoretical material for sonar signal processing should consult references such as Burdic (6). Furthermore, the large volume of radar literature on filtering, detection, and beam forming also serves as foundational material for sonar signal processing.

Sonar signal processing algorithmic development is faced with inherent difficulties. First, the oceanic environment is hostile and highly variable: sound does not always travel in straight lines, important environmental parameters are often unknown in situ, and the knowledge of surface and bottom scattering mechanisms is incomplete and highly site-dependent. This makes it difficult to develop reliable detection and classification systems for general use. Second, practical systems are plagued by high sensor cost, difficulty in array deployment and recovery, power limitations, and communication constraints. Consequently, good target localization and reliable in situ environmental parametric estimation are difficult to achieve because there are often an insufficient number of projectors and hydrophones arranged in an inadequate array configuration.

Despite the difficulties cited, new developments in materials and electronics will allow the development of low-cost sensors, compact deployment systems, and high-speed signal multiplexing and processing electronics. This, in turn, will create new demands for sonar signal processing algorithmic development and present opportunities for improving sonar system performance.

BIBLIOGRAPHY

1. I. Tostoy and C. S. Clay, *Ocean Acoustics,* Washington, DC: American Institute of Physics, 1987.
2. P. C. Etter, *Underwater Acoustic Modeling,* New York: Elsevier Applied Science, 1991.
3. R. J. Urick, *Principles of Underwater Sound,* New York: McGraw–Hill, 1983.
4. M. L. Honig and D. G. Messerschmitt, *Adaptive Filters: Structures, Algorithms, and Applications,* Boston, MA: Kluwer, 1984.
5. A. Tolstoy, *Matched Field Processing for Underwater Acoustics,* Singapore: World Scientific, 1993.
6. W. Burdic, *Underwater Acoustic System Analysis,* New York: Prentice–Hall, 1984.
7. B. D. Van Veen and K. M. Buckley, Beamforming: A versatile approach to spatial filtering, *IEEE ASSP Mag.,* **5** (2): 4–24, 1988.
8. S. W. Golomb and H. Taylor, Construction and properties of Costas arrays, *Proc. IEEE,* **72**: 1143–1163, 1984.
9. L. J. Ziomek, *Underwater Acoustics: A Linear Systems Theory Approach,* New York: Academic Press, 1985.

DAVID M. DRUMHELLER
Naval Research Laboratory
CHARLES F. GAUMOND
Naval Research Laboratory
BRIAN T. O'CONNOR
Naval Research Laboratory