UNCLASSIFIED

AD NUMBER

AD468869

NEW LIMITATION CHANGE

TO

Approved for public release, distribution unlimited

FROM

Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative/Operational Use; 17 MAY 1965. Other requests shall be referred to Air Warfare Research Department, NADC, Johnsville, PA.

AUTHORITY

NADC ltr, 15 Feb 1966

THIS PAGE IS UNCLASSIFIED

SECURITY MARKING

The classified or limited status of this report applies to each page, unless otherwise marked.
Separate page printouts MUST be marked accordingly.

THIS DOCUMENT CONTAINS INFORMATION AFFECTING THE NATIONAL DEFENSE OF THE UNITED STATES WITHIN THE MEANING OF THE ESPIONAGE LAWS, TITLE 18, U.S.C., SECTIONS 793 AND 794. THE TRANSMISSION OR THE REVELATION OF ITS CONTENTS IN ANY MANNER TO AN UNAUTHORIZED PERSON IS PROHIBITED BY LAW.

NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

468869

U.S. Naval Air Development Center

Johnsville, Pennsylvania

AIR WARFARE RESEARCH DEPARTMENT

REPORT NO. NADC-WR-6509 17 MAY 1965

AD 468 869

LECTURE NOTES ON UNDERWATER ACOUSTICS

DDC
NOV 8 1965
TISIA B



Limitations concerning the distribution of this report and revelation of its contents appear on the inside of this cover.

U. S. NAVAL AIR DEVELOPMENT CENTER
JOHNSVILLE, PA. 18974

AIR WARFARE RESEARCH DEPARTMENT

REPORT NO. NADC-WR-6509 17 MAY 1965

LECTURE NOTES ON UNDERWATER ACOUSTICS

Prepared by: C. L. Bartberger
C. L. Bartberger

Approved by: W. H. Raber
W. H. Raber
Director

FOREWORD

These notes have been prepared for a course entitled, "The Technology of Underwater Sound," given at the U. S. Naval Air Development Center by Mr. Charles L. Bartberger. As originally written they were intended as a simplified presentation of some of the material contained in J. W. Horton's book, "Fundamentals of Sonar," published by the U. S. Naval Institute. They have since been revised and a considerable amount of new material has been added.

It will be evident, upon examining the contents of these notes, that the approach is largely theoretical. The intent has been primarily to present basic principles and for this reason the portions dealing with hardware and practical applications are rather sketchy.

A number of topics, such as explosive echo-ranging and passive listening, have been omitted. It is hoped to include these topics and to improve the organization and the consistency of notation in a later revision.

TECHNOLOGY OF UNDERWATER SOUND
REVISED NOTES

TABLE OF CONTENTS

	Page
INTRODUCTION	1
PHYSICAL PROPERTIES OF ACOUSTIC WAVES IN WATER	5
A. One-Dimensional Plane Waves	5
1. Introduction	5
2. Differential Equation of Motion	6
3. Solution of the Differential Equation	10
4. Sinusoidal Waves; Acoustic Pressure	12
5. Particle Displacement and Particle Velocity	15
6. Pressure Gradient	17
7. Intensity	17
8. Note on Units of Measurement	21
9. Average Values	22
10. Acoustic Impedance	23
11. Standing Waves	24
B. Three-Dimensional Waves	28
1. The Differential Equation	28
2. Three-Dimensional Plane Waves	34
3. Two-Dimensional Plane Waves	35
4. Spherical Waves	36
5. Cylindrical Waves	40
6. Unit Area Acoustic Impedance and Admittance	42
7. Acoustic Impedance and Mechanical Impedance	45
8. Noise Pressure; Equivalent Sine Wave Intensity	46
C. Waves with a Continuous Distribution of Frequencies	48
1. Spectra	48
2. Power per Unit Band; Intensity per Unit Band	49
3. Note on Units	50
RELATIVE MAGNITUDES AND TRANSMISSION LOSS	51
1. Introduction	51
2. Relative Magnitudes	53
3. Transmission Loss - The Decibel	53

		Page
4.	Transmission Level	56
5.	Sound Pressure Level	58
6.	Note Regarding the Pressure Reference	60
7.	Change of Reference	61
8.	Spectrum Level; Band Level	62
9.	Band Level Correction	65
10.	Source Level; Index Level	68
THE PROPAGATION OF UNDERWATER SOUND		70
A.	Propagation Loss	70
1.	Definition	70
2.	General Discussion	71
3.	Spreading Loss	73
4.	Attenuation Loss	74
5.	Combined Loss	75
6.	The Measurement of Propagation Loss	76
7.	Evaluation of the Propagation Loss Components	76
8.	Values of the Attenuation Coefficient	77
9.	Fluctuations in Sound Level	79
10.	Propagation Anomalies	79
B.	Refraction	80
1.	The Ray Solution of the Wave Equation	80
2.	Plane Waves and the Ray Concept	80
3.	The Eikonal Equation	82
4.	Criterion for the Validity of the Eikonal Equation	85
5.	Application to Typical Ocean Conditions - Vertical Velocity Gradient	88
6.	Snell's Law	92
7.	The Speed of Sound in the Ocean; Velocity Gradients	101
8.	Ray Paths in a Constant-Gradient Medium	107
9.	Multiple Constant-Gradient Layers; Boundary Effects	114
10.	Spreading Loss	123
11.	The Naval Air Development Center Ray Tracing Program	131
12.	Limitations of the Constant-Gradient Approximation	134
13.	Method of Curvilinear Gradients	141
14.	Differential Equation Method	144
C.	Reflection	149
1.	Introduction	149
2.	Reflected and Transmitted Pressure and Intensity	149
3.	Effects of Boundary Conditions	156

		Page
4.	Image Sources	163
5.	The Lloyd Mirror Effect	164
6.	Normal Mode Propagation	168
D.	Long-Range Propagation Paths in the Ocean	177
1.	Introduction	177
2.	Bottom-Bounce Path	177
3.	SOFAR Path	180
4.	Convergence Zone Path	183
5.	Reliable Acoustic Path (RAP)	187
6.	Refracted Surface-Reflected (RSR) Path	191
TRANSDUCERS		193
A.	The Nature of Sonar Transducers	193
1.	Introduction	193
2.	Mechanical Impedance	195
3.	Radiation Impedance	199
4.	Piezoelectric Transducers	203
5.	Magnetostrictive Transducers	233
B.	Directional Characteristics	238
1.	Introduction	238
2.	Directivity Patterns for Specific Transducer Types	242
3.	Shading of Transducers	263
4.	The Near Field	267
5.	Directivity Factor; Directivity Index	272
6.	Noise Rejection by Directional Hydrophones and Arrays	283
7.	Reverberation Factor; Reverberation Index	286
8.	Capture Area	288
C.	Thermal Noise	290
1.	Introduction	290
2.	Thermal Equivalent Water Noise	293
3.	Thermal Equivalent Transducer Noise	295
D.	Calibration of Transducers	296
1.	Transducer Sensitivity	297
2.	Pressure Due to a Simple Source; Extended Source	298
3.	The Reciprocity Theorem	300
4.	Direct Calculation of J for Quartz Crystal Transducer	302
5.	Absolute Calibration Using Reciprocity Theorem	306

Page

6. Requirements for Free-Field Calibration 308
7. Special Calibration Techniques 309
8. Near-Field Calibration Techniques 313

E. Transducer Systems 319

1. Directivity Patterns of Simple Multi-Element Systems 319
2. Electrically-Steered Arrays 321
3. Bearing Deviation Indicators 335

INDICATORS AND RECORDERS; DETECTION OF SIGNALS 338

A. Types of Indicators and Recorders 338

1. Audible Presentations 338
2. Visual Presentations 338

B. Signal Detection 340

1. General Remarks Concerning the Observation of Acoustic Signals 340
2. Signal-to-Noise Ratio; Signal Differential; Observed Differential 343
3. Recognition Differential 343
4. Statistical Detection Theory 349

ECHO-RANGING 369

A. Echo Level 369

1. Target Strength 370
2. Effect of Deviation Loss 373
3. Note on Positive Values of Target Strength 375
4. Bi-Static Echo-Ranging 375

B. Noise-Limited Ranges 375

1. Noise 375
2. The Sonar Equation 381
3. Figure of Merit 382
4. Echo-Excess; Detection Probability 383
5. Noise-Limited Ranges 386

C. The Doppler Shift 386

1. Target Doppler 387
2. Self-Doppler 389
3. Combined Doppler 390
4. Minimum Required Bandwidth 390

		Page
D.	Reverberation-Limited Ranges	391
	1. Volume Reverberation	391
	2. Surface and Bottom Reverberation	396
	3. Recognition Differential	401
	4. Reverberation-Limited Ranges	401
E.	Special Types of Pulses; Some Signal Processing Considerations	403
	1. CW Pulse	403
	2. Linear FM Pulse	408
	3. Doppler-Invariant Pulse	411
	4. Pseudo-Random Noise Pulse	412

TECHNOLOGY OF UNDERWATER SOUND

INTRODUCTION

1. Uses of Underwater Sound

a. To detect, track, and localize submerged objects (chiefly submarines and mines).

b. Communication. Sound propagation through the water may be used for voice and other communication between submarines, ships, and (with the aid of sonobuoys) aircraft.

c. Navigation. Navigation by underwater sound is becoming increasingly important as submergence time increases.

d. Surveying. In the field of oceanography, sound is being used to survey the oceans to permit underwater navigation.

In this course we shall be concerned chiefly with the detection of submarines.

2. Methods of Detecting Submarines

a. Electromagnetic

(1) Visual. Up to the present time the eye has been the best detector of submarines. Limited to surfaced or snorkeling submarines or to wakes behind recently submerged submarines. Optical path in water is very short. Limited to conditions of good visibility.

(2) Radar. Subject to same limitations as visual detection, except that radar is applicable under conditions of poor visibility (at night, etc.) and provides greater detection ranges under favorable conditions. Back scattering from sea surface is a serious limitation on snorkel detection in high sea states. Radiation emitted by radar gives warning to submarines at distances beyond radar detection range.

d. Active Sonar (Echo Ranging). Acoustic energy supplied by the sonar in the form of a pulse, such as a ping or an explosion, Pulse is transmitted to target and is returned in the form of an echo. Two-way transmission. Signal depends upon the strength of the pulse, the back-scattering properties of the target, and the two-way propagation loss. Interference consists not only of sea noise and self-noise of listening platform, but also on amount of radiated energy which is scattered back by the ocean environment - reverberation.

(1) A Monostatic echo ranging system is one in which both the transmitting and receiving elements are located at approximately the same place.

(2) A Bistatic echo ranging system is one in which the transmitting and receiving elements are separated from one another,

e. Principal Advantages and Disadvantages

Passive Sonar

Advantages	Disadvantages
Does not alert target	Depends upon cooperation of target
One-way propagation loss	Poor range information
No reverberation	
Ability to classify target	

Active Sonar

Advantages	Disadvantages
Good vs. a quiet target	Alerts target
Gives accurate range information	Range limited by two-way propagation loss
Speed information from doppler shift	Signal strength depends on target aspect
	Reverberation

f. Factors Affecting Sonar Performance

(1) Frequency of Acoustic Waves. Frequency affects attenuation of waves in the ocean medium. High-frequency waves are heavily absorbed--high attenuation--short ranges. Frequency also affects size of equipment. Low frequency (required for long ranges) has long wave length; necessitates large equipment--bulky, heavy, expensive. Frequency also affects resolving power (ability to distinguish two objects located close together).

(2) Velocity of Sound. Velocity of sound is about 5000 ft/sec. Very slow compared with electromagnetic waves. Imposes limitation on information rate.

(3) Refraction of Sound Waves. Variation of velocity of sound with depth causes sound rays to travel in curved paths. Creates "shadow zones." Severely limits detection ranges under unfavorable conditions.

(4) Location of Sonar Elements. Limitations imposed by refraction can be alleviated to some extent by locating the transmitting and receiving elements at the optimum depth. The problem is complicated, however, by the fact that the target, unless committed by tactical considerations, is free to choose his best depth to avoid detection.

(5) The Sonar Platform. In some cases the self-noise generated by the sonar platform itself is a critical item limiting sonar performance. An example is the noise generated by a destroyer or by a helicopter with dipping sonar.

TECHNOLOGY OF UNDERWATER SOUND
(REVISED NOTES)

Physical Properties of Acoustic Waves in Water

A. One-Dimensional Plane Waves

1. Introduction.

We shall begin our discussion of acoustic waves in water by considering the simple case of one-dimensional plane waves in an ideal infinite, homogeneous ocean. Imagine a hypothetical infinite rigid plane which by some means is made to oscillate back and forth in this medium with sinusoidal motion, the direction of motion being at right angles to the plane. If the water were totally incompressible, the whole ocean would move simultaneously with the driving plane. This would correspond to an infinite speed of propagation.

Actually, however, water is slightly compressible. The inertia of the water tends to resist the motion of the plane. When, for example, the plane moves to the right, it compresses the immediately adjacent water particles on the right. This increased pressure then acts on the particles of water farther to the right, causing them to move and exert pressure on the particles beyond them. The increase in pressure is thus propagated as a wave through the water. The speed of propagation is determined by the density and elasticity of the medium.

When the plane moves toward the left, it causes a reduced pressure tending to drag the water particles back again. The motion of these particles in turn reduces the pressure on the particles to their right, and so on. The wave of increased pressure is thus followed by a wave of reduced pressure.

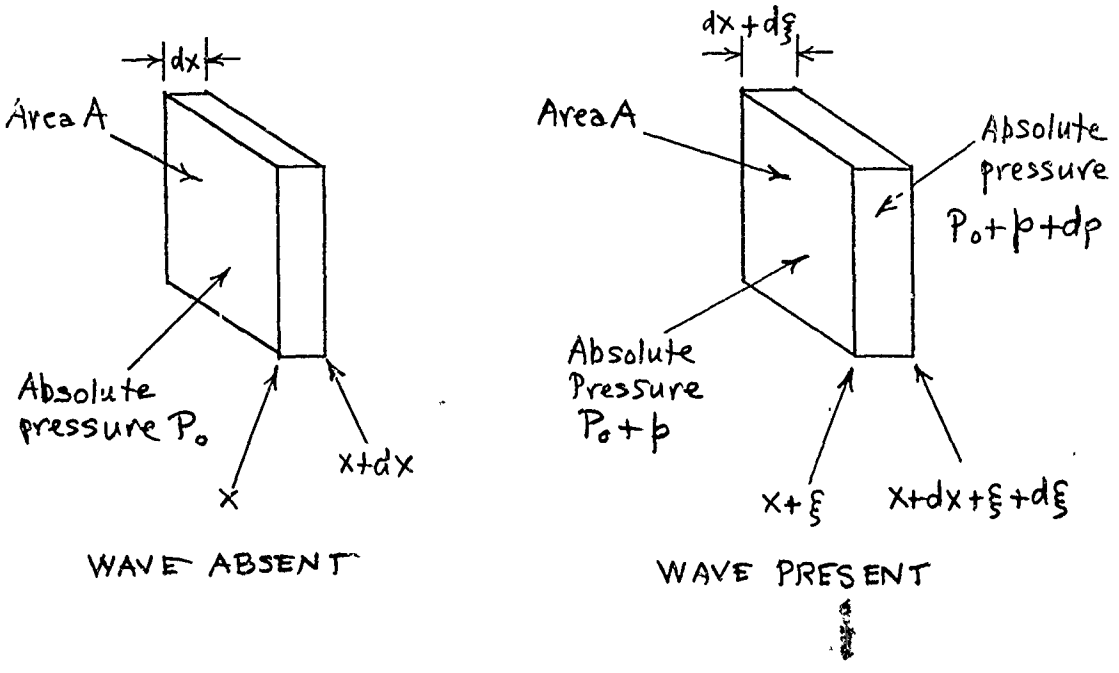
By this action the driving plane transmits its sinusoidal motion to all the particles of the medium. Because of the finite speed of propagation the phase of the oscillations is delayed by an amount proportional to the distance the wave has traveled from the source.

The motion of the particles in an acoustic wave in water is parallel to the direction of propagation. Such waves are called longitudinal waves, as contrasted with the transverse waves of a vibrating string where the motion of the particles is at right angles to the direction of propagation.

2. Differential Equation of Motion

Before discussing quantitatively the characteristics of plane acoustic waves, we shall first derive the basic differential equation and shall then show that its solution leads to a mathematical description of the observed phenomena.

For this purpose let us consider a very thin slab of water bounded by two planes parallel to the hypothetical driving plane. Let the area of the faces of the slab be A. (The value selected for A is immaterial, since by our hypothesis the motion of the water is the same at all points over the entire plane.) In the absence of sound waves the thickness of the



slab is dx, the left face being at a distance x from the driving plane. The mass of the slab is ρAdx, where ρ is the density of the water. The absolute pressure on the slab is P_o.

In the presence of acoustic waves the slab will move back and forth about the position x, and in addition the water will be compressed and expanded, so that the right-hand face will experience a slightly different motion than the left-hand face. Let ξ denote the displacement of the left-hand face from its equilibrium position at x. If we were to photograph the slab at a particular instant of time, we should find that the left-hand face has moved from x to $x + \xi$ and the right-hand face has moved from $x + dx$ to $x + dx + \xi + d\xi$. The pressure on the left is now $P_o + p$, where p is the fluctuating component of the pressure due to the presence of the waves, called the acoustic pressure. The pressure on the right face is $P_o + p + dp$.

In ordinary sound waves encountered in the sea, as distinguished from the shock waves resulting from explosions, the displacement is exceedingly small. It is therefore to be understood that

$$\xi \ll x$$

$$d\xi \ll dx$$

$$p \ll P_o$$

We are now in a position to calculate the acceleration of the slab by Newton's second law,

$$F = ma$$

The net force acting toward the right is the difference between the forces on the two faces,

$$F = (P + p)\,A - (P + p + dp)\,A = -A\,dp$$

The mass is

$$m = \rho A dx$$

The acceleration is

$$a = \frac{\partial^2 \xi}{\partial t^2}$$

(It will be noted that since the displacement ξ is a function of both x and t, we must use partial derivatives to indicate rates of change with time at a fixed value of x.) Therefore

$$-A\,dp = \rho\,A\,dx\,\frac{\partial^2 \xi}{\partial t^2}$$

or

$$\frac{\partial p}{\partial x} = -\rho\,\frac{\partial^2 \xi}{\partial t^2}$$

Note that the ratio of dp to dx is also a partial derivative, since the pressure increment corresponds to a change in x at a fixed instant of time. This equation says that the force exerted on a unit volume of the liquid is equal to the (negative of the) pressure gradient.

We must now evaluate the pressure gradient in terms of the elasticity of the medium. This is done by application of Hooke's law:

Stress = Constant x Strain

In the present situation the stress is the difference between the pressure on the slab when the waves are present and the equilibrium pressure in the absence of waves

$$\text{Stress} = (P_o + p) - P_o = p$$

The strain is the corresponding fractional change in volume. The equilibrium (unstressed) volume is

$$V_o = Adx$$

The volume when the waves are present is

$$V = A(dx + d\xi)$$

Since an increase in pressure causes the volume to decrease, the strain is

$$\text{Strain} = \frac{V_o - V}{V_o} = \frac{Adx - A(dx + d\xi)}{Adx} = -\frac{\partial \xi}{\partial x}$$

In the case of volume compression of a liquid, the constant of proportionality is the bulk modulus of elasticity, denoted by E.

Hooke's law therefore takes the form

$$p = E\,\frac{V_o - V}{V_o}$$

or

$$p = -E\,\frac{\partial \xi}{\partial x} \qquad (2)$$

If we take the partial derivative of (2) with respect to x, we obtain

$$\frac{\partial p}{\partial x} = -E\,\frac{\partial^2 \xi}{\partial x^2} \qquad (3)$$

Substitution of (3) into (1) yields

$$\frac{\partial^2 \xi}{\partial x^2} = \frac{\rho}{E}\,\frac{\partial^2 \xi}{\partial t^2} \qquad (4)$$

which is the basic differential equation for the particle displacement.

It will be seen from equation (2) that the bulk modulus E has the dimensions of pressure, which is force/area, or mass · acceleration/area. Density is mass/volume. Using the symbols M, L, and T for mass, length, and time we see that

$$E = \frac{ML}{T^2 L^2} = \frac{M}{LT^2}$$

$$\rho = \frac{M}{L^3}$$

and

$$\frac{E}{\rho} = \frac{M}{LT^2} \times \frac{L^3}{M} = \frac{L^2}{T^2}$$

The ratio $\sqrt{\frac{E}{\rho}}$ thus has the dimensions of velocity. It is, in fact, the speed of propagation of the waves, which will be denoted by the symbol c.

$$c = \sqrt{\frac{E}{\rho}} \tag{5}$$

This relation was derived by Sir Isaac Newton several centuries ago and within the past 50 years has been repeatedly checked by measurement of all three physical quantities.

The speed of sound in sea water is of the order of 5000 feet/second. Its value is not constant everywhere, but depends somewhat upon the temperature, the pressure (which of course is a function of the depth), and, to a somewhat lesser extent, upon the salinity, or salt content. The maximum variation is from about 4700 to about 5100 ft/sec. Although the variations are only a few percent, small changes in the speed of propagation are capable of exerting a profound influence on the performance of underwater acoustic systems, as we shall see later.

Equation (4), which we may now write in the form

$$\frac{\partial^2 \xi}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 \xi}{\partial t^2} \tag{4a}$$

is valid for a one dimensional wave such as we are now considering, but it does not hold in general for more complicated waves. The corresponding equation involving pressure instead of displacement is generally valid, however, and can be obtained by taking the first partial derivative of (1) with respect to x and the second partial derivative of (2) with respect to t. Thus

$$\frac{\partial^2 p}{\partial x^2} = -\rho \frac{\partial^3 \xi}{\partial x \partial t^2}$$

and $$\frac{\partial^2 p}{\partial t^2} = -E \frac{\partial^3 \xi}{\partial x \partial t^2}$$

so that $$\frac{\partial^2 p}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} \qquad (6)$$

3. Solution of the Differential Equation.

Since both (4a) and (6) have the same form, the solutions will be similar. There are two types of solutions, one of the form

$$\phi (t - x/c)$$

and the other of the form

$$\phi (t + x/c)$$

where ϕ can be any suitable function having continuous derivatives. To verify that these are solutions of the differential equation, all that is necessary is to take the derivatives and substitute them into the differential equation. For example, in the case of the first function $\phi (t - x/c)$, let

$$w = t - \frac{x}{c}$$

and $$\phi' = \frac{d\phi}{dw} ; \qquad \phi'' = \frac{d\phi'}{dw} = \frac{d^2\phi}{dw^2}$$

Then $$\frac{\partial \phi}{\partial x} = \frac{d\phi}{dw} \frac{\partial w}{\partial x} = -\frac{1}{c} \phi'$$

$$\frac{\partial^2 \phi}{\partial x^2} = -\frac{1}{c} \frac{d\phi'}{dw} \frac{\partial w}{\partial x} = \frac{1}{c^2} \phi''$$

and $$\frac{\partial \phi}{\partial t} = \frac{d\phi}{dw} \frac{\partial w}{\partial t} = \phi'$$

$$\frac{\partial^2 \phi}{\partial t^2} = \frac{d\phi'}{dw} \frac{\partial w}{\partial t} = \phi''$$

so that $$\frac{\partial^2 \phi}{\partial x^2} = \frac{1}{c^2} \phi'' = \frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2}$$

The function $\phi (t - \frac{x}{c})$ represents a wave traveling in the positive x direction. To see this we note that the function has a constant value for all values of x and t which satisfy the relation

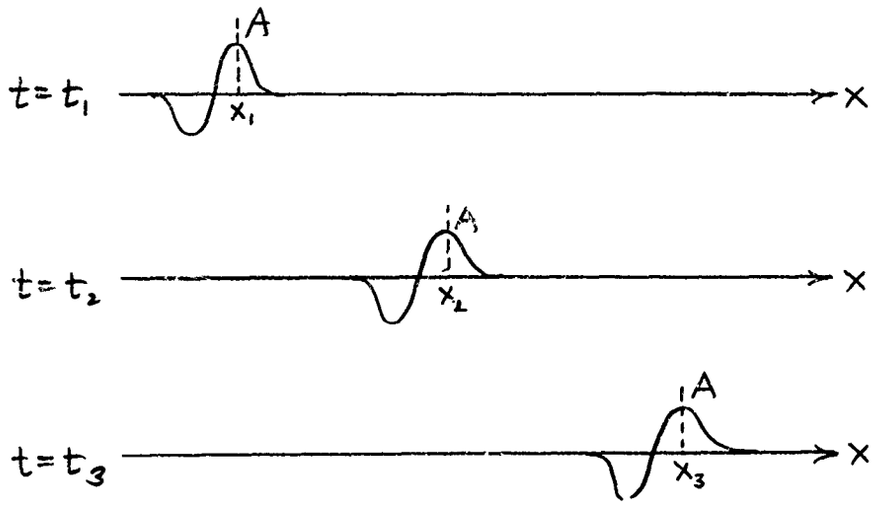
$$t - \frac{x}{c} = \text{constant}$$

or $$x = ct + \text{constant}$$

The behavior is illustrated in the following sketch which shows a wave "photographed" at three successive instants of time t_1, t_2, t_3. The location of any given point on the wave, such as the peak at A, progresses from x_1 at t_1 to x_2 at t_2, to x_3 at t_3, etc., where

$$x_1 - ct_1 = x_2 - ct_2 = x_3 - ct_3 = \text{constant}$$

Hence the shape of the wave remains fixed as the wave progresses uniformly along the x axis with speed c.



The function $\phi\,(t + x/c)$, on the other hand, represents a wave traveling with speed c in the negative x direction, as may be seen from the fact that the function has a constant value for all values of x and t which satisfy the relation

$$x = -ct + \text{constant}$$

Since the two waves differ only in the direction of propagation, but otherwise have the same physical properties, we shall for the present restrict our attention to the first type.

4. Sinusoidal Waves; Acoustic Pressure

In this survey of underwater acoustics we shall be concerned chiefly with sinusoidal waves or with more complicated waves which, by Fourier analysis, can be synthesized from sinusoidal waves. The pressure, displacement, and other related characteristics of sinusoidal waves are expressed in terms of sines and cosines of the angle $2\pi f(t - x/c)$ or $\omega(t - x/c)$, where f is the frequency in cycles per second and $\omega = 2\pi f$ is the "angular frequency" in radians per second.

Following the conventional practice in a.c. circuit theory, we shall use the complex exponential notation

$$e^{2\pi jf(t-x/c)} \quad \text{or} \quad e^{j\omega(t-x/c)}$$

which is useful because it tends to simplify the mathematical operations. The relationship between the exponential and the sine and cosine is given by the basic mathematical theorems

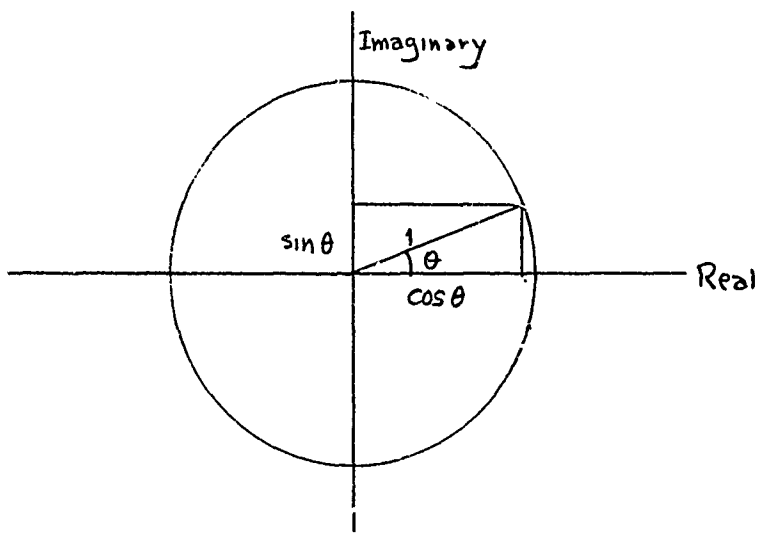
$$e^{j\theta} = \cos\theta + j\sin\theta \tag{7}$$

$$\cos\theta = \tfrac{1}{2}(e^{j\theta} + e^{-j\theta}) \tag{8}$$

$$\sin\theta = \frac{1}{2j}(e^{j\theta} - e^{-j\theta}) \tag{9}$$

where $j = \sqrt{-1}$

The complex exponential $e^{j\theta}$ is thus a complex number whose real part is cos θ and whose imaginary part is sin θ. When plotted in the complex



plane, $e^{j\theta}$ is seen to be a vector of unit length, making an angle θ with the positive real axis. As θ is varied from 0 to 2π radians, the tip of the vector traces out a circle of unit radius. Of particular interest are the following four values of θ:

θ degrees	θ radians	$e^{j\theta}$
-90	$-\pi/2$	$-j$
0	0	$+1$
90	$\pi/2$	$+j$
± 180	$\pm\pi$	-1

When a characteristic of a wave, such as the pressure, is expressed as a complex exponential, it is to be understood that the actual physical quantity is represented by the real part of the complex number. For example, if the pressure is given as

$$p = A\, e^{j\omega(t - x/c)} \quad \text{(complex)}$$

where A is a real number, then the actual physical pressure is

$$p = A \cos \omega(t - x/c) \quad \text{(real)}$$

The solution of a second order differential equation involves two constants of integration. In a continuous wave these constants can be interpreted as the amplitude and phase. The general expression for the pressure is thus

$$p = p_m\, e^{j\phi}\, e^{2\pi jf(t - \frac{x}{c})}$$

$$= p_m\, e^{j\left[2\pi f(t - \frac{x}{c}) + \phi\right]}$$

where p_m is the amplitude and ϕ is the phase. The real pressure is

$$p = p_m \cos\left[2\pi f(t - \frac{x}{c}) + \phi\right]$$

Note: From a strict mathematical standpoint, different symbols should be used for the complex and real quantities. However, it is thought that the additional complexity is not warranted in these notes.

From a practical standpoint, when we are considering a particular physical quantity by itself, the phase angle is of no significance, since without loss of generality it can be eliminated by a slight shift in the origin

of t (or of x). However, when we are investigating the relationship between one physical quantity and another, such as pressure and displacement, the relative phase between the two quantities is of great importance. For our present discussion we shall treat the pressure as having zero phase angle and shall relate the phases of all other quantities to that of the pressure.

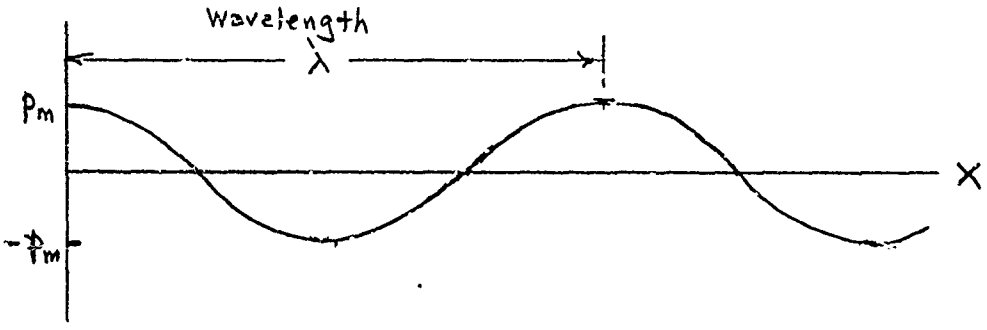
The complex instantaneous pressure is expressed by the formula

$$p = p_m e^{2\pi jf(t - \frac{x}{c})} = p_m e^{j\omega(t - \frac{x}{c})} \qquad (10)$$

the real part being

$$p = p_m \cos 2\pi f(t - \frac{x}{c}) = p_m \cos \omega(t - \frac{x}{c}) \qquad (10a)$$

If we were to sit at a fixed location in space and were to measure the instantaneous pressure as a function of time, we should see that it oscillates sinusoidally between $- p_m$ and $+ p_m$. Conversely, if we were to "freeze" the wave at any instant of time, we should observe that it varies sinusoidally with x, as indicated in the following sketch, which is drawn at time t = 0. The distance corresponding to one cycle is called the



wavelength, λ. To determine the relation between the wavelength and other parameters, we note that when x changes by an amount λ, the phase angle in (10) and (10a) changes by 2π radians. Thus,

$$2\pi f\,(t - \frac{x}{c}) - 2\pi f\,(t - \frac{x + \lambda}{c}) = 2\pi$$

or

$$\lambda f = c \qquad (11)$$

Equation (11) is an extremely important fundamental relationship. By use of this formula, the phase angle is frequently expressed in a variety of equivalent forms, such as

$$\begin{aligned} 2\pi f(t - \frac{x}{c}) &= \omega(t - \frac{x}{c}) \\ &= 2\pi(ft - \frac{x}{\lambda}) \\ &= \omega t - \frac{2\pi x}{\lambda} \qquad (12) \\ &= \omega t - kx \\ &= k(ct - x) \end{aligned}$$

where $k = \frac{2\pi}{\lambda}$ = the wave number (13)

All of the above forms are used interchangeably.

5. Particle Displacement and Particle Velocity.

The particle displacement ξ has already been defined as the displacement of the particles of water from their static equilibrium position due to the passage of a wave. Its direction is parallel to the direction of propagation, which, in our one-dimensional plane wave, is along the x axis. The particle velocity, which will be denoted by the symbol u, is the rate of change of ξ with respect to time,

$$u = \frac{\partial \xi}{\partial t} \qquad (14)$$

The relationship between the displacement and the acoustic pressure is expressed by equation (2). In applying this equation, it will be desirable to express the bulk modulus E in terms of the density and the speed of sound by substitution of (5). Thus,

$$p = -\rho c^2 \frac{\partial \xi}{\partial x} \qquad (15)$$

We know that since ξ is a solution of the differential equation (4a), it will have the form

$$\xi = B\, e^{j\omega(t - \frac{x}{c})}$$

The unknown constant B can be evaluated by differentiating ξ with respect to x and substituting into (15)

$$\frac{\partial \xi}{\partial x} = -\frac{j\omega B}{c} e^{j\omega(t - \frac{x}{c})} = -\frac{p}{\rho c^2}$$

$$= -\frac{p_m}{\rho c^2} e^{j\omega(t - \frac{x}{c})}$$

The value of B is

$$B = \frac{p_m}{j\omega\rho c} = -j\frac{p_m}{\omega\rho c}$$

The factor - j indicates that the displacement is 90 degrees out of phase with the pressure. If we let ξ_m denote the amplitude of the displacement, we have

$$\xi = -j\xi_m e^{j\omega(t - \frac{x}{c})} \tag{16}$$

from which it is seen that

$$\xi_m = \frac{p_m}{\omega\rho c} \tag{17}$$

To observe the phase relationship we note that

$$-j = e^{-j\frac{\pi}{2}}$$

Hence

$$\xi = \xi_m e^{j\left[\omega(t - \frac{x}{c}) - \frac{\pi}{2}\right]}$$

The term $-\frac{\pi}{2}$ signifies that the displacement lags behind the pressure in phase by 90°. The real displacement is

$$\xi = \xi_m \sin\omega(t - \frac{x}{c}) \tag{16a}$$

Taking the derative of (16) we obtain for the particle velocity

$$u = \omega\xi_m e^{j\omega(t - \frac{x}{c})}$$

If u_m denotes the amplitude of the particle velocity, so that

$$u = u_m e^{j\omega(t - \frac{x}{c})} \tag{18}$$

we see that

$$u_m = \omega\xi_m \tag{19}$$

and hence, from (17)

$$p_m = \rho c\, u_m \tag{20}$$

It is further observed from (10) and (18) that the pressure and particle velocity are in phase, so that the preceding relationship holds not only for amplitudes but for the instantaneous values as well. Thus,

$$p = \rho c u \tag{21}$$

We shall learn later that although equation (21) holds for plane waves, it does not hold in general for other types of waves.

6. Pressure Gradient

The pressure gradient is the rate of change of the pressure with respect to space coordinates. In the special case of our one-dimensional plane wave, the pressure gradient is simply the rate of change of p with respect to x. From equation (10) we have

$$\frac{\partial p}{\partial x} = -\frac{j\omega p_m}{c} e^{j\omega(t - \frac{x}{c})}$$

The pressure gradient lags the pressure 90 degrees in phase. In terms of its amplitude the pressure gradient may be written

$$\frac{\partial p}{\partial x} = -j \left(\frac{\partial p}{\partial x}\right)_m e^{j\omega(t - \frac{x}{c})} \tag{22}$$

the real part being

$$\frac{\partial p}{\partial x} = \left(\frac{\partial p}{\partial x}\right)_m \sin \omega \left(t - \frac{x}{c}\right) \tag{22a}$$

The relation between $\left(\frac{\partial p}{\partial x}\right)_m$ and p_m is

$$\left(\frac{\partial p}{\partial x}\right)_m = \frac{\omega p_m}{c} = k p_m \tag{23}$$

7. Intensity

The utility of acoustic waves to man depends fundamentally upon their property of propagating energy from one location to another. In order to convey information, for example, one must transmit a certain amount of energy. The propagation of acoustic energy through a medium is expressed in terms of acoustic intensity, which is defined as the rate of flow of energy

across a unit area normal to the direction of propagation. To be more precise, let A represent the area of a small portion of such a normal surface surrounding the point at which the intensity is defined, and let dE/dt represent the rate of flow of energy across A. Then the intensity is

$$\text{Intensity} = \lim_{A \to 0} \frac{dE/dt}{A}$$

This definition is required in the general case where the energy flow varies from point to point over the normal surfaces. In the special case of plane waves, of course, the surface normal to the direction of propagation is a plane, and the intensity is uniform over the plane.

The relationship between the intensity and other properties of the wave which have already been discussed may be derived from two different points of view. Since each provides a different insight into the situation, we shall consider them both.

First, the propagation of energy along the wave may be thought of in terms of the work done on a sheet of water particles (in a plane normal to the direction of propagation) by the pressure exerted on them by their neighbors "upstream." The work dW by a force F acting through a distance dx is Fdx, and the power expended by the force is the work per unit time,

$$\frac{dW}{dt} = F \frac{dx}{dt}$$

If now we focus our attention on a unit area of the sheet, the force exerted on unit area is simply the pressure p, and the work per unit area per unit time is the intensity, while the velocity dx/dt is the particle velocity u. Hence the intensity is

$$I = pu \qquad (24)$$

It will be recalled that in a plane wave the acoustic pressure and the particle velocity are in phase, so that a positive acoustic pressure occurs when the water particles are moving in the positive x direction, whereas a negative acoustic pressure occurs when they are moving in the negative x direction. In other words, the particles move to the right under

higher-than-normal absolute pressure and to the left under lower-than-normal absolute pressure. Both of these conditions are accompanied by a transfer of energy to the right. At the instant when both p and u are zero there is no transfer. The propagation of energy thus occurs in pulses at the rate of two pulses per cycle.

It is interesting to note that if we were to consider a wave traveling in the opposite direction, the pressure and particle velocity would be functions of $t + x/c$ instead of $t - x/c$, and we should find that p is positive when u is negative, and vice versa (180 degrees out of phase), confirming the physical concept that energy is flowing in the negative x direction.

In the second approach we consider the energy density (i.e., energy per unit volume) associated with the wave and note that the intensity is equal to the product of the energy density and the velocity of propagation. Energy is present in two forms - kinetic and potential.

Kinetic energy. The familiar formula for kinetic energy is $\frac{1}{2}$(mass) x (velocity)^2. The kinetic energy density is therefore

$$\tfrac{1}{2}(\text{mass per unit volume}) \times (\text{velocity})^2 = \tfrac{1}{2}\rho u^2$$

This result can be written in terms of pressure instead of density by application of (21).

$$\text{K. E. density} = \frac{pu}{2c} \qquad (25)$$

Potential energy. It will be recalled that the passage of a wave is accompanied by a periodic compression and expansion of the water. It takes work to compress or expand the water from its original volume at the equilibrium static pressure to the new volume associated with the transient pressure due to the wave. The incremental work required to change the volume from any value V to V-dV is -pdV, where p is the acoustic pressure (differential pressure) corresponding to the volume V. The total work required is computed by integrating -pdV over the range from the initial to the final volume. The computation can be simplified by noting that according to Hooke's law the pressure is proportional to

the change between the original and final volume, and hence the average pressure is one-half the final pressure. The total work done on an original volume V_o is therefore

$$\tfrac{1}{2} p\,(V_o - V)$$

Applying Hooke's Law

$$p = E\,\frac{V_o - V}{V_o}$$

and noting that the potential energy density is the work done on unit volume, i.e., V_o = 1 unit, we obtain

$$\text{P.E. density} = \frac{p^2}{2E} = \frac{p^2}{2\rho c^2} = \frac{pu}{2c} \qquad (26)$$

(equations (5) and (21)).

Comparison of (25) and (26) shows that the kinetic energy density and potential energy density are equal, and the total energy density is pu/c. The intensity is obtained by multiplying the energy density by the velocity of propagation c, leading to the same result as we had before.

With the aid of equation (21), $p = \rho c u$, the intensity may be expressed in two alternate forms

$$I = pu \qquad (24)$$

$$= \frac{p^2}{\rho c} \qquad (24a)$$

$$= \rho c u^2 \qquad (24b)$$

In acoustic measurements it is not the usual practice to measure intensity directly. Most hydrophones in common use are of one of three types - sensitive to pressure, particle velocity, or pressure gradient. The intensity must be computed from the appropriate formulas. Actually, however, intensity, as expressed explicitly in units of rate of energy flow per unit area, is seldom used in underwater acoustics. The basic measurement is one of pressure, and the concept of intensity is involved only implicitly as "that intensity which is equivalent to such and such number of microbars." This practice appears to be quite satisfactory for those already working in the field, but it is a little awkward from the standpoint of indoctrinating new people.

8. Note on Units of Measurement

The unit in which underwater acoustic pressure is measured is the dyne/cm^2 or microbar (the two units are equivalent). Since the dyne/cm^2 is the unit of pressure in the cgs (centimeter-gram-second) system of units, it is desirable in working with the theoretical formulas of the preceding paragraphs to express all the physical quantities in the same system of units. The various quantities and their respective units are as follows:

time, t - sec

frequency, f - cycles/sec

angular frequency, ω - radians/sec

linear dimensions, x, ξ, λ - cm

velocity, u, c - cm/sec

acceleration - cm/sec^2

mass - gm

density, ρ - gm/cm^3

force (mass x accel.) - gm cm/sec^2 = dyne

pressure, p - dyne/cm^2 = microbar

modulus of elasticity, E - dyne/cm^2

pressure gradient, $\frac{\partial p}{\partial x}$ - dyne/cm^3

energy (work) - dyne cm = erg

also, 1 joule = 10^7 erg

energy density - erg/cm^3 or joule/cm^3

power - erg/sec or watt

1 watt = 1 joule/sec = 10^7 erg/sec

intensity, I - erg/cm^2 sec or watt/cm^2

1 watt/cm^2 = 10^7 erg/cm^2 sec

It should be pointed out that although the above set of units is useful in providing a consistent basis for theoretical computations, most practical measurements of distances, wavelengths, sound speed, etc. are made in more familiar units, as will be discussed later.

9. Average Values

In making acoustic measurements we are in general interested in average values rather than instantaneous values. Hydrophones which are sensitive to pressure or particle velocity measure the root mean square value of these quantities. The root mean square is the square root of the time average of the squares. In computing the square of a quantity we must work with real values rather than complex values, since the square of the imaginary part leads to a real number. The mean square pressure is the time average of

$$p_m^2 \cos^2 \omega \left(t - \frac{x}{c}\right)$$

The average of $\cos^2 \omega \left(t - \frac{x}{c}\right)$ over a cycle is $\frac{1}{2}$, so that

$$p_{rms} = \frac{1}{\sqrt{2}} p_m \tag{27}$$

Similarly,

$$u_{rms} = \frac{1}{\sqrt{2}} u_m \tag{27a}$$

$$\xi_{rms} = \frac{1}{\sqrt{2}} \xi_m \tag{27b}$$

and

$$\left(\frac{\partial p}{\partial x}\right)_{rms} = \frac{1}{\sqrt{2}} \left(\frac{\partial p}{\partial x}\right)_m \tag{27c}$$

The instantaneous value of the intensity, according to (24), is

$$I = p_m u_m \cos^2 \omega \left(t - \frac{x}{c}\right)$$

Since the average of $\cos^2 \omega \left(t - \frac{x}{c}\right)$ is $\frac{1}{2}$, the average intensity is

$$I_{av} = \tfrac{1}{2} p_m u_m \tag{27d}$$

Combining these results, we find that the following formulas apply to the average values

$$u_{rms} = \omega \xi_{rms} \tag{28}$$

$$p_{rms} = \rho c\, u_{rms} \tag{29}$$

$$\left(\frac{\partial p}{\partial x}\right)_{rms} = \frac{\omega}{c}\, p_{rms} = k p_{rms} \tag{30}$$

$$I_{av} = p_{rms}\, u_{rms} = \frac{p_{rms}^2}{\rho c} = \rho c u_{rms}^2 \tag{31}$$

In the later portions of this course we shall be dealing exclusively with rms values and shall then drop the subscripts to simplify the notation. For the time being, however, we shall retain the current notation. It should be stressed that all formulas expressed in the form

$$\text{Instantaneous value} = (\text{Amplitude})\, e^{j\omega\left(t - \frac{x}{c}\right)}$$

can also be written in the equivalent form

$$\text{Instantaneous value} = \sqrt{2}\ (\text{RMS value})\, e^{j\omega\left(t - \frac{x}{c}\right)}$$

10. Acoustic Impedance

It is instructive to note the analogy between the formulas for acoustic waves electrical a.c. circuits

p	$= \rho c u$	e	$= Ri$
I	$= pu$	P	$= ei$
	$= \rho c u^2$		$= Ri^2$
	$= p^2 / \rho c$		$= e^2 / R$

The acoustic formulas apply to plane waves and the electrical formulas apply to a purely resistive a.c. circuit. Note the correspondence:

Acoustic	Electrical
pressure, p	voltage, e
velocity, u	current, i
intensity, I	power, P
ρc	resistance, R

It is seen that the product ρc, which is a property of the acoustic medium, is the analog of electrical resistance. On the basis of this

analogy, the quantity ρc is called the specific acoustic impedance of the medium. It is measured in units of specific acoustic ohms. In the fundamental cgs system of units the density ρ is expressed in gm/cm^3 and the speed of sound c in cm/sec, so that

1 specific acoustic ohm = 1 (gm/cm^3)(cm/sec)

Typical values in sea water at 60°F and in air at 20°C and one atmosphere pressure are given in the following table:

	Water	Air
ρ (gm/cm^3)	1.026	0.00121
c (cm/sec)	150,000	34,400
ρc (sp.ac.ohms)	154,000	42

The specific acoustic impedance is a very important factor in the design of sonar equipment.

Since the above analogy involves only resistive (non-reactive) circuits, it may be wondered why the term "impedance" is used rather than "resistance." The answer is that thus far we have been considering only plane waves. When we come to other types of waves, such as spherical waves, we shall find that the pressure and the particle velocity are no longer in phase, and hence the wave is analogous to an electric circuit containing both resistance and reactance. The ratio of pressure to particle velocity is then analogous to an impedance.

11. Standing Waves

It frequently happens, both in the ocean and in transducer calibration tanks, that reflections from the boundaries of the medium give rise to additional waves, so that waves traveling in more than one direction may be present in the water at the same time. This phenomenon produces marked changes in the acoustical pattern and unless suitable precautions are taken can introduce significant errors into acoustic measurements.

To illustrate the effect we shall consider the ideal case of perfect reflection from a boundary at right angles to the direction of

propagation, producing two wave trains of equal intensity, traveling in opposite directions. (It will be noted that to achieve such a condition in practice it would be necessary to have two such reflectors separated by the proper distance, like the case of a violin string, but we shall not be concerned with these details in the present discussion.)

To distinguish the two waves let us employ the subscript 1 to identify the wave traveling in the positive x direction and the subscript 2 for the wave traveling in the negative x direction. For the first wave the pressure, particle velocity, and pressure gradient are

$$u_1 = u_m e^{j\omega(t - \frac{x}{c})}; \quad p_1 = p_m e^{j(t - \frac{x}{c})}$$

$$(\frac{\partial p}{\partial x})_1 = -j(\frac{\partial p}{\partial x})_m e^{j\omega(t - \frac{x}{c})}$$

where

$$p_m = \rho c u_m$$

$$u_m = \omega \xi_m$$

$$(\frac{\partial p}{\partial x})_m = \frac{\omega p_m}{c} = k p_m$$

For the second wave the pressure is

$$p_2 = p_m e^{j\omega(t + \frac{x}{c})}$$

In deriving expressions for the other quantities we follow the same procedure as before. The only difference between the two results is that derivatives with respect to x will have the opposite sign. Since one such derivative is involved in each case, the results are

$$u_2 = -u_m e^{j\omega(t + \frac{x}{c})}$$

$$(\frac{\partial p}{\partial x})_2 = +j(\frac{\partial p}{\partial x})_m e^{j\omega(t + \frac{x}{c})}$$

The sign reversal is quite logical, since both the particle velocity and pressure gradient are vectors whose direction is the direction of propagation.

Pressure, on the other hand, is a scalar quantity, independent of direction. According to Pascal's law, pressure is transmitted equally in all directions.

The resultant pressure, particle velocity, and pressure gradient are obtained by adding the contributions from the two waves. The pressure is

$$p = p_1 + p_2 = p_m \left[e^{j\omega\left(t - \frac{x}{c}\right)} + e^{j\omega\left(t + \frac{x}{c}\right)} \right]$$

$$= p_m \left(e^{\frac{j\omega x}{c}} + e^{-\frac{j\omega x}{c}} \right) e^{j\omega t}$$

By equation (8)

$$p = 2p_m \cos kx \, e^{j\omega t} \tag{32}$$

Similarly

$$u = -2j \, u_m \sin kx \, e^{j\omega t} \tag{33}$$

$$\frac{\partial p}{\partial x} = -2 \left(\frac{\partial p}{\partial x} \right)_m \sin kx \, e^{j\omega t} \tag{34}$$

where $$k = \frac{2\pi}{\lambda}$$

Considering first the pressure, it will be seen that the amplitude $2 p_m \cos kx$ is a sinusoidal function of the distance x along the wave. At the points where

$$kx = \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \ldots$$

i.e., where

$$x = \frac{\lambda}{4}, \frac{3\lambda}{4}, \frac{5\lambda}{4}, \ldots$$

the amplitude is zero and there is no acoustic pressure at any time. These points are called nodes, pressure nodes in this case. At intermediate points where

$$x = 0, \frac{\lambda}{2}, \lambda, \frac{3\lambda}{2}, \ldots$$

we find that $\cos kx = \pm 1$ and the amplitude is a maximum. (A negative amplitude, of course, signifies a phase reversal.) Locations of maximum

amplitude are called loops. The pressure at a loop is twice the pressure of either of the two waves separately. This pressure field, which consists of a stationary pattern of loops and nodes, is an example of the well-known phenomenon of standing waves. The presence of such a condition in a transducer calibration tank can lead to serious errors because the reading of a measuring hydrophone depends critically upon where the instrument is located relative to the loops and nodes. Therefore, great care must be taken in designing calibration facilities to avoid interference by reflections from the water boundaries.

A velocity-sensitive hydrophone will also experience loops and nodes, but the pattern, as may be seen from (33) is different. The amplitude of the particle velocity is proportional to sin kx instead of cos kx. Therefore, velocity nodes occur at pressure loops and velocity loops at pressure nodes. The factor - j in (33) indicates that the particle velocity oscillations (with respect to time) are 90 degrees out of phase with the pressure oscillations.

Equation (34) shows that the pressure gradient has loops and nodes at the same places as the particle velocity but that the oscillations are 90 degrees out of phase.

As a final observation, it will be noted that since the pressure and particle velocity are 90 degrees out of phase, the average intensity is zero, just as zero power is dissipated in a purely reactive electric circuit. This is readily verified by writing down the product of the real parts of p and u

$$I = 4\, p_m\, u_m \sin kx \cos kx \sin \omega t \cos \omega t$$

Note that $2 \sin \omega t \cos \omega t = \sin 2\, \omega t$, whose average is zero. The physical significance of the zero intensity is that equal amounts of energy are being carried in opposite directions by the two waves.

Physical Properties of Acoustic Waves in Water (cont'd)

B. Three Dimensional Waves

In this section we shall derive the differential equation in three dimensions, expressing it not only in the rectangular but also in the cylindrical and spherical polar coordinate systems. We shall discuss briefly the characteristics of plane waves in two and three dimensions, and the simplest case of spherical waves, that is, waves whose properties are a function only of the radial coordinate and are independent of the angular coordinates. We shall also say a few words about cylindrical waves.

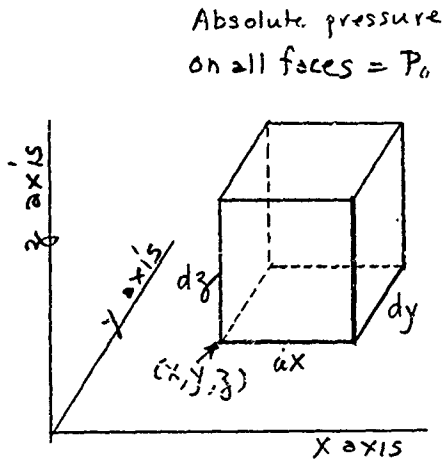
1. The Differential Equation.

The physical concepts involved in the derivation of the differential equation in three dimensions are basically the same as in the one-dimensional case. The principal difference is that the mathematics becomes somewhat more complicated.

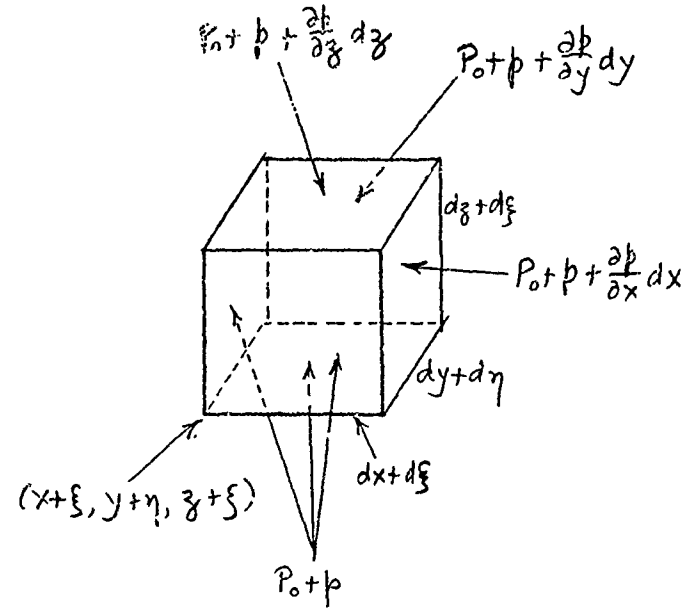
The basic element of water on which we focus our attention is an infinitesimal cube with dimensions dx, dy, dz, located at the point (x, y, z) in a rectangular coordinate system. These remarks pertain to the static condition of the cube, i.e., in still water with no waves present. When the wave is present, the cube is displaced from its static position (x, y, z) to a new position $(x + \xi, y + \eta, z + \zeta)$, and its dimensions are changed from dx, dy, dz to $dx + d\xi$, $dy + d\eta$, $dz + d\zeta$. It should be noted that the displacements associated with ordinary acoustic waves are exceedingly small, so that

$$\xi << x, \quad \eta << y, \quad \zeta << z, \quad d\xi << dx, \text{ etc.}$$

As in the one-dimensional case, the absolute static pressure is P_o and the absolute pressure in the presence of the wave is $P_o + p$, where p is the acoustic pressure.



STATIC CONDITION
(No wave)



WAVE PRESENT

The method of procedure is basically the same as in the one-dimensional problem. We shall formulate two equations: the first, based on Newton's second law of motion, will relate the pressure gradient to the acceleration of the little cube of water, and the second, based on Hooke's law, will relate the acoustic pressure to the change in volume. The final result will be obtained by suitably combining these two equations.

Since force and acceleration are vector quantities, we shall have to deal with three components. Consider first the x-component. The acceleration of the cube in the x-direction is caused by the differential pressure between the two faces perpendicular to the x-axis. Assuming that $d\xi$, $d\eta$, $d\zeta$ are negligible in comparison with dx, dy, dz, this difference of pressure is $\frac{\partial p}{\partial x}dx$, and the area of each face is dy dz. The net force in the positive x direction is $-\frac{\partial p}{\partial x}$ dx dy dz. The mass of the cube is $\rho dx dy dz$, and the acceleration is $\partial^2\xi/\partial t^2$. In the equation the factor dx dy dz appears on both sides and may be cancelled out. Therefore the x-component is

$$\frac{\partial p}{\partial x} = -\rho \frac{\partial^2 \xi}{\partial t^2} \tag{101}$$

(which is identical with equation (1) of the one-dimensional case). A similar argument shows that the corresponding y- and z-components are

$$\frac{\partial p}{\partial y} = -\rho \frac{\partial^2 y}{\partial t^2} \tag{102}$$

$$\frac{\partial p}{\partial z} = -\rho \frac{\partial^2}{\partial t^2} \tag{103}$$

By the use of conventional vector notation these three equations can be combined into a single vector equation. The displacement is in reality a vector having components ξ, η, and ζ. If we denote this vector by $\vec{d}$, it may be written in the form

$$\vec{d} = \hat{i}\xi + \hat{j}\eta + \hat{k}\zeta \tag{104}$$

where $\hat{i}$, $\hat{j}$, $\hat{k}$ are unit vectors pointing along the x, y, z axes. To simplify the notation let us indicate time differentiation of $\vec{d}$ with a dot (second derivative with a double dot). The three component equations may be combined to give

$$\hat{i}\frac{\partial p}{\partial x} + \hat{j}\frac{\partial p}{\partial y} + \hat{k}\frac{\partial p}{\partial z} = -\rho \ddot{\vec{d}} \tag{105}$$

The commonly used symbol for the vector differential operator on the left-hand side of (102) is ∇ (*del*)

$$\nabla = \hat{i}\frac{\partial}{\partial x} + \hat{j}\frac{\partial}{\partial y} + \hat{k}\frac{\partial}{\partial z} \tag{106}$$

When this vector operates on a scalar function, such as pressure, the resulting derivative is called a <u>gradient</u>. Thus

$$\nabla p = \text{grad } p$$

so that (105) may be rewritten in the form

$$\text{grad } p = -\rho \ddot{\vec{d}} \tag{105a}$$

This is the three-dimensional equivalent of equation (1). The gradient is a vector describing the rate of change of a scalar function with respect to space coordinates. Its direction is the direction in which the rate of change is a maximum, and its magnitude is the value of that rate of change.

In the Hooke's law equation the stress, as before, is the acoustic pressure. To find the strain we note that the original (static) volume is

$$dV_o = dx\, dy\, dz$$

and the stressed volume is

$$dV = (dx + d\xi)(dy + d\eta)(dz + d\zeta) = (1 + \frac{\partial \xi}{\partial x})(1 + \frac{\partial \eta}{\partial y})(1 + \frac{\partial \zeta}{\partial z})\, dx\, dy\, dz$$

Since these derivatives are numerically very small in comparison with unity, products of two or three of them may be neglected giving approximately

$$dV = (1 + \frac{\partial \xi}{\partial x} + \frac{\partial \eta}{\partial y} + \frac{\partial \zeta}{\partial z})\, dx\, dy\, dz$$

The strain is therefore

$$\frac{dV_o - dV}{dV_o} = -(\frac{\partial \xi}{\partial x} + \frac{\partial \eta}{\partial y} + \frac{\partial \zeta}{\partial z})$$

and Hooke's law takes the form

$$p = -E(\frac{\partial \xi}{\partial x} + \frac{\partial \eta}{\partial y} + \frac{\partial \zeta}{\partial z}) \qquad (107)$$

The sum of the three derivatives above can be put in more compact form through application of the operator ∇. It is in fact the dot product (or scalar product) of the two vectors ∇ and $\vec{d}$. The dot product of two vectors A and B is defined as follows

$$\begin{aligned} \vec{A} \cdot \vec{B} &= (\hat{i} A_x + \hat{j} A_y + \hat{k} A_z) \cdot (\hat{i} B_x + \hat{j} B_y + \hat{k} B_z) \\ &= A_x B_x + A_y B_y + A_z B_z \end{aligned} \qquad (108)$$

The dot product of ∇ operating on a vector function is called the <u>divergence</u> of the vector. Therefore

$$\nabla \cdot \vec{d} = \text{div}\, \vec{d} = \frac{\partial \xi}{\partial x} + \frac{\partial \eta}{\partial y} + \frac{\partial \zeta}{\partial z} \qquad (109)$$

and (107) may be rewritten as

$$p = -E\, \text{div}\, \vec{d} \qquad (107a)$$

We now have two equations (105a) and (107a) relating p and $\vec{d}$. We shall obtain a single equation in p by eliminating $\vec{d}$. To do this let us first take the second time derivative of (107a), obtaining

$$\frac{\partial^2 p}{\partial t^2} = - E \operatorname{div} \ddot{\vec{d}} \tag{110}$$

Next, we take the divergence of (105a), obtaining

$$\operatorname{div} \operatorname{grad} p = - \rho \operatorname{div} \ddot{\vec{d}}$$

Combining this with (110) yields

$$\operatorname{div} \operatorname{grad} p = \frac{\rho}{E} \frac{\partial^2 p}{\partial t^2} = \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2}$$

The divergence of a gradient is

$$\begin{aligned} \text{div grad} &= \nabla \cdot \nabla = \nabla^2 \\ &= \left(\hat{i} \frac{\partial}{\partial x} + \hat{j} \frac{\partial}{\partial y} + \hat{k} \frac{\partial}{\partial z}\right) \cdot \left(\hat{i} \frac{\partial}{\partial x} + \hat{j} \frac{\partial}{\partial y} + \hat{k} \frac{\partial}{\partial z}\right) \\ &= \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \end{aligned} \tag{111}$$

The final result is therefore

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} = \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} \tag{112}$$

or, in vector notation

$$\nabla^2 p = \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} \tag{113}$$

The differential equation (112) is an obvious generalization of the one-dimensional equation (6). However, if one attempts to obtain a vector differential equation in $\vec{d}$ by eliminating the pressure, one does not achieve a correspondingly simple result. It is noted in passing that an equation of this type is satisfied by a function called the velocity potential defined by the vector equation

$$- \text{grad (veloc. pot.)} = \text{particle velocity}$$

The particle velocity is the time derivative of the displacement

$$\vec{u} = \dot{\vec{d}} = \frac{\partial \vec{d}}{\partial t} \tag{114}$$

It is a vector whose components we shall designate as u, v, w.

$$\vec{u} = \hat{i}\,u + \hat{j}\,v + \hat{k}\,w \tag{115}$$

A useful relation between the particle velocity and the pressure is obtained from (105a), namely

$$\text{grad } p = -\rho \frac{\partial \vec{u}}{\partial t} \tag{116}$$

One of the benefits which accrue from the use of vector notation is that a vector has a fixed magnitude and direction independent of the coordinate system in which it is described. The components are different in different coordinate systems, but the vector is the same. The differential equation (113), expressed in terms of the vector operator ∇, is an example. By using the proper expression for ∇^2, we can describe acoustic waves in any coordinate system.

The two most widely useful non-Cartesian coordinate systems are the cylindrical and spherical polar systems. The cylindrical coordinates r, ϕ, z are related to the Cartesian coordinates as follows

$$\begin{aligned} x &= r\cos\phi \\ y &= r\sin\phi \\ z &= z \end{aligned} \tag{117}$$

Let a be a scalar function and $\vec{A}$ a vector function. Then, in the cylindrical coordinate system,

$$\text{grad } a = \hat{r}\frac{\partial a}{\partial r} + \hat{\phi}\frac{1}{r}\frac{\partial a}{\partial \phi} + \hat{z}\frac{\partial a}{\partial z} \tag{118}$$

$$\text{div } \vec{A} = \frac{1}{r}\frac{\partial}{\partial r}(rA_r) + \frac{1}{r}\frac{\partial A\phi}{\partial \phi} + \frac{\partial Az}{\partial z} \tag{119}$$

and the wave equation is

$$\nabla^2 p = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial p}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 p}{\partial \phi^2} + \frac{\partial^2 p}{\partial z^2} = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} \tag{120}$$

Note: $\hat{r}, \hat{\phi}, \hat{z}$ are unit vectors in the radial, tangential, and axial directions respectively, and A_r, A_ϕ, A_z are the components of $\vec{A}$ in these directions.

The spherical polar coordinates r, θ, φ are related to the Cartesian coordinates as follows

$$\begin{aligned} x &= r \sin\theta \cos\phi \\ y &= r \sin\theta \sin\phi \\ z &= r \cos\theta \end{aligned} \qquad (121)$$

Here we have

$$\text{grad } a = \hat{r}\frac{\partial a}{\partial r} + \hat{\theta}\frac{1}{r}\frac{\partial a}{\partial \theta} + \hat{\phi}\frac{1}{r\sin\theta}\frac{\partial a}{\partial \phi} \qquad (122)$$

$$\text{div}\,\vec{A} = \frac{1}{r^2}\frac{\partial}{\partial r}(r^2 A_r) + \frac{1}{r\sin\theta}\frac{\partial}{\partial \theta}(A_\theta \sin\theta) + \frac{1}{r\sin\theta}\frac{\partial A_\phi}{\partial \phi} \qquad (123)$$

$$\nabla^2 p = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial p}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta\frac{\partial p}{\partial \theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 p}{\partial \phi^2} \qquad (124)$$

Note: Spherical coordinates are related to geographical coordinates on the earth. r is the radial coordinate, θ is the co-latitude, measured from the north pole, and φ is the longitude. $\hat{r}$ is a unit vector in the radial direction, $\hat{\theta}$ is a unit vector directed along a meridian toward the south, and $\hat{\phi}$ is a unit vector directed along a parallel of latitude toward the east.

2. Three-Dimensional Plane Waves.

The solution of (112) in terms of sinusoidal functions is

$$p = p_m e^{j\omega\left[t - \frac{1}{c}(\alpha x + \beta y + \gamma z)\right]} \qquad (125)$$

Substitution of this solution into the differential equation yields a relation among the three constants α, β, γ.

$$\alpha^2 + \beta^2 + \gamma^2 = 1 \qquad (126)$$

These three numbers are thus the direction cosines of a line perpendicular to the plane

$$\alpha x + \beta y + \gamma z = \text{constant} \qquad (127)$$

The plane defined by (127) is a surface of constant phase, that is, a wave-front, and the line perpendicular to it indicates the direction in which the wave is moving. Therefore α, β, γ are the direction cosines of the direction of propagation.

The pressure gradient is found by taking derivatives of (125). In vector notation the gradient is

$$\text{grad } p = -\, j\, k\, p_m \,(\hat{i}\alpha + \hat{j}\beta + \hat{k}\gamma)\, e^{j\omega\left[t - \frac{1}{c}(\alpha x + \beta y + \gamma z)\right]} \qquad (128)$$

It is seen that the vector

$$\hat{i}\alpha + \hat{j}\beta + \hat{k}\gamma$$

is pointed along the direction of propagation and has the magnitude

$$\sqrt{\alpha^2 + \beta^2 + \gamma^2} = 1$$

The magnitude of the pressure gradient is thus the same as for the one-dimensional wave. (that is, $k\phi_m$).

The particle velocity vector can be obtained from the pressure gradient by dividing by $-\rho$ and integrating with respect to time, as may be seen from (116). Integration of the complex exponential with respect to time yields a factor $1/j\omega$. Hence

$$\vec{u} = -\frac{1}{j\omega\rho} \text{ grad } p \qquad (129)$$

or, from (128), since $k = \omega/c$,

$$\vec{u} = \frac{p_m}{\rho c}(\hat{i}\alpha + \hat{j}\beta + \hat{k}\gamma)\, e^{j\omega\left[t - \frac{1}{c}(\alpha x + \beta y + \gamma z)\right]} \qquad (130)$$

The particle displacement is obtained by one further integration, yielding

$$\vec{d} = -\frac{j p_m}{\omega \rho c}(\hat{i}\alpha + \hat{j}\beta + \hat{k}\gamma)\, e^{j\omega\left[t - \frac{1}{c}(\alpha x + \beta y + \gamma z)\right]} \qquad (131)$$

Again we see that both $\vec{d}$ and $\vec{u}$ point along the direction of propagation and that their magnitudes agree with the previous results for a one-dimensional wave.

3. Two-Dimensional Plane Waves.

The formulas take on a simple and useful form when we restrict the problem to two dimensions. If we limit the motion to the xy-plane, the direction cosines can be expressed in the form

$$\alpha = \cos\theta$$
$$\beta = \sin\theta$$
$$\gamma = 0$$

For this case the time-space factor in the exponent takes the form

$$t - \frac{1}{c}(x\cos\theta + y\sin\theta)$$

and the vector indicating direction of propagation is

$$\hat{i}\cos\theta + \hat{j}\sin\theta$$

These changes convert equations (125), (128), (130), and (131) to two dimensions. We shall write out explicitly the equations for the pressure and the two components, u and v, of the particle velocity

$$p = p_m e^{j\omega\left[t - \frac{1}{c}(x\cos\theta + y\sin\theta)\right]} \tag{132}$$

$$u = \frac{p_m}{\rho c}\cos\theta\, e^{j\omega\left[t - \frac{1}{c}(x\cos\theta + y\sin\theta)\right]} \tag{133}$$

$$v = \frac{p_m}{\rho c}\sin\theta\, e^{j\omega\left[t - \frac{1}{c}(x\cos\theta + y\sin\theta)\right]} \tag{134}$$

It is interesting to note that when $\theta = 0$ or 180 degrees, the two-dimensional wave reduces to one dimension. When $\theta = 0$, the wave travels in the positive x direction; when $\theta = 180$ degrees, the direction is reversed.

4. Spherical Waves.

We shall consider only the simple case in which the waves are functions of the radial coordinate only. Since only one of the three space coordinates is involved, the problem is reduced to one dimension, and vector notation will not be required. The differential equation (124) reduces to

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\phi}{\partial r}\right) = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} \tag{135}$$

The solution is

$$p = \frac{A}{r} e^{j(\omega t - kr)} \tag{136}$$

where A is a constant. It may be readily verified by differentiation that (136) satisfies the differential equation. The pressure gradient, as may be seen from (122), is the derivative with respect to r.

$$\frac{\partial p}{\partial r} = -\frac{A}{r^2} (1 + jkr) e^{j(\omega t - kr)} \tag{137}$$

The particle velocity, according to (129), is

$$u = -\frac{1}{j\omega\rho} \frac{\partial p}{\partial r}$$

or

$$u = \frac{A}{\rho c r} \left(1 - \frac{j}{kr}\right) e^{j(\omega t - kr)} \tag{138}$$

Here we see two respects in which spherical waves differ from plane waves. In the first place, the amplitude of the waves varies with the radius. Mathematically speaking, these waves spread out from a point source at the origin, though in the real world they can be generated by a small pulsating sphere. The pressure amplitude varies inversely as the first power of the radius.

The second distinguishing characteristic is a phase lag between the particle velocity and the pressure, indicated by the complex amplitude in (138). This is analogous to the phase shift between the current and voltage in an electric circuit containing both resistance and reactance, and leads to the concept of acoustic impedance, which will be discussed below.

The (average) intensity is the average of the product of the real parts of p and u. To simplify the notation, let

$$p_m = \frac{A}{r} \tag{139}$$

so that

$$p = p_m e^{j(\omega t - kr)} \tag{140}$$

The real pressure is then

$$p = p_m \cos(\omega t - kr) \tag{141}$$

and the real particle velocity is

$$u = \frac{p_m}{\rho c} \cos(\omega t - kr) + \frac{p_m}{\rho c k r} \sin(\omega t - kr) \tag{142}$$

The second term in (142) is 90 degrees out of phase with the pressure and corresponds to the reactive component in an electrical circuit, which dissipates no power. The average of the product $\sin(\omega t - kr)\cos(\omega t - kr)$ is zero. The average of $\cos^2(\omega t - kr)$ is $\frac{1}{2}$ and the intensity is therefore

$$I_{av} = \frac{p_m^2}{2\rho c} \tag{143}$$

or, in terms of the original constant A,

$$I_{av} = \frac{A^2}{2\rho c r^2} \tag{144}$$

showing that the intensity varies inversely as the square of the distance from the source. This is an illustration of the well-known inverse square law or spherical spreading law.

The constant A may be evaluated in terms of the total power radiated by the source. Since the intensity is the power transmitted across unit area of a sphere of radius r, and has a uniform value over the entire spherical surface, whose area is $4\pi r^2$, the total power is

$$\mathcal{P} = 4\pi r^2 I_{av} \tag{145}$$

Therefore, from (144),

$$A^2 = \frac{\rho c \mathcal{P}}{2\pi} \tag{146}$$

In our study of plane waves we derived three equivalent formulas (equation (31)) relating the intensity to the rms pressure and the rms particle velocity. Recalling that

$$p_{rms} = \frac{1}{\sqrt{2}} p_m \; ; \; u_{rms} = \frac{1}{\sqrt{2}} u_m$$

we see from (143) that one of these formulas,

$$I_{av} = \frac{p^2_{rms}}{\rho c} \tag{147}$$

is valid for spherical waves as well. It is of interest to inquire whether the other two formulas are also valid. The amplitude of the particle velocity of a spherical wave is the absolute value of the coefficient of the complex exponential in (138), i.e.,

$$u_m = \frac{A}{\rho c r}\left|1 - \frac{j}{kr}\right|$$

$$= \frac{A}{\rho c r}\sqrt{1 + \frac{1}{k^2 r^2}} = \frac{p_m}{\rho c}\sqrt{1 + \frac{1}{k^2 r^2}} \tag{148}$$

Therefore

$$p_{rms} = \frac{\rho c}{\sqrt{1 + \frac{1}{k^2 r^2}}}\, u_{rms} \tag{149}$$

Substituting this into (147), we obtain the results listed in the following table:

Intensity, I_{av}	
For plane waves	For spherical waves
$\frac{p^2_{rms}}{\rho c}$	$\frac{p^2_{rms}}{\rho c}$
$p_{rms}\, u_{rms}$	$\frac{p_{rms}\, u_{rms}}{\sqrt{1 + \frac{1}{k^2 r^2}}}$
$\rho c u^2_{rms}$	$\frac{\rho c u^2_{rms}}{1 + \frac{1}{k^2 r^2}}$

Thus the only generally valid formula is (147).

A few remarks are in order regarding the magnitude of the effect. The ratio of the imaginary to the real term in the complex amplitude of u is

$$\frac{1}{kr} = \frac{\lambda}{2\pi r}$$

For a phase shift of 5 degrees, which may be considered negligible, 1/kr is approximately 0.087. The corresponding value of r is

$$r = \frac{\lambda}{2\pi\ (0.087)} \approx 2\lambda$$

At distances greater than a few wavelengths away from a point sound source the phase shift due to the 1/kr term may be neglected and spherical waves may be treated as plane waves whose amplitude varies inversely as the first power of the distance from the source and whose intensity varies inversely as the square.

The phase relationships discussed above are important only in the immediate vicinity of a sonar transducer. (or other sound source)

5. Cylindrical Waves.

The solution of the wave equation in cylindrical coordinates (120) leads to Bessel's functions and is outside the scope of the present discussion. A useful approximate solution can be obtained, however, for large values of the radial coordinate r. Let us assume that the wave travels radially outward from an infinite line at r = 0 and is independent of the other two coordinates, ϕ and z. The differential equation for such a wave is

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial p}{\partial r}\right) = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2}\ , \qquad (150)$$

The approximate solution is

$$p = \frac{A}{\sqrt{r}}\ e^{\,j(\omega t - kr)} \qquad (151)$$

Taking derivatives, we find for the left side of (150)

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial p}{\partial r}\right) = -\frac{k^2 A}{\sqrt{r}}\left(1 - \frac{1}{4k^2r^2}\right)e^{\,j\,(\omega t - kr)}$$

and for the right side, since $k = \omega/c$,

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = -\frac{k^2 A}{\sqrt{r}}\ e^{\,j\,(\omega t - kr)}$$

If (151) is to be a good solution, the condition which must be met is

$$\frac{1}{4k^2 r^2} << 1 \qquad (152)$$

At distances beyond two wavelengths from the source, i.e., $r > 2\lambda$, we find

$$kr > 4\pi$$

or

$$\frac{1}{4k^2r^2} < \frac{1}{64\pi^2} \approx .0016$$

The error term is less than 0.2 percent. Therefore, at large distances of many wavelengths the solution (151) is extremely good.

It can be shown that the particle velocity of the wave is approximately* $u = \frac{A}{\rho c\sqrt{r}} e^{j(\omega t - kr)}$ (153)

and the (average) intensity is

$$I_{av} = \frac{A^2}{2\rho cr} \tag{154}$$

At large distances from the source, cylindrical waves may be treated as plane waves whose amplitude varies inversely as the square root of the radius and whose intensity varies inversely as the first power.

In the ideal problem we are considering the source is an infinite line. The intensity (154) is the rate of energy flow across unit area of the side of a cylinder of radius r. If P_1 is the total power radiated per unit length by the line source, then

$$P_1 = 2\pi r I_{av} \tag{155}$$

and the relationship between the power and the constant A is

$$A^2 = \frac{\rho c P_1}{\pi} \tag{156}$$

We may also obtain a solution of the cylindrical wave equation for the more general case where the direction of propagation has an axial component as well as radial. In this case the pressure is a function of z as well as r and the approximate solution for large r is

$$p = \frac{A}{\sqrt{r}} e^{j[\omega t - k(r\cos\theta + z\sin\theta)]}$$

When viewed in the rz plane, these waves behave like two-dimensional plane waves traveling in a direction making an angle θ with the r axis,

*If (152) is met.

except of course that the intensity decreases with the radial distance. The waves are spreading out on a cone.

6. Unit Area Acoustic Impedance and Admittance.

The analogy between acoustic waves and electrical circuits has been mentioned several times. It has been pointed out that acoustic pressure is the analog of an a.c. voltage, particle velocity the analog of current, and intensity the analog of electrical power. For plane waves the ratio of pressure to particle velocity was found to be ρc, a property of the medium. This is the analog of an electrical resistance and is called the specific acoustic impedance of the medium.

For spherical waves the ratio of the complex pressure to the complex particle velocity is a complex quantity and is thus analogous to electrical impedance. We may therefore define an impedance by the relation

$$p = \bar{z}_a u \tag{157}$$

where $\bar{z}_a$ is called the unit area acoustic impedance. The bar over the z_a denotes a complex quantity. (Strictly speaking, p and u should be similarly marked.) Since the real and imaginary parts of $\bar{z}_a$ are analogous to resistance and reactance, we may express the impedance in the conventional form

$$\bar{z}_a = r_a + j\, x_a \tag{158}$$

The magnitude of the impedance is the absolute value of $\bar{z}_a$,

$$z_a = |\bar{z}_a| = \sqrt{r_a^2 + x_a^2} \tag{159}$$

and the phase angle between the particle velocity and pressure is

$$\phi = \tan^{-1} \frac{x_a}{r_a} \tag{160}$$

so that

$$\cos \phi = \frac{r_a}{z_a}; \qquad \sin \phi = \frac{x_a}{z_a} \tag{161}$$

An alternate expression for the impedance is

$$\bar{z}_a = z_a e^{j\phi} \tag{162}$$

The magnitude of the impedance z_a expresses the ratio of the amplitudes of the pressure and particle velocity, and likewise, of course, the ratio of the rms values.

$$p_{rms} = z_a u_{rms} \tag{163}$$

The angle ϕ measures the phase lag between the particle velocity and pressure.

To compute the intensity let us express p and u in a simplified form. All that is significant is that the rms values are p_{rms} and u_{rms} and that they differ in phase by an angle ϕ. Hence we may write

$$u = \sqrt{2}\, u_{rms} \cos \omega t$$

$$p = \sqrt{2}\, p_{rms} \cos (\omega t + \phi)$$

It is readily shown that the average value of the product is

$$I_{av} = p_{rms}\, u_{rms} \cos \phi = \frac{r_a}{z_a}\, p_{rms}\, u_{rms} \tag{164}$$

Substitution of (163) yields the two alternate forms

$$I_{av} = \frac{r_a\, p_{rms}^2}{z_a^2} \tag{164a}$$

and

$$I_{av} = r_a\, u_{rms}^2 \tag{164b}$$

For a spherical wave $\bar{z}_a$ is obtained from (136) and (138) and is found to be

$$\bar{z}_a = \frac{\rho c}{1 - \frac{j}{kr}} = \rho c\, \frac{1 + \frac{j}{kr}}{1 + \frac{1}{k^2 r^2}}$$

from which

$$r_a = \frac{\rho c}{1 + \frac{1}{k^2 r^2}} \tag{165}$$

$$x_a = \frac{\frac{\rho c}{kr}}{1 + \frac{1}{k^2 r^2}} \tag{166}$$

$$z_a = \frac{\rho c}{\sqrt{1 + \frac{1}{k^2 r^2}}} \tag{167}$$

$$\phi = \tan^{-1} \frac{1}{kr} \tag{168}$$

It is readily shown that these formulas lead to the results derived previously.

The form in which the pressure and particle velocity are expressed in (136) and (138) suggests a more natural interpretation in terms of the complementary concept of admittance. In electrical circuit theory admittance is defined as the reciprocal of impedance, so that in the acoustic case we have

$$u = \bar{y}_a p \tag{169}$$

where $\bar{y}_a$ is the unit area acoustic admittance. The real and imaginary parts are the conductance g_a and the susceptance b_a.

$$\bar{y}_a = g_a + j\, b_a \tag{170}$$

$$y_a = |\bar{y}_a| = \sqrt{g_a^2 + b_a^2} \tag{171}$$

Comparing the admittance with the impedance we see that

$$\bar{y}_a = \frac{1}{\bar{z}_a} = \frac{1}{r_a + j\, x_a} = \frac{r_a - j\, x_a}{r_a^2 + x_a^2} \tag{172}$$

Therefore

$$g_a = \frac{r_a}{r_a^2 + x_a^2} \tag{173}$$

$$b_a = \frac{-x_a}{r_a^2 + x_a^2} \tag{174}$$

and

$$\phi = -\tan^{-1} \frac{b_a}{g_a} \tag{175}$$

$$\cos \phi = \frac{g_a}{y_a} \; ; \; \sin \phi = -\frac{b_a}{y_a} \tag{176}$$

$$\bar{y}_a = y_a e^{-j\phi} \tag{177}$$

In terms of admittance the pressure, particle velocity, and intensity are related thus

$$u_{rms} = y_a p_{rms} \tag{178}$$

$$I_{av} = g_a p_{rms}^2 \tag{179}$$

$$= \frac{g_a}{y_a} p_{rms} u_{rms} \tag{179a}$$

$$= \frac{g_a}{y_a^2} u_{rms}^2 \tag{179b}$$

For a spherical wave the admittance takes on a simpler form than the impedance.

$$\bar{y}_a = \frac{1}{\rho c}\left(1 - \frac{j}{kr}\right) \tag{180}$$

$$g_a = \frac{1}{\rho c} \tag{181}$$

$$b_a = -\frac{1}{\rho c k r} \tag{182}$$

$$y_a = \frac{1}{\rho c}\sqrt{1 + \frac{1}{k^2 r^2}} \tag{183}$$

It is readily shown that these formulas lead to the same results as before.

7. Acoustic Impedance and Mechanical Impedance.

The concept of unit area acoustic impedance (and admittance) has arisen in connection with a study of the propagation of acoustic waves through a homogeneous medium (water). The sonar designer is faced with another problem, the interaction between the water and acoustic hardware. In working with mechanical radiating elements having a finite area, one encounters the concepts of force and volume velocity. The relationship between the force on a transducer face and the acoustic pressure of the adjoining water is obviously

$$F = pA$$

where A is the area of the face. The volume velocity, which is the total volume of water displaced per unit time by the area A, is uA.

Two types of impedance are encountered in the interaction

$$\text{acoustic impedance} = \frac{\text{pressure}}{\text{volume velocity}} = \frac{p}{uA} = \frac{\bar{z}_a}{A}$$

$$\text{mechanical impedance} = \frac{\text{force}}{\text{particle velocity}} = \frac{pA}{u} = A\bar{z}_a$$

8. Noise Pressure; Equivalent Sine Wave Intensity

Acoustic noise is a random vibration which may be considered as a composite of sinusoidal waves of different frequencies, travelling in different directions. The frequencies are spread out over a continuous spectrum (discussed later), and in general the directions are spread out continuously in space. When a pressure-sensitive hydrophone is placed in such a sound field, the reading it gives is the resultant of the contributions of all frequencies within the bandwidth of its response.

To investigate the nature of this process we shall use a simple-minded approach. We shall first consider two plane waves of different angular frequencies, whose rms pressures are p_1 and p_2. The instantaneous pressures will be denoted by the symbols p_{i1} and p_{i2}. If p_{i1}, etc. we restrict our attention to a fixed location in space, the space coordinates in the phase angle will be constant, and we may write the instantaneous pressures in the simple form

$$p_{i1} = \sqrt{2}\, p_1 \cos(\omega_1 t + \phi_1)$$

$$p_{i2} = \sqrt{2}\, p_2 \cos(\omega_2 t + \phi_2)$$

The square of the resultant instantaneous pressure is

$$(p_{i1} + p_{i2})^2 = 2\Big[p_1^2 \cos^2(\omega_1 t + \phi_1) + 2p_1 p_2 \cos(\omega_1 t + \phi) \cos(\omega_2 t + \phi_2) + p_2^2 \cos^2(\omega_2 t + \phi_2)\Big]$$

In computing the time average of this quantity, we note that the square of the cosine averages to $\frac{1}{2}$ while the cross term averages to zero if the two frequencies are not the same. The latter statement may be proved from the trigonometric identity

$$\cos(\omega_1 t + \phi_1)\cos(\omega_2 t + \phi_2) = \cos[(\omega_1 + \omega_2)t + (\phi_1 + \phi_2)]$$

$$+ \cos[(\omega_1 - \omega_2)t + (\phi_i - \phi_2)]$$

Both terms on the right average to zero so long as $\omega_1 \neq \omega_2$. The resultant mean square pressure of the two waves is therefore

$$p^2 = p_1^2 + p_2^2$$

If there are present additional waves of other frequencies, the mean square value of the resultant of all of them may be computed in a similar way, and all the cross terms will average to zero, leaving

$$p^2 = p_1^2 + p_2^2 + p_3^2 + \cdots = \sum_i p_i^2 \qquad (184)$$

The rms resultant pressure is of course the square root of this. Thus, when waves of different frequencies are present, a pressure-sensitive hydrophone measures not the sum of the pressures of the individual waves, but the square root of the sums of the squares.

If both sides of (184) are divided by ρc, we obtain

$$\frac{p^2}{\rho c} = \frac{p_1^2}{\rho c} + \frac{p_2^2}{\rho c} + \frac{p_3^2}{\rho c} + \cdots \qquad (185)$$

The individual terms on the right-hand side of (185) represent the intensities of the respective component waves. If all these waves were traveling in the same direction, the sum would represent the actual intensity of the composite wave. Since, however, this is not the case in a noise field, the sum, $p^2/\rho c$, does not represent the intensity of a real physical wave. It is called the equivalent sine wave intensity. It is merely a hypothetical intensity inferred from a pressure measurement.

As indicated above, a noise field consists of waves traveling in all directions. If, for example, we had an ideal transducer having a very narrow beam with no side lobes, we should receive a finite amount of energy regardless of how we orient the transducer. If the characteristics of the noise field are such that we receive the same amount

from every direction, the noise field is called isotropic. Ambient sea noise in the near-surface region is approximately isotropic.

C. Waves with a Continuous Distribution of Frequencies.

1. Spectra

Many waves encountered in the sea are not simple sine waves; they are complex in form. As indicated previously, a complex wave may be considered as made up of a large number of sine wave components, each of a different frequency. The distribution of these components is called the spectrum of the wave.

If the spectrum consists of a finite number of discrete frequencies, i.e., not a continuous distribution, it is called a line spectrum.

If the spectrum has a continuous distribution of frequencies, it is called a continuous spectrum. Those components which lie between any two specified frequencies, f_a and f_b, are said to occupy a frequency band. The difference between the two frequencies, $f_b - f_a$, is called the bandwidth. The ratio of the two frequencies, f_b/f_a is called the band ratio. A band ratio of $f_b/f_a = 2$ is called an octave. A band ratio of $f_b/f_a = 10$ is called a denary band. (It is frequently called a decade, although a decade is properly defined as a band whose width is $f_b - f_a = 10$ cps.) A band ratio of $f_b/f_a = \sqrt[3]{2}$ is called a one-third octave band.

A more precise way of defining a continuous spectrum is as follows. Suppose we had a hypothetical idea filter whose bandwidth we could control as we desired. Consider first a line spectrum and begin with a bandwidth large enough to encompass several of the discrete frequencies; then reduce the bandwidth gradually. As soon as one of the frequencies is filtered out, we notice a sudden drop in the transmitted power. As this process is continued, the frequencies will be filtered out one at a time until, when the last one disappears, the power will suddenly drop to zero. On the other hand, if the nature of the spectrum is such that as we reduce the bandwidth the transmitted power decreases continuously,

so that as long as the bandwidth is finite, no matter how small, a finite amount of power is transmitted, the spectrum is a continuous spectrum.

2. Power per Unit Band; Intensity per Unit Band

Let P_{ab} be the power transmitted in the frequency band $f_b - f_a$. Then the power per unit band is

$$\text{Power per Unit Band} = U = \lim_{f_b - f_a \to 0} \frac{P_{ab}}{f_b - f_a} = \frac{dP}{df} \tag{184}$$

A curve of power per unit band plotted against frequency for any given wave is called the spectrum characteristic of the wave.

Similarly, if I_{ab} is the intensity in the frequency band $f_b - f_a$, the intensity per unit band is

$$\text{Intensity per Unit Band} = J = \lim_{f_b - f_a \to 0} \frac{I_{ab}}{f_b - f_a} = \frac{dI}{df} \tag{185}$$

The intensity per unit band may be expressed in another form. Recall that the intensity is the power transmitted per unit area normal to the direction of propagation. The relationship between power and intensity is therefore

$$I = \lim_{A \to o} \frac{P}{A} = \frac{dP}{dA} \tag{186}$$

where A is the area, and therefore

$$J = \frac{dU}{dA} \tag{187}$$

The term "intensity per unit band" is seldom used in underwater acoustics. It is usually referred to as the "intensity in a 1 cps band." Although from a practical point of view these terms are considered equivalent, and nobody worries about it, they are fundamentally different. The intensity in any band, whether 1 cps band or otherwise, has the dimensions of intensity, which is different from intensity per cps. Intensity is related to intensity per unit band in the same way as displacement is related to velocity. Also, when the intensity per unit band varies with frequency, there may be a difference in the numerical values of the two quantities.

3. Note on Units: For practical measurements the unit of energy is the joule and the unit of power is the watt. These units are defined as follows:

$$1 \text{ joule} = 10^7 \text{ erg}$$

$$1 \text{ watt} = 10^7 \text{ erg/sec.}$$

These various quantities are summarized in the following table:

Quantity	Symbol Relationship	Practical Units
Energy	W	joule
Time	t	sec.
Area	A	cm^2
Frequency	f	cps
Power	$P = dW/dt$	watt
Intensity	$I = dP/dA$	$watt/cm^2$
Power Unit Band	$U = dP/df$	watt/cps = joule/cyc
Intensity Unit Band	$J = dI/df = dU/dA$	$watt/cm^2$ cps = $joule/cm^2$ cyc

The concept of "per unit band" applies only to a continuous spectrum. In the case of a pure sine wave, a finite amount of power is transmitted at a single discrete frequency. In this case the power per unit band is infinite, the bandwidth is zero, and the product of the two is finite.

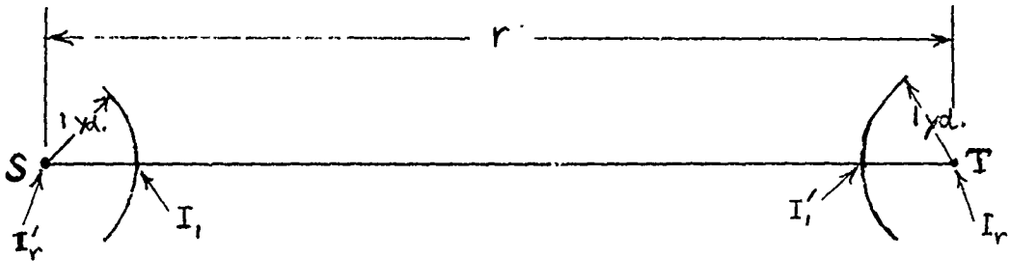
TECHNOLOGY OF UNDERWATER SOUND
(REVISED NOTES)

Relative Magnitudes and Transmission Loss

Note: All pressures referred to in this section are to be interpreted as rms values and all intensities as average values.

1. Introduction

To introduce the subject, let us consider the process involved in echo-ranging. Sound waves are generated by a transducer at S. Spreading out from S, they travel to a target located at T, at a distance r away. Some of this sound energy is scattered by the target. It spreads out from T and a part of it travels back to the transducer at S.



In the conventional sonar equation the transmission of energy is described in terms of the intensities at certain key points along the route, beginning with the so-called source intensity, or index intensity, at a standard distance of one yard from the source. These are as follows:

I_1 = intensity of outgoing pulse at 1 yard from source

I_r = intensity of outgoing pulse at target

I_1' = intensity of reflected pulse at 1 yard from target

I_r' = intensity of reflected pulse received at transducer (echo intensity)

To compute the echo intensity we begin with the source intensity and progressively multiply by intensity ratios

$$I_1 \cdot \left(\frac{I_r}{I_1}\right) \cdot \left(\frac{I_1'}{I_r}\right) \cdot \left(\frac{I_r'}{I_1'}\right) = I_r' \qquad (201)$$

Let us illustrate with some typical numerical values.

$$6.3 \times 10^{-3} \cdot 4.0 \times 10^{-8} \cdot 5.0 \times 10 \cdot 4.0 \times 10^{-8} = 5.0 \times 10^{-16} \qquad (201a)$$
$$\text{watt/cm}^2 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{watt/cm}^2$$

Numerical computations could be greatly simplified if we were to express (201) in such a form that the answer is obtained by adding terms rather than multiplying factors. This is the property of logarithms, since

$$\log a\, b = \log a + \log b$$

Strictly speaking, however, whenever we are working with physical quantities, we must always take logarithms of dimensionless ratios. Examination of (201) or (201a) shows that we are starting and ending with actual values of intensity —— I_1 watt/cm^2 and I_r' watt/cm^2. The transmission processes between the two are expressed as dimensionless ratios. We can express I_1 and I_r' also as dimensionless ratios if we select some arbitrary value of intensity I_o, as for example

$$I_o = 1 \text{ watt/cm}^2$$

and divide (201) by this. Thus

$$\left(\frac{I_1}{I_o}\right) \cdot \left(\frac{I_r}{I_1}\right) \cdot \left(\frac{I_1'}{I_r}\right) \cdot \left(\frac{I_r'}{I_1'}\right) = \left(\frac{I_r'}{I_o}\right) \qquad (202)$$

We may now take the logarithm of (202). However, instead of working with the logarithm directly, it has become standard practice to multiply the logarithm by 10. Common logarithms (to the base of 10) are used for this

purpose. Equation (202) and its numerical example thus become

$$10 \log \frac{I_1}{I_o} + 10 \log \frac{I_r}{I_1} + 10 \log \frac{I'_1}{I_r} + 10 \log \frac{I_r'}{I'_1} = 10 \log \frac{I'_r}{I_o} \qquad (203)$$

and (approximately)

$$-22 \quad -74 \quad +17 \quad -74 = -153 \qquad (203a)$$

The units in which equations (202) and (202a) are expressed are called decibels. We shall have more to say about this in the following paragraphs.

2. Relative Magnitudes

The transmission of power in acoustic systems, as illustrated by the previous example, is expressed in terms of ratios of physical quantities. Once the actual value of the power (or intensity) at one point in the system is specified, the actual values at other points can be obtained by ratios. Such ratios are called relative magnitudes. The actual values of the quantities are called absolute magnitudes.

3. Transmission Loss - The Decibel

The unit employed by electrical engineers for measuring the transmission loss in a power transmission system is the decibel. If P_{in} is the input power to any given part of the system and P_{out} is the output power from the same part, the transmission factor, which is really a measure of the efficiency, is P_{out}/P_{in}. The transmission loss, N, expressed in decibels, is defined in terms of logarithms to the base 10 as follows:

$$N = 10 \log \frac{1}{P_{out}/P_{in}} = 10 \log \frac{P_{in}}{P_{out}} = -10 \log \frac{P_{out}}{P_{in}} \qquad (204)$$

Note that to obtain the transmission loss we take the logarithm of the

reciprocal of the transmission factor. A positive transmission loss means that the output power is less than the input power.

The ratio r corresponding to one decibel is

$$10 \log r = 1 \text{ decibel (db)}$$

or

$$r = 10^{0.1} = 1.2589 \qquad (205)$$

The relationship between ratios and decibels is illustrated in the following table:

Ratio	d_b	d_b	Ratio
1	0	0	1
2	3.01	3	1.995
3	4.77	6	3.981
4	6.02	10	10.
5	6.99	13	19.95
10	10.	16	39.81
20	13.01	20	100.
50	16.99	23	199.5
100	20.	26	398.1
200	23.01	30	1000.
etc.		etc.	

Although it is convenient to express any ratio in terms of 10 times the logarithm, the name decibel is properly applied only to ratios of energy, power, and related quantities, such as intensity, power per unit band, and intensity per unit band. In certain specific cases this application can be extended somewhat. For example, the acoustic intensity can be expressed as $p^2/\rho c$. The ratio of two intensities I_1 and I_2 is

$$\frac{I_2}{I_1} = \frac{p_2^2/\rho_2 c_2}{p_1^2/\rho_1 c_1}$$

So long as we are dealing with sea water where changes in ρ and c are relatively small, the specific acoustic impedance may, for intensity computations,

be considered to be constant. Therefore, within this approximation,

$$\frac{I_2}{I_1} = \frac{p_2^2}{p_1^2} \tag{206}$$

so that

$$N = 10 \log \frac{I_2}{I_1} = 10 \log \frac{p_2^2}{p_1^2}$$

or

$$N = 20 \log \frac{p_2}{p_1} \text{ db} \tag{207}$$

Remember that when we wish to express a transmission loss in terms of a pressure ratio, we must use 20 log (pressure ratio), because the intensity is proportional to the square of the pressure. Remember also that these formulas do not apply to the transmission of sound across boundary between two media having different values of ρc, such as water and air.

It should be noted that the concept of the decibel has physiological significance, because the response of the eye and ear to changes in light and sound intensity is logarithmic.

There is no generally accepted terminology for expressing 10 log of ratios in general. Dr. J. W. Horton in his book, Fundamentals of Sonar, suggests the name logit. For example, suppose we wished to consider the ratio of two frequencies, $f_1 = 20,000$ cps and $f_2 = 1000$ cps. Then we would say

$$10 \log f_1/f_2 = 10 \log \frac{20,000}{1,000} = 10 \log 20$$

$$= 13 \text{ frequency logits}$$

Similarly, for the ratio of two velocities, $v_1 = 30$ knots and $v_2 = 10$ knots, we would say

$$10 \log \frac{v_2}{v_1} = 10 \log \frac{30}{10} = 4.8 \text{ velocity logits}$$

In these notes we shall rarely use the term logit. We shall from time to time have occasion to work with logarithms of bandwidth (frequency) ratios, and since these will be associated with intensity computations, we shall follow the common practice of calling them decibels.

4. Transmission Level

The concept of transmission loss is associated only with a ratio of transmitted powers; it says nothing about the absolute magnitude of either power. If we wish to specify the actual power at any point in the system, we must select some arbitrary reference value and work with the ratio of the actual power to the reference power. For example, suppose the power at some point in an electrical transmission system is P watts and the reference value is P_o watts. Then the actual value of P is expressed in terms of a ratio in the following manner

$$P = \left(\frac{P}{P_o}\right) P_o \quad \text{watts} \tag{208}$$

Equation (208) says that the actual power P is greater (or less) than the reference value P_o by a factor (P/P_o). When the power at this point is expressed in decibel notation, it is called the transmission level or power level and is written as follows:

$$L = 10 \log \frac{P}{P_o} \ \text{db//}P_o \ \text{watts} \tag{209}$$

where L is the power level corresponding to P watts. We use a double slant line in this application to indicate that the symbol following it denotes the reference value on which the ratio is based.

Similarly, if we wish to express an intensity I in watts/cm^2 relative to a reference intensity I_o watts/cm^2, we write

$$I = (\frac{I}{I_o})I_o \quad \text{watts/cm}^2 \tag{210}$$

and

$$L = 10 \log \frac{I}{I_o} \text{ db//}I_o \quad \text{watts/cm}^2 \tag{211}$$

In this case the level is called an intensity level.

To illustrate this concept, let us return to the example in section 4 above. Note that in (201) and (201a) the index intensity I_1 and the echo intensity I_r' appear as absolute values, whereas the other factors appear only as intensity ratios. In order to express this equation in decibel form, we first divided both sides by the reference intensity I_o watts/cm^2, as shown by (202), and then took logarithms, obtaining (203). The numerical values shown in (203a) are based on I_o = 1 watt/cm^2.

The first term is an intensity level

$$L_1 = 10 \log \frac{I_1}{I_o} = -22 \text{db//watt/cm}^2 *$$

The other terms on the left are pure loss terms and are expressed simply as db (no reference). Thus, for the outgoing wave the transmission loss is

$$N = -10 \log \frac{I_r}{I_1} = -10 \log (4 \times 10^{-8}) = +74 \text{ db}$$

(Note that in the sonar equation the loss is subtracted from the source level.) Finally, the term on the right side of the equation is an intensity level

$$L_r' = 10 \log (\frac{I_r'}{I_o}) = -153 \text{ db//watt/cm}^2$$

Note also that in an algebraic equation such as (201a) the units must balance; watts/cm^2 on the left must equal watts/cm^2 on the right. Similarly, in the decibel equation (203a) the units of the reference must balance. There

* When the number of units is not specified for the reference, it is understood to be one, just as in algebra x stands for 1x. Thus, db//watt/cm^2 signifies db//1 watt/cm^2.

is one term on the left with a reference in watts/cm², and one term on the right with a similar reference. All other terms are purely relative. In other words, a level must equal a level.

In the above example we may find the level at any intermediate point in the system. For example, at the point where the outgoing wave strikes the target the intensity is

$$I_r = 6.3 \times 10^{-3} \cdot 4.0 \times 10^{-8} = 2.5 \times 10^{-10} \text{ watt/cm}^2$$

and the intensity level is

$$L_r = 10 \log \frac{I_r}{I_o} = 10 \log \frac{2.5 \times 10^{-10}}{1} = -96 \text{ db//1 watt/cm}^2$$

This same result can be obtained more simply by starting with the source level and subtracting the transmission loss. Thus

$$L_r = -22 \text{ db//1 watt/cm}^2 - 74 \text{ db}$$

$$= -96 \text{ db//1 watt/cm}^2$$

In fact, the intensity level at any point in the system can be obtained by starting with the source level and progressively subtracting losses up to that point.

5. Sound Pressure Level

Most acoustic measurements are made with pressure-sensitive hydrophones and are expressed units of pressure (microbars). For this reason it has become the usual practice in underwater acoustics to refer levels to a pressure of one microbar (although some measuring instruments are calibrated in terms of 0.0002 microbar). If we let p_o represent the reference pressure, then from (207) we see that the level is expressed as

$$L = 20 \log \frac{p}{p_o} \text{ db}//p_o \text{ μb} \qquad (212)$$

A level expressed in this manner is called a sound pressure level. For a single plane wave traveling in a single direction, sound pressure level and intensity level are equivalent and can be converted from one to the other by adding or subtracting the proper number of decibels. In more complicated noise fields, as has been pointed out earlier, the pressure measured by a hydrophone is in many instances the resultant pressure due to many waves traveling in different directions, and the intensity $p^2/\rho c$ computed from this pressure does not represent an actual energy flow in any one direction. It is termed an "equivalent" intensity (equivalent to the intensity of a single hypothetical wave having the same pressure). In such a sound field the relation between the sound pressure level and the equivalent intensity level is to be interpreted in the same manner.

It is interesting to note that in air acoustics the reference used for intensity levels is 10^{-10} microwatt/cm^2 (or 10^{-9} erg/cm^{2}sec), while the reference for sound pressure level is 0.0002 dyne/cm^2 (or microbar). Since the specific acoustic impedance of air under standard conditions is $\rho c = 42$ specific acoustic ohms, the intensity corresponding to this pressure is

$$\frac{p^2}{\rho c} = \frac{(.0002)^2}{42} = 0.95 \times 10^{-9} \text{ erg/cm}^2 \text{ sec}$$

The two references are thus almost identical.

In underwater acoustics, levels are almost never referred to intensity units. They are commonly expressed as sound pressure levels, usually referred to a pressure of one microbar, and occasionally 0.0002 microbar.

The latter is rather widely used in reporting radiated noise measurements of ships, apparently because the measuring equipment is calibrated that way. The .0002 μb reference is a carry-over from air acoustics.

6. Note Regarding the Pressure Reference

Although the practice of referring transmission levels to pressure units is widely used, it leads to some confusion -- or at least to some pedagogical problems -- when extended to the concept of spectrum level (which will be discussed below). It turns out that the pressure reference for spectrum level is applied to a physical quantity which is not pressure at all, but something different. In the discussion which follows we shall point out this difficulty and shall overcome it by using a slightly different notation which, although more complicated, reflects more accurately the physical concepts involved.

Since the actual energy flow in an acoustic wave is proportional to the square of the pressure, it is more logical to work in units of p^2 than p, and equation (212) is more logically expressed in the form

$$L = 10 \log \frac{p^2}{p_o^2} \quad db//p_o^2\ \mu b^2 \qquad (213)$$

For a reference pressure of 1 μb, we are expressing intensity levels in

decibels referred to 1 μb^2

The only difference thus far is that we use 1 μb^2 in place of 1 μb. No other change is involved. The number of decibels is the same.

7. Change of Reference

When the same level (i.e., the level corresponding to the same actual intensity) is expressed in decibels referred to different references, it will have different numerical values. To change from one level to another we must add or subtract a constant.

Let us consider the change from 1 μb to 0.0002 μb. Suppose the pressure is 1 μb. Then the level referred to 1 μb^2 is

$$L_{1\ \mu b} = 10 \log \frac{1}{1} = 0 \text{ db//1 } \mu b^2 \tag{214}$$

and the level referred to $(0.0002\ \mu b)^2$ is

$$L_{.0002\ \mu b} = 10 \log \frac{1}{(.0002)^2}$$

$$= 74 \text{ db//}(.0002\ \mu b)^2 \tag{215}$$

From this we see that <u>levels expressed in db//$(.0002\ \mu b)^2$ are 74 db higher than levels expressed in db//1 μb^2</u>. (The 1 in front of μb^2 may be ignored.) A level of, say, 10 db//μb^2 would be equivalent to a level of 84 db//$(.0002\ \mu b)^2$.

Next, let us consider the change from 1 μb^2 to 1 watt/cm^2. In sea water with a specific acoustic impedance of 154,000 ohms the intensity corresponding to one microbar is

$$I = \frac{p^2}{\rho c} = \frac{1}{154,000} = 6.5 \times 10^{-6} \text{ erg/cm}^2 \text{ sec} \tag{216}$$

$$= 6.5 \times 10^{-13} \text{ watt/cm}^2 \tag{216a}$$

The reference intensity is

$$I_o = 1 \text{ watt/cm}^2$$

so that the level corresponding to 0 db//μb^2 is

$$L_{w/cm^2} = 10 \log \frac{I}{I_o} = 10 \log (6.5 \times 10^{-13})$$

$$= -121.9 \text{ db//watt/cm}^2 \tag{217}$$

Levels referred to 1 watt/cm^2 are approximately 122 db lower than levels referred to 1 μb^2.

8. Spectrum Level; Band Level

The energy distribution in a continuous spectrum is described in terms of the intensity per unit band, J, which we have defined as the rate of change of intensity with frequency

$$J = \frac{dI}{df} \tag{219}$$

In terms of watts, J has the dimensions watt/cm^2 cps (or its equivalent, joule/cm^2 cycle). The decibel equivalent of intensity per unit band is called the spectrum level, L_S. It is defined by the relation

$$L_S = 10 \log \frac{J}{J_o} \quad db//J_o \tag{220}$$

If J is expressed in the units indicated above, the logical reference value is

$$J_o = 1 \text{ watt/cm}^2 \text{ cps}$$

To derive the relation between spectrum level and intensity level, we see that the amount of energy being propagated depends upon the range of frequencies involved, that is, upon the bandwidth. The intensity I in a band of frequencies from f_a to f_b is obtained by integrating J. Thus

$$I = \int_{f_a}^{f_b} J \, df \tag{221}$$

For the present let us make the simplifying assumption that J is constant, so that

$$I = J (f_b - f_a) = JW \tag{222}$$

where

$$W = f_b - f_a = \text{the bandwidth}$$

If J is not constant, we could use an average value over the band in (222). The intensity in a band is thus the product of the intensity per unit band and the bandwidth. The reference quantities are related by the equation

$$I_o = J_o W_o$$

If we select

$$I_o = 1 \text{ watt/cm}^2$$

and

$$J_o = 1 \text{ watt/cm}^2 \text{ cps}$$

we see that

$$W_o = 1 \text{ cps}$$

A dimensionless equation is obtained by dividing by the reference values

$$\frac{I}{I_o} = \frac{J}{J_o} \cdot \frac{W}{W_o}$$

The corresponding levels are

$$10 \log \frac{I}{I_o} = 10 \log \frac{J}{J_o} + 10 \log \frac{W}{W_o}$$

or, since W_o has a numerical value of 1,

$$L = L_s + 10 \log W \qquad (223)$$

In the above expression L is the band level, or intensity level in the band of width W. Strictly speaking, the term 10 log W should be expressed in frequency logits//1 cps, but it is usually expressed simply in decibels.

In the preceding discussion we have defined the various physical quantities in terms of watts and have carefully avoided the use of microbars. We must now return to the "real world." By virtue of the relation

$$I = \frac{p^2}{\rho c}$$

the intensity in any band is proportional to the square of the pressure in that

band. This is true for any number of small bands which make up a large band. If we were to measure the pressures p_1, p_2, p_3 ..., in each of the small bands, and the resultant pressure p in the large band, we should find that

$$p^2 = p_1^2 + p_2^2 + p_3^2 + \ldots$$

If we were to break up the large band into more and more bands of smaller and smaller bandwidth, we should ultimately arrive at the concept of (pressure)^2 per unit bandwidth, and this quantity would be proportional to the intensity per unit bandwidth, as indicated by the following equation.

$$J = \frac{\text{(pressure)}^2 \text{ per unit band}}{\rho c} \qquad (224)$$

Unfortunately this quantity does not appear to have any official recognition and no symbol exists for it, but nevertheless it is the fundamental basis upon which spectrum levels (when expressed in terms of microbars) rest. For pressures measured in microbars the units are of press^2/unit band μb^2/cps. Thus, the pressure equivalent of equation (222) is

$$p^2 = \left[\text{(pressure)}^2 \text{ per unit band}\right] \cdot W \qquad (225)$$

and the units are

$$\mu b^2 = (\mu b^2/\text{cps}) \cdot \text{cps}$$

On the basis of the above discussion, we see that intensity levels should be expressed in db//μb^2 and spectrum levels in db//μb^2/cps, and they are related to each other by equation (223), which is reproduced below with the reference specified.

$$L \text{ db//}\mu b^2 = L_s \text{ db//}\mu b^2/\text{cps} + 10 \log W \qquad (223a)$$

In most of the literature on underwater acoustics both intensity levels and spectrum levels are expressed simply as db//μb. Although the meaning of these symbols is understood by workers in the field, and the objection to them is probably not serious enough to justify a change, they are nevertheless sloppy and can be confusing to newcomers. It is encouraging to see the more precise notation being used by a few writers.

The same conversions apply to spectrum level as to intensity level, namely

$$L_S \text{ db}//(.0002\ \mu\text{b})^2/\text{cps} = L_S \text{ db}//\mu\text{b}^2/\text{cps} + 74 \text{ db} \tag{226}$$

$$L_S \text{ db}//\text{watt}/\text{cm}^2 \text{ cps} = L_S \text{ db}//\mu\text{b}^2/\text{cps} - 121.9 \text{ db} \tag{227}$$

9. Band Level Correction

In the preceding discussion of spectrum level and band level we made the assumption that the intensity per unit band, J, is constant over the band or, if not, we assumed we knew the average value. In real noise fields J is not constant, nor is its average value accurately known. The normal practice is to use the value J_m at the geometric mean of the upper and lower frequencies of the band; that is, at the frequency f_m such that

$$f_m = \sqrt{f_a f_b} \tag{228}$$

where b and a refer to the upper and lower frequencies, respectively. One does not use the actual J, of course, but the corresponding spectrum level. For example, if in equation (223) we wished to compute the intensity level in a band from 1 to 3 kc, we should use the spectrum level at a frequency of

$$f_m = \sqrt{1 \times 3} = 1.73 \text{ kc}$$

If the spectrum level at this frequency were -35 db//μb^2/cps, the intensity level in the band would be

$$L = -35 + 10 \log 2000 = -2 \text{ db//}\mu b^2$$

Although this procedure is sufficiently accurate for most practical applications, it is in general only an approximation, as we shall now show. Over most bands of practical interest, J can be considered as varying with some constant power, n, of the frequency. Ambient sea noise, for instance, varies as the -1.7 power over a wide range of frequencies. Let J_a, J_b, and J_m denote the values of J at frequencies f_a, f_b, and f_m. The relation between J and f may then be expressed in the form

$$\frac{J}{J_a} = \left(\frac{f}{f_a}\right)^n \tag{229}$$

The average value of J over the band from f_a to f_b is

$$\overline{J} = \frac{J_a}{(f_b - f_a) f_a^{\,n}} \int_{f_a}^{f_b} f^n \, df$$

which integrates to

$$\overline{J} = \frac{J_a (f_b^{\,n+1} - f_a^{\,n+1})}{(n+1)(f_b - f_a) f_a^{\,n}} \quad \text{if } n \neq -1$$

or

$$\overline{J} = \frac{J_a f_a}{f_b - f_a} \log_e \left(\frac{f_b}{f_a}\right) \quad \text{if } n = -1$$

The value of J_m is found by letting $f = f_m$ in (229)

$$\frac{J_m}{J_a} = \frac{f_m^{\,n}}{f_a^{\,n}} \tag{230}$$

To simplify the writing, let

$$K = \sqrt{f_b/f_a} \tag{231}$$

so that

$$f_b = K f_m \tag{232}$$

and

$$f_a = f_m/K \tag{233}$$

Substituting (230), (232) and (233) into the expressions for $\overline{J}$, we obtain

$$\overline{J} = J_m \left[\frac{K}{K^2 - 1} \cdot \frac{K^{2(n+1)} - 1}{(n+1)K^{n+1}} \right] \quad \text{if } n \neq -1 \tag{234}$$

or

$$\overline{J} = J_m \left[\frac{K}{K^2 - 1} \cdot 2 \log_e K \right] \quad \text{if } n = -1 \tag{235}$$

The factor in brackets represents the correction which must be applied to J_m to obtain the correct average value $\overline{J}$. Let this factor be called $F(K, n)$, and let the spectrum level at the mean frequency f_m be called L_m. Then the band level equation (223) becomes

$$L = L_m + 10 \log W + 10 \log F(K, n) \tag{236}$$

The last term, $10 \log F(K, n)$, is called the band level correction. The correction is zero when

$$n = 0 \text{ or } -2$$

as may be verified by substitution in (234). The condition $n = 0$, of course, corresponds to J = constant. The correction function $10 \log F(K, n)$ has been plotted by Horton in Fundamentals of Sonar. The value of n for most noise sources lies between 0 and -3, and if the bandwidth does not exceed an octave or so, the band level correction amounts to less than one decibel.

10. Source Level; Index Level

In underwater acoustic systems the magnitude of the acoustic output of a sound source is specified in terms of the intensity at a standard distance of one yard from the source. This is known as the source intensity, or index intensity of the source. The corresponding transmission level is called the source level or index level.

For the present we shall limit the discussion to omnidirectional sources, that is, sources which radiate equally in all directions, so that the intensity has a uniform value over the entire surface of any small sphere surrounding the source. Let P denote the total power in watts radiated by the source, I_1 the source intensity in watts/cm^2, and r_1 the standard 1 yard reference distance, expressed in cm. Then if the source is a point source (or if its dimensions are very much less than one yard), the intensity is equal to the ratio of the power to the area of a sphere of radius r_1 (see equation (145)). Thus

$$I_1 = \frac{P}{4\pi r_1^2} \tag{237}$$

Substituting the numerical value of r_1, i.e.,

$$r_1 = 1 \text{ yd} = 91.44 \text{ cm},$$

into (237) yields for the source intensity

$$I_1 = \frac{P}{4\pi\,(91.44)^2} = 0.952 \times 10^{-5} P \text{ watt/cm}^2$$

The source level S_o for an omnidirectional source is

$$S_o = 10 \log \frac{I_1}{I_o} \qquad (I_o = 1 \text{ watt/cm}^2)$$

$$S_o = 10 \log P - 50.2 \qquad \text{db//watt cm}^2 \tag{238}$$

To convert to db//μb^2 we add 121.9 db, obtaining

$$S_o = 10 \log P + 71.7 \text{ db//}\mu b^2 \tag{239}$$

The above formulas for S_o apply only to omnidirectional sources. The modifications required for directional sources will be derived later.

Even though these results were derived for point sources, they may be applied to extended sources as well, so long as their significance is understood. The concept of source level is concerned with the process of describing the flow of energy from a source to a distant receiver. At distances of many hundreds or thousands of yards, the waves from even a large source appear to be spreading out from a point. The standard reference distance is to be interpreted as being one yard from this apparent point source.

Note on the Bandwidth of a 1/3 Octave Band

The bandwidth of any band may be expressed in terms of the geometric mean frequency f_m and the ratio K by subtracting (233) from (232).

$$W = f_b - f_a = (K - \frac{1}{K}) f_m \tag{240}$$

The value of K for a 1/3 octave band is $2^{1/6}$ and the bandwidth is

$$W = (2^{1/3} - 2^{-1/3}) f_m = 0.2316 f_m \tag{241}$$

or

$$10 \log W = 10 \log f_m - 6.4 \text{ db} \tag{242}$$

The Propagation of Underwater Sound

A. Propagation Loss

1. Definition

The flow of acoustic energy from a source to a receiver is described in terms of the intensity at a standard distance of one yard from the source, and the reduction in intensity between this point and the receiver. The transmission level corresponding to the source intensity is called the source level and has been discussed earlier. The transmission loss corresponding to the reduction in intensity between the reference point and the receiver is called the propagation loss.

If I_1 denotes the source intensity and I the intensity at any distant point, the propagation loss to that point is

$$N_w = 10 \log \frac{I_1}{I} \qquad (301)$$

The level L at the distant point is obtained by subtracting the propagation loss from the source level S_o,

$$L = S_o - N_w \qquad (302)$$

It should be noted that the choice of the standard reference distance is purely arbitrary and in no way affects the value of the intensity level at the distant point. The choice of one yard is convenient because ranges are customarily measured in yards (or thousands of yards). If a different reference distance were chosen, the numerical values of both the source level and the propagation loss would be changed by the same amount. For example, if a distance of 10 yards were selected for the reference, the intensity at 10 yards would be 1/100 as large as the intensity at 1 yard (inverse square law). The source level would thus be 20 db smaller, but the propagation loss would also be 20 db smaller, so that the difference

between the two, which gives the level at the distant point, would remain unchanged. A reference distance of 1 meter is sometimes used in scientific work.

2. General Discussion

The propagation loss is affected by a number of factors, of which one of the most important is the spreading loss. As the waves spread out from a source, the surface across which the energy is being propagated becomes larger and larger. Even if no energy is lost by absorption or scattering, the energy becomes diluted as it spreads over larger and larger areas, and the intensity is correspondingly reduced. Near a source the spreading is spherical and the intensity is proportional to the inverse square of the distance. At larger distances the spreading is affected by refraction, which causes the rays, or paths along which the waves travel, to curve. Refraction is capable of producing drastic changes in the spreading loss, including focusing effects and shadow zones. This phenomenon will be discussed in a later section.

A second factor affecting propagation loss is the attenuation loss. Attenuation usually refers to a gradual loss of energy from the wave as it moves through the medium. Attenuation is caused both by absorption and scattering. Absorption refers to the process of converting acoustic energy into heat and is associated largely with frictional effects. Scattering, as the name suggests, refers to the process whereby objects in the water (bubbles, marine life, or other foreign particles) cause a portion of the energy in a wave to be deflected in various directions in the form of incoherent radiation and thereby rendered useless. The processes included under the heading of attenuation normally possess the characteristic that the rate of loss per unit distance traveled is proportional to the amount of intensity present, and this leads to the familiar exponential decay law. The attenuation loss is dependent upon the acoustic frequency and is very severe at high frequencies.

Propagation loss is also affected by reflection from the surface and bottom of the ocean. This process is more correctly described as scattering, since when a sound wave impinges on either of these boundaries, a certain amount of energy is scattered in all directions. Usually, however, the greater part of the scattered energy is scattered forward in the direction of specular reflection (angle of reflection equal to angle of incidence). Multiple reflections, especially in shallow water, tend to produce cylindrical spreading at ranges appreciably larger than the water depth.

There is a propagation loss associated with each reflection from the ocean boundaries. This loss is due in part to the scattering of sound in other directions and in part to transmission of a portion of the energy into the adjoining medium. The reflection loss at the surface, which is due almost entirely to scattering, is influenced by the roughness of the surface and is therefore a function of the sea state. It tends to be rather small, of the order of 1 db or so under most conditions. The bottom loss is strongly affected by the nature of the bottom, the angle of incidence, and the acoustic frequency. Hard bottoms, such as sand and rock, are good reflectors. Soft mud is a poor reflector. Roughness causes losses due to increased scattering. Bottom losses are usually small at angles of incidence near grazing but increase rapidly at steeper angles. Beyond 20 to 30 degrees they tend to level off. The loss increases with increasing frequency. Bottom losses can be quite severe - sometimes as much as 20 db.

The reflection loss at a boundary causes a discrete jump in the propagation loss. It is a lump-sum item which is added to the propagation loss each time a reflection occurs. However, in shallow water and in long-range transmission of low frequency sound, where a large number of reflections occur, the discrete losses are practically equivalent to a continuous attenuation process and are often treated as such.

3. Spreading Loss

In this paragraph we shall consider only the ideal cases of spherical and cylindrical spreading, postponing a discussion of the effects of refraction until a later section. For spherical waves the intensity is inversely proportional to the square of the distance from the source. Let I_1 represent the source intensity at the standard distance r_1 (= 1 yard), and let I represent the intensity at any distance r (normally hundreds or thousands of yards).

Then by the inverse square law the intensity ratio is

$$\frac{I}{I_1} = \frac{r_1^2}{r^2} \tag{303}$$

and the spreading loss is

$$N_{spr} = 10 \log \frac{I_1}{I} = 20 \log \frac{r}{r_1}$$

If r is measured in kiloyards (thousands of yards), then

$$r_1 = 10^{-3} \text{ kyd}$$

and

$$N_{spr} = 60 + 20 \log r \quad \text{db} \tag{304}$$

Note that spreading loss (and propagation loss as well) involves merely a ratio of two quantities and is therefore expressed simply in decibels without any reference.

Under certain conditions the propagation of sound in the ocean can be represented fairly well with a model in which the spreading is spherical out to some range r_2, and cylindrical beyond. For a cylindrical wave the intensity varies inversely as the first power of the distance. If I_2 is the intensity at the range r_2 and I is the intensity at the range r ($> r_2$), then

$$\frac{I_2}{I_1} = \frac{r_1^2}{r_2^2} \quad \text{and} \quad \frac{I}{I_2} = \frac{r_2}{r}$$

so that

$$\frac{I}{I_1} = \frac{I_2}{I_1} \cdot \frac{I}{I_2} = \frac{r_1^2}{r_2^2} \cdot \frac{r_2}{r} = \frac{r_1^2}{r_2 r}$$

and the spreading loss is

$$N_{spr} = 20 \log \frac{r_2}{r_1} + 10 \log \frac{r}{r_2}$$

or, if the ranges are expressed in kiloyards,

$$N_{spr} = 60 + 10 \log r_2 + 10 \log r \quad \text{db} \qquad (305)$$

4. Attenuation Loss

In the attenuation process the loss of intensity dI which occurs when the wave traverses a distance dx is proportional to dx and to the existing intensity I. For a plane wave traveling in the x direction the reduction is

$$dI = - \alpha' I \, dx$$

where α' is a constant of proportionality. This differential equation solves to

$$\log_e I = - \alpha' x + \text{constant}$$

The constant of integration may be evaluated if we know the intensity I_1 at a fixed point x_1. Thus

$$\log_e \frac{I}{I_1} = - \alpha' (x - x_1) \qquad (306)$$

or

$$\frac{I}{I_1} = e^{- \alpha' (x - x_1)} \qquad (307)$$

The intensity decays exponentially with the distance traveled.

The attenuation loss between x_1 and x is

$$N_{att} = 10 \log \frac{I_1}{I} \qquad (308)$$

where the logarithm is taken to the base 10. The change of base is accomplished by the well-known formula

$$\log_{10} \frac{I_1}{I} = (\log_{10} e)(\log_e \frac{I_1}{I}) = 0.434 \log_e \frac{I_1}{I} \qquad (309)$$

Therefore, combining (306), (308), and (309)

$$N_{att} = (10 \log_{10} e) \log_e \frac{I_1}{I}$$

$$= 4.34 \alpha' (x - x_1)$$

Let

$$\alpha = 10 \alpha' \log_{10} e = 4.34 \alpha' \qquad (310)$$

Then the attenuation loss is

$$N_{att} = \alpha (x - x_1) \qquad (311)$$

The coefficient α is called the attenuation coefficient. It is usually expressed in decibels per kiloyard, db/kyd. The distance x, of course, must be expressed in kyd.

Although (311) was derived for plane waves, it may be applied with negligible error to spherical or cylindrical waves in the ocean.

5. Combined Loss - Spreading plus Attenuation

Although we have treated spreading and attenuation separately, both effects are present together in the propagation of waves in the ocean. If the index intensity is I_1 at the standard reference distance r_1, the intensity of a spherically spreading wave at distance r is obtained by combining (303) and (307)

$$\frac{I}{I_1} = \frac{r_1^2}{r^2} e^{-\alpha' (r - r_1)} \qquad (312)$$

The corresponding propagation loss is

$$N_w = 10 \log \frac{I_1}{I}$$

$$N_w = 60 + 20 \log r + \alpha (r - r_1)$$

In the usual applications to underwater acoustic systems we are concerned with propagation over distances of the order of one or more kiloyards. Since the reference distance r_1 is only one yard or 0.001 kyd, it may usually be neglected in comparison with r. Therefore, for practical purposes the propagation loss for spherical spreading is

$$N_w = 60 + 20 \log r + \alpha r \qquad (313)$$

For the case of spherical spreading out to a range r_2 and cylindrical spreading beyond, the spreading loss is given by (305), and the combined loss is

$$N_w = 60 + 10 \log r_2 + 10 \log r + \alpha r \qquad (314)$$

6. The Measurement of Propagation Loss

Because the attenuation coefficient varies with the acoustic frequency, it is advisable to measure propagation loss using sinusoidal waves of a fixed frequency. To measure the propagation loss between two points it is necessary to measure the intensity level at both points. The level at the far point is measured with a calibrated hydrophone. The level at the one yard point may be determined either by calibrating the sound projector or by measuring with a calibrated hydrophone. When using the hydrophone, it is of course not practical to measure the pressure at one yard. Measurements are taken at some larger distance (still close enough that the spreading is spherical and the attenuation may be neglected) and the inverse square correction is applied.

7. Evaluation of the Propagation Loss Components

From a single measurement of propagation loss it is impossible to determine how much of the loss is due to spreading and how much to attenuation. In order to separate the one from the other it is necessary to take measurements at a number of different ranges. In the general case where the spreading varies as the inverse n^{th} power of the range (n = 2 for spherical spreading), the propagation loss is

$$N_w = 30 n + 10 n \log r + \alpha r \qquad (315)$$

The nature of the log function is such that it increases rapidly with r at short ranges, but slowly at long ranges. The 10 n log r term tends to be dominant at short ranges and the αr term at long ranges.

Therefore, if we plot N_W vs log r, we should expect to get a straight line having a slope of 10 n at short ranges. At longer ranges the attenuation term will begin to take over, causing the line to curve in the direction of higher losses. The spreading loss component can be determined from the slope of the straight-line portion of this curve.

On the other hand, if we plot N_W vs r, we should expect to get a straight line with slope α at the longer ranges. At shorter ranges the spreading loss term will begin to take over, causing the loss to decrease more rapidly than linear.

Typical curves showing these effects are given by J. W. Horton in Fundamentals of Sonar, pp 78, 79.

It should be pointed out that conditions in the ocean are seldom conducive to accurate separation of the two effects.

8. Values of the Attenuation Coefficient

It has been known for a long time that in pure water the attenuation of sound waves is due to viscous friction of the water molecules. Measurements have confirmed the theory developed by G. G. Stokes a century ago which predicts that the attenuation coefficient should vary as the square of the frequency. In sea water an additional effect is observed at low frequencies. This is due to the dissolved salts. When both effects are included, the attenuation coefficient can be expressed by an equation of the following form,

$$\alpha = \frac{A f^2}{B + f^2} + Cf^2 \qquad (316)$$

where f is in kc. The constants A, B, and C are functions of temperature such that α decreases with increasing temperature. The U. S. Navy Underwater Sound Laboratory, on the basis of extensive measurements in the Atlantic Ocean (Project AMOS), gives the following values:

$$\begin{aligned} A &= 0.651\, f_T \\ B &= f_T^2 \\ C &= 0.0269/f_T \end{aligned} \qquad (317)$$

where

$$f_T = 1.23 \times 10^{\left[6 - \frac{2100}{T + 459.6}\right]} \qquad (318)$$

T = temperature, °F.

For T = 50°F these equations lead to

$$A = 60.6; \quad B = 8760; \quad C = 0.000289 \qquad (319)$$

M. J. Sheehy and P. Halley found from measurements of sound generated by an underwater atomic explosion in the Pacific Ocean that the attenuation coefficient at very low frequencies from 16 to 250 cps can be expressed as

$$\alpha = 0.033\, f^{3/2} \quad \text{db/kyd} \qquad (320)$$

According to Horton, when losses due to reflections at the surface and bottom are included, the attenuation coefficient is considerably larger at low frequencies. He gives the following formula for the overall coefficient

$$\alpha = 0.20\, f + 0.00015\, f^2 \qquad (321)$$

The following table summarizes the value of α computed from the various formulas.

Attenuation Coefficient (db kyd)

f (kc)	AMOS Data, 50°F	Eq. (320)	Eq. (321)
0.1	.00007	.0010	.020
0.2	.00029	.0029	.040
0.5	.0018	.0117	.100
1.0	.0073	.033	.200
2.0	.0291		.401
3.0	.0655		.601
5.0	.1815		1.004
10.0	.720		2.01
20.0	2.79		4.06
30.0	5.96		6.14

The largest discrepancies occur at very low frequencies where the coefficient is very small. It is extremely difficult to measure a small coefficient, since the sound must travel very large distances before it is

attenuated appreciably, and in doing so it undergoes repeated reflections from the ocean boundaries. The 3/2 power law has been observed more recently by other observers and appears to represent reasonably well the attenuation of low frequency waves over long distances in the ocean. The AMOS values, on the other hand, represent the attenuation due only to the sea water itself and do not take into account any other effects.

An important conclusion that can be drawn from the above information is that in order to achieve large sonar detection ranges it is necessary to operate at low frequencies; otherwise the attenuation loss becomes prohibitive.

9. Fluctuations in Sound Level

Sound levels in the sea are subject to extremely large rapid fluctuations. It is thought that an important cause of these fluctuations is the interference between waves arriving by different paths. One bit of evidence for this conclusion is that when several frequencies are transmitted simultaneously over the same path, the fluctuations in the different waves occur at different times. Fluctuations have also been observed on longer time scales. Variations have been observed in the average level over a period of several minutes or even several hours. Values obtained for propagation loss are based on the statistical average of a large number of measurements.

10. Propagation Anomalies

It is convenient to consider spherical spreading plus normal attenuation as a nominal standard for computing propagation loss. Any deviations from this nominal standard which are observed in acoustic propagation in the sea are often spoken of as propagation anomalies. These may be of two types: spreading anomalies (i.e., other than inverse square spreading) and attenuation anomalies.

The Propagation of Underwater Sound (cont'd)

B. Refraction

1. The Ray Solution of the Wave Equation; Introduction

In the real ocean, where the speed of sound varies as a function of the space coordinates, solution of the basic differential equation - the "wave equation" -

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} \qquad (401)^*$$

generally involves insurmountable mathematical difficulties. There are, however, several techniques for obtaining very useful approximate solutions. One of the most valuable of such techniques is the method of rays. The concept of rays follows logically from the characteristics of wave solution. Families of rays are obtained as the solution of a simpler differential equation, called the eikonal equation, which in special cases is itself a solution of the wave equation and under rather broad conditions is a very good approximation to a solution. The relation of ray acoustics to wave acoustics is similar to the relation of geometric optics to physical optics.

2. Plane Waves and the Ray Concept

To introduce the subject we shall review the solution of the wave equation which we obtained previously in the special case where the speed of sound c is constant. The solution is of the form

$$p = F\left\{2\pi f\left[t - \frac{1}{c}(\alpha x + \beta y + \gamma z)\right]\right\} \qquad (402)$$

* This equation was designated previously as (112).

where the expression in braces is the argument of the function F and represents the phase of the wave, and α, β, γ are constants satisfying the relation

$$\alpha^2 + \beta^2 + \gamma^2 = 1 \tag{126}$$

It is clear that the wave has a constant phase for all values of x, y, z, and t satisfying the relation

$$t - \frac{1}{c}(\alpha x + \beta y + \gamma z) = \text{constant} \tag{403}$$

Focusing our attention on the wave at a fixed instant of time, we see that the plane

$$\alpha x + \beta y + \gamma z = s = \text{constant} \tag{404}$$

is the locus of points of <u>constant arrival time</u>. Such a surface is called a <u>wave front</u>. If we now watch the wave front move forward in time, we see from (403) that

$$\frac{d}{dt}(\alpha x + \beta y + \gamma z) = \frac{ds}{dt} = c \tag{405}$$

The wave moves forward with speed c. The direction in which the point (x, y, z) on the wave front moves is the direction of the gradient of s,

$$\begin{aligned} \text{grad } s &= \hat{\imath}\frac{\partial s}{\partial x} + \hat{\jmath}\frac{\partial s}{\partial y} + \hat{k}\frac{\partial s}{\partial z} \\ &= \hat{\imath}\alpha + \hat{\jmath}\beta + \hat{k}\gamma \end{aligned} \tag{406}$$

This is a unit vector defining a line whose direction cosines are α, β, γ, and which is therefore perpendicular to the plane defined by (404). The direction of propagation is therefore normal to the wave front.

We thus see that as time increases, any given point on a wave front moves with speed c along a line normal to the wave front. The line along which the point moves is called a ray. For the plane wave discussed above, the rays are parallel straight lines.

3. The Eikonal Equation

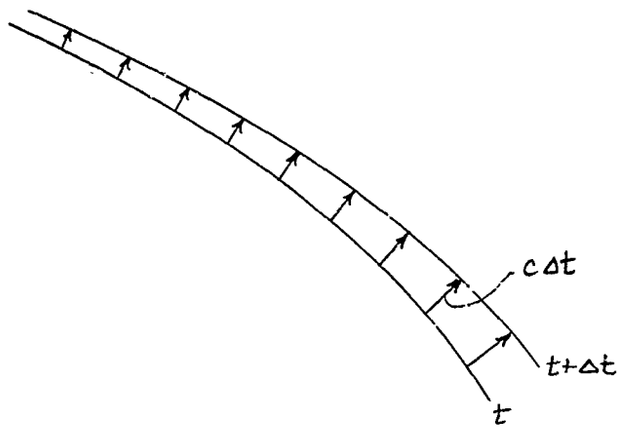
In the real ocean our basic assumption that c is a constant is not true. The speed of sound varies with the space coordinates, chiefly with the depth. When c is variable, equation (402) is no longer a solution of the wave equation. However, if the variations of c are small, then (402) should be a reasonable approximation to a solution. Let us assume this to be true and investigate its properties under the hypothesis that c is a function of the space coordinates x, y, z.

The immediate consequence is that α, β, γ are also variables and, in fact, are no longer the direction cosines of the normal to the wave front. The wave fronts are no longer planes but warped surfaces, and the rays are no longer straight lines but curved lines in space.

For mathematical convenience let us express the frequency f in terms of a constant reference value of the sound speed c_o and its associated wavelength λ_o ($\lambda_c f = c_o$), so that the phase of the wave is

$$\frac{2\pi}{\lambda_o}\left[c_o t - \frac{c_o}{c}(\alpha x + \beta y + \gamma z)\right]$$

A wave front is a surface on which the entire second term in the above expression has a constant value, and it propagates at speed c. Since c has different values at different locations along the wave front, some portions of the front move faster than others. A picture of a wave front at two instants of time close together might look as shown in the sketch.



The perpendicular distance between the two surfaces is c Δt.

Since the space function in the preceding expression for the phase angle has lost its simple interpretation as a plane, we may just as well designate it as a general function W(x, y, z) so that the phase angle may be written

$$\frac{2\pi}{\lambda_o}\left[c_o t - W(x, y, z)\right] \tag{407}$$

As any particular wave front moves along, the phase remains constant, so that the value of the function W increases linearly with time, i.e.,

$$W = c_o t + \text{constant}$$

or

$$\frac{dW}{dt} = c_o \tag{408}$$

Now, any point on the wave front, as we have seen, moves along a ray path at speed c, so that if s designates the distance traveled by this point, then

$$\frac{ds}{dt} = c \tag{409}$$

and hence

$$\frac{dW}{ds} = \frac{\frac{dW}{dt}}{\frac{ds}{dt}} = \frac{c_o}{c} \tag{410}$$

Equations (408) and (409) say that while the wave front is moving physically through space at speed c, the value of the function $W(x, y, z)$ increases at a constant rate c_o.

We must now determine the rate of change of W with respect to s. Recalling that the direction of a ray path is normal to the surface $W(x, y, z) = \text{constant}$, we see that the rate of change we seek is given by the gradient of W, whose magnitude is $\frac{dW}{ds}$ and whose direction is the direction of the ray path. The gradient is

$$\text{grad } W = \hat{i} \frac{\partial W}{\partial x} + \hat{j} \frac{\partial W}{\partial y} + \hat{k} \frac{\partial W}{\partial z} \tag{411}$$

The magnitude A of any vector

$$\vec{A} = \hat{i} A_x + \hat{j} A_y + \hat{k} A_z$$

is

$$A = \sqrt{A_x^2 + A_y^2 + A_z^2}$$

Therefore

$$\frac{dW}{ds} = \sqrt{\left(\frac{\partial W}{\partial x}\right)^2 + \left(\frac{\partial W}{\partial y}\right)^2 + \left(\frac{\partial W}{\partial z}\right)^2} \tag{412}$$

and the direction cosines of the ray path are

$$\frac{\frac{\partial W}{\partial x}}{\frac{dW}{ds}}, \quad \frac{\frac{\partial W}{\partial y}}{\frac{dW}{ds}}, \quad \frac{\frac{\partial W}{\partial z}}{\frac{dW}{ds}} \tag{413}$$

Inserting (412) into (410) and squaring yields the desired differential equation

$$\left(\frac{\partial W}{\partial x}\right)^2 + \left(\frac{\partial W}{\partial y}\right)^2 + \left(\frac{\partial W}{\partial z}\right)^2 = \frac{c_o^2}{c^2} \tag{414}$$

Equation (414) is called the eikonal equation. It is of fundamental importance because it is the basis upon which ray theory rests. Families of rays are derived as solutions of this equation.

Before proceeding to apply these results to a practical problem, we shall inquire into the question of the validity of the eikonal equation.

4. Criterion for the Validity of the Eikonal Equation

It has been pointed out that the function (402) is not a solution of the wave equation when c is variable, but is a good approximation if c does not vary too rapidly. The same comments apply to the eikonal equation which was derived from it. In this section we shall attempt to determine what is meant by "not too rapidly." To this end we shall assume a general solution to the wave equation, of the form

$$p = A(x,y,z)\, e^{\frac{2\pi j}{\lambda_0}\left[c_0 t - W(x,y,z)\right]} \tag{415}$$

where the amplitude A is an undetermined function of the space coordinates. When we take the necessary derivatives and substitute them into the wave equation (401), we get both real and imaginary terms which must be equated separately. The two resulting equations are

$$\left(\frac{\partial W}{\partial x}\right)^2 + \left(\frac{\partial W}{\partial y}\right)^2 + \left(\frac{\partial W}{\partial z}\right)^2 - \frac{c_0^2}{c^2} = \frac{\lambda_0^2}{4\pi^2 A}\left(\frac{\partial^2 A}{\partial x^2} + \frac{\partial^2 A}{\partial y^2} + \frac{\partial^2 A}{\partial z^2}\right) \tag{416}$$

$$\frac{\partial^2 W}{\partial x^2} + \frac{\partial^2 W}{\partial y^2} + \frac{\partial^2 W}{\partial z^2} + \frac{2}{A}\left(\frac{\partial A}{\partial x}\frac{\partial W}{\partial x} + \frac{\partial A}{\partial y}\frac{\partial W}{\partial y} + \frac{\partial A}{\partial z}\frac{\partial W}{\partial z}\right) = 0 \tag{417}$$

Comparison with (414) shows that if W satisfies the eikonal equation, it will also be a solution of the wave equation only if the term on the right-hand side of (416), i.e.,

$$\frac{\lambda_0^2}{4\pi^2 A}\left(\frac{\partial^2 A}{\partial x^2}+\frac{\partial^2 A}{\partial y^2}+\frac{\partial^2 A}{\partial z^2}\right)$$

is zero. This will be true in general only in the limit of very high frequencies where $\lambda_0 \to 0$. The approximation will be good, however, if this term is small compared with c_0^2/c^2 (or with $\left(\frac{\partial W}{\partial x}\right)^2+\left(\frac{\partial W}{\partial y}\right)^2+\left(\frac{\partial W}{\partial z}\right)^2$, which amounts to the same thing).

A rough idea of how small λ_0 must be can be obtained by some order-of-magnitude computations. We saw earlier (412) that the sum of the squares of the space derivatives of W is equal to the square of tne derivative along the direction of maximum variation. The same general concept may be applied to the other groups of derivatives appearing in (416) and (417). Since we are interested only in the order of magnitude, we can loosely treat the sum of the three partial derivatives as a single derivative with respect to some space coordinate, even though we do not know its direction. A first derivative will be indicated by a prime symbol and a second derivative with a double prime. Because of the roughness of the estimates we may ignore the constants $4\pi^2$ and 2. Under these conditions (336) and (337) may be written as follows

$$(W')^2-\left(\frac{c_0}{c}\right)^2 \sim \lambda_0^2\,\frac{A''}{A} \qquad (418)$$

$$W''+\frac{W'A'}{A}\sim 0 \qquad (419)$$

The variation of c amounts to only a few percent, so that if we select a representative value for c_0 the ratio c_0/c is of the order of unity. This fact, together with the eikonal equation shows that

$$(W')^2 \sim \left(\frac{c_0}{c}\right)^2 \sim 1$$

or

$$W' \sim \frac{c_0}{c} \sim 1 \qquad (420)$$

From this we see that our criterion is

$$\lambda_0^2\,\frac{A''}{A} << 1 \qquad (421)$$

This result needs to be expressed in more tangible terms. To do this, let us obtain A'' by taking the derivative of (419). In all of these operations it must be noted that algebraic signs have no significance since we do not know the signs of the various quantities. The object is to determine which terms are significant and which are not. Taking the derivative, ignoring signs, and noting that $W' \sim 1$, we obtain

$$\frac{A''}{A} \sim W''' + \frac{W'' A'}{A} + \left(\frac{A'}{A}\right)^2 \qquad (422)$$

We now wish to express W''' in terms of derivatives of c and to show that the last two terms are negligible. The derivative of (420) is

$$W'' \sim \frac{c_o c'}{c^2} \sim \frac{c'}{c} \qquad (423)$$

and from (419)

$$\frac{A'}{A} \sim W'' \sim \frac{c'}{c} \qquad (424)$$

Thus the last two terms of (422) are each of the order of magnitude of $(c'/c)^2$. Since the fractional rate of change in c, that is c'/c, is normally quite small, we may neglect its square, so that (422) becomes

$$\frac{A''}{A} \sim W'''$$

Finally we take the derivative of W'', neglecting the term involving $(c'/c)^2$ obtaining

$$W''' \sim \frac{c''}{c} \sim \frac{A''}{A} \qquad (425)$$

The criterion (421) is thus roughly equivalent to

$$\frac{\lambda_o^2 c''}{c} << 1 \qquad (426)$$

Now c' is the sound velocity gradient, which is commonly denoted by the symbol g. c'' is the rate of change of the gradient, and $\lambda_o c''$ is the change in the gradient over the space of one wavelength, which might logically be written as Δg. The condition (426) is then

$$\frac{\lambda_o \Delta g}{c} << 1 \qquad (427)$$

From these results we see that the validity of the ray solution is dependent upon the rate of change of the velocity gradient. So long as the change in the gradient over a wavelength is small (compared with c/λ_0), ray diagrams will give a good description of the sound field. On the other hand, in regions of rapid change, the extreme example of which is a boundary between different media, ray theory breaks down.

5. Application to Typical Ocean Conditions - Vertical Velocity Gradient

In most ocean areas the speed of sound varies chiefly as a function of the depth, the changes in the horizontal directions being relatively small. We shall apply the eikonal equation to the ideal case where c depends only upon the y coordinate. It will turn out as a consequence of this assumption that sound rays are refracted only in a vertical plane - no horizontal refraction. The problem therefore reduces effectively to two dimensions. We shall designate the horizontal coordinate as x and the vertical coordinate as y, and the z coordinate will drop out. Our basic assumption is that c is a function only of y.

We shall employ the following conventions and notation

x = horizontal distance from sound source,

y = vertical distance measured downward from the ocean surface,

s = distance measured from source along ray path,

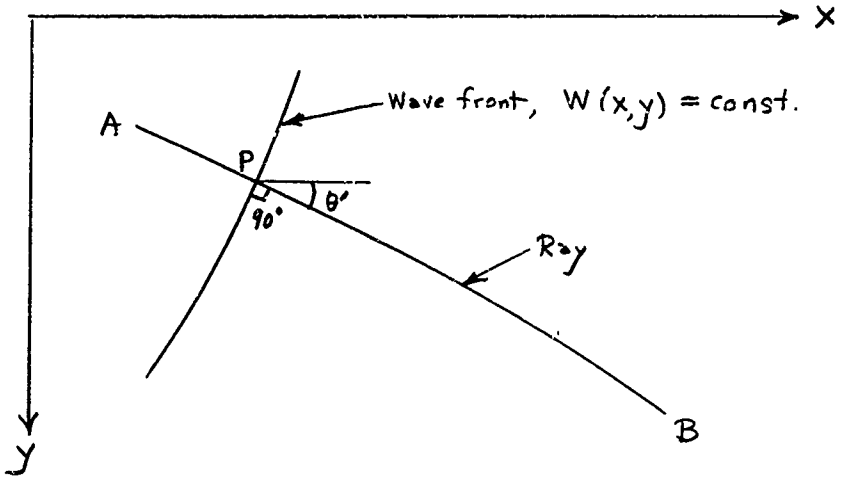
t = travel time from source along ray path,

θ' = angle of ray path relative to the horizontal, measured positive downward,*

$g = \frac{dc}{dy}$ = velocity gradient, a function of y only.

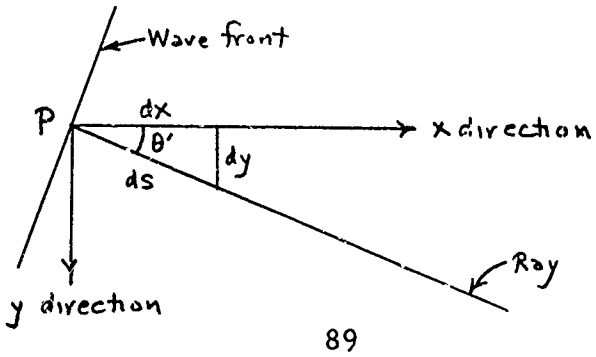
*The symbol θ' is used in this section to denote the angle measured positive downward. We shall reserve the plain symbol θ to denote the angle measured positive upward.

The accompanying sketch shows a wave front $W(x, y) =$ constant, and a ray AB which intersects the wave front at the point P, where it is perpendicular to the wave front. The angle of the ray with the horizontal at this point is θ'.



In the analysis which follows we shall seek to determine the characteristics of a ray path by deriving relations among the variables $x, y, s, t,$ and θ'. We shall begin with the function $W(x, y)$ defined by the eikonal equation and shall derive a set of differential equations whose solutions yield the desired relationships. Analytical solutions can be obtained if certain simplifying assumptions are made. The differential equations will be expressed in terms of full (rather than partial) derivatives, it being understood that all differentials dx, dy, ds, etc., refer to increments taken along the ray path.

The sketch below shows an enlarged view of the portion of the preceding diagram in the vicinity of the point P. Let ds be



an elemert of length along the ray and let dx and dy be its components along the x and y axes. Then

$$dx = ds \cos \theta' \tag{428}$$

$$dy = ds \sin \theta' \tag{429}$$

The rate of change of θ' along the ray may be expressed in terms of partial derivatives with respect to x and y.

$$\frac{d\theta'}{ds} = \frac{\partial \theta'}{\partial x}\frac{dx}{ds} + \frac{\partial \theta'}{\partial y}\frac{dy}{ds}$$

$$= \frac{\partial \theta'}{\partial x}\cos\theta' + \frac{\partial \theta'}{\partial y}\sin\theta' \tag{430}$$

We now return to the W function. Since the problem has been reduced to two dimensions, the wave front $W(x, y) = \text{constant}$ is a curved line in the xy-plane and the gradient of W is a vector (in the xy-plane) which points in the direction of the ray, and whose magnitude, according to (410), is

$$\frac{dW}{ds} = \frac{c_o}{c}$$

The direction cosines of the ray, i.e., cosines of the angles with the x and y axes, are $\cos\theta'$ and $\sin\theta'$ respectively, and these must be equal to the respective direction cosines of the gradient, as given by (413). Therefore, relative to the x axis we have

$$\frac{\frac{\partial W}{\partial x}}{\frac{dW}{ds}} = \frac{\frac{\partial W}{\partial x}}{\frac{c_o}{c}} = \cos\theta'$$

or

$$\frac{\partial W}{\partial x} = \frac{c_o \cos\theta'}{c} \tag{431}$$

and relative to the y axis

$$\frac{\partial W}{\partial y} = \frac{c_o \sin\theta'}{c} \tag{432}$$

If we take the partial derivative of (431) with respect to y and the partial derivative of (432) with respect to x, we obtain in each case the second partial of W with respect to x and y, and resulting expressions will be equal. Recalling that c is a function of y only, we obtain

$$\frac{\partial^2 W}{\partial x \partial y} = -\frac{c_0 \sin\theta'}{c}\frac{\partial \theta'}{\partial y} - \frac{c_0 \cos\theta'}{c^2}\frac{dc}{dy} = \frac{c_0 \cos\theta'}{c}\frac{\partial \theta'}{\partial x}$$

or

$$c\left(\frac{\partial \theta'}{\partial x}\cos\theta' + \frac{\partial \theta'}{\partial y}\sin\theta'\right) = -\frac{dc}{dy}\cos\theta'$$

Substitution of (430) gives

$$c\frac{d\theta'}{ds} = -\frac{dc}{dy}\cos\theta' \tag{433}$$

and, from (429)

$$\frac{dc}{c} = -\frac{\sin\theta' d\theta'}{\cos\theta'} \tag{434}$$

Equation (434) integrates directly to

$$\log c = \log \cos\theta' + \text{constant}$$

If we write the constant as $\log c_v$, we obtain

$$c = c_v \cos\theta' \tag{435}$$

Equation (435), which says that the cosine of the ray inclination angle is proportional to the speed of sound, is of fundamental importance in ray acoustics. It is a generalized form of Snell's Law, which is familiar to students of geometric optics. The more familiar form of Snell's Law is encountered when a wave travels across a discrete boundary between two homogeneous media having distinctly different speeds of propagation, and experiences a sharp change in direction. Equation (435) applies not only to this type of discrete refraction but also to the continuous refraction which results when the speed of propagation varies in a continuous manner within the medium. The derivative dc/dy is the (vertical) velocity gradient g.

Inserting this symbol in (433) and applying Snell's Law, we obtain a relation between ds and $d\theta'$.

$$ds = -\frac{c_v}{g}\, d\theta' \tag{436}$$

Equations for dx and dy in terms of $d\theta'$ are obtained by applying (428) and (429).

$$d_x = -\frac{c_v}{g} \cos\theta'\, d\theta' \tag{437}$$

$$dy = -\frac{c_v}{g} \sin\theta'\, d\theta' \tag{438}$$

Finally, the increment in travel time is

$$dt = \frac{ds}{c} = \frac{ds}{c_v \cos\theta'}$$

or

$$dt = -\frac{d\theta'}{g \cos\theta'} \tag{439}$$

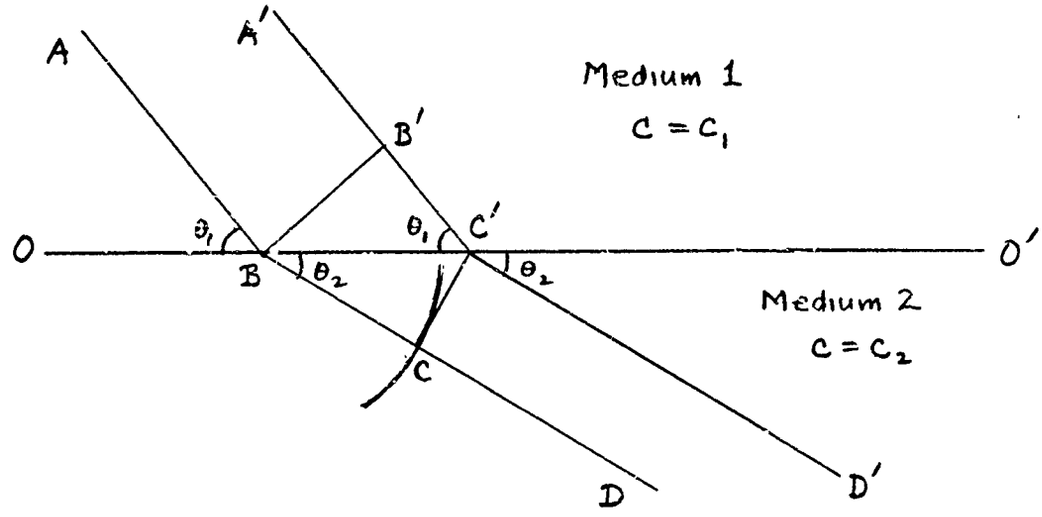
Equations (436), (437), (438), and (439) are particularly useful in the special case where the gradient is constant, since they may be integrated directly to give a set of parametric equations expressing x, y, s, and t as functions of the angle θ'.

6. Snell's Law

Snell's Law may be derived in its more familiar form by considering the passage of a wave across a boundary separating two homogeneous media, such as water and air, in which the speed of propagation has distinctly different values. The effect here is to cause a sudden change in the direction of propagation, as can be seen in the case of visible light when a straw is placed at an oblique angle in a glass of water. This is the basic meaning of the word refraction, although the meaning has been extended to include the continuous curving of rays which occurs when the speed of propagation changes gradually within a medium.

In the sketch below, the line OO' indicates the boundary between two homogeneous media. The speed of propagation is c_1 in the first medium

and c_2 in the second. A plane wave is traveling downward in the first medium at an angle θ_1 with the boundary. The wave front at the instant when the ray AB reaches the boundary is shown by the line BB', which is



perpendicular to AB. As the wave crosses the boundary the speed of propagation is suddenly changed from c_1 to c_2. Therefore, while the ray A'B'C'D' is traversing the distance B'C', the ray ABCD traverses a different distance BC. When the ray A'B'C'D' reaches the point C', the wave front will lie along the line CC'. To locate the point C, we swing an arc of radius BC about the point B, and then draw a tangent line from C' to this arc. This procedure will determine the direction θ_2 in which the ray BCD must travel in the second medium in order to be perpendicular to the wave front CC'.

The travel time from B to C is the same as from B' to C', and is equal to the distance divided by the speed. Hence

$$\frac{BC}{c_2} = \frac{B'C'}{c_1}$$

But

$$BC = BC' \cos \theta_2$$

and

$$B'C' = BC' \cos \theta_1$$

Therefore

$$\frac{BC' \cos \theta_2}{c_2} = \frac{BC' \cos \theta_1}{c_1}$$

or

$$\frac{\cos \theta_2}{c_2} = \frac{\cos \theta_1}{c_1} \qquad (440)$$

Equation (440) is the conventional form of Snell's Law. It is clear that if there were other media in the form of parallel layers, whose boundaries were parallel to OO', the above relation would hold for a ray traveling through all of them, so that

$$\frac{c_1}{\cos \theta_1} = \frac{c_2}{\cos \theta_2} = \frac{c_3}{\cos \theta_3} = \cdots = \text{constant} = c_v \qquad (441)$$

Suppose now that in the ocean the speed of sound varies continuously with the depth. We could approximate the true situation by dividing the ocean into a number of layers in each of which the speed of sound is assumed to be constant and to have a value corresponding to the center of the layer. Equation (441) would then apply. We may now go through a limiting process of letting the number of layers go to infinity and the thickness of each layer go to zero, and thereby arrive at a continuous variation of θ with c, given by

$$c = c_v \cos \theta \qquad (435)$$

This, of course, is the same result which we obtained from the eikonal equation.*

*The prime symbol on θ has been omitted here because we are not concerned about whether the angle is positive upward or positive downward. As a matter of fact the cosine is the same whether the angle is positive or negative.

Two facts must be borne in mind in applying Snell's Law. First, equation (435) is valid only in physical situations where the speed of sound is a one-dimensional space function. The angle θ is measured relative to the direction along which c is constant. In the ocean, c must vary only with depth and θ is measured from the horizontal.

Second, the constant c_v applies only to a particular ray. When, for example, we wish to consider a family of rays which spread out from a sound source, different rays will have different values of c_v, depending upon the angle at which they leave the source. The physical significance of c_v may be seen by letting $\theta = 0$. From this it is seen that c_v is the speed of propagation at the depth at which the ray is horizontal. As we shall see, the existence of a velocity gradient causes rays to curve either upward or downward. An upward-curving ray, for example, will descend to a maximum depth, at which it is horizontal, and will then curve back upward toward the surface. The point at which a ray reaches a maximum or minimum depth is called a vertex, and the speed of sound, c_v, at this depth is called the vertex velocity. It is clear that rays which leave a source at different angles (relative to the horizontal) will vertex at different depths and will therefore have different values of c_v.

Snell's Law applies not only to the refraction of sound rays in a stratified ocean, but also to reflection from the top and bottom (assuming, of course, that the reflection is specular). An interesting problem arises in the case of a sloping bottom, which violates the basic requirement that c be independent of the horizontal coordinate. In this case the reflected ray acquires a new value of c_v which depends upon the tilt angle of the bottom.

Refraction at Boundary; Critical Angle

To investigate further the properties of refraction at the interface between two different media, let us assume that the speed of propagation in the second medium is larger than in the first, that is, $c_2 > c_1$, as implied by the preceding diagram. From (440) we see that

$\cos \theta_2 > \cos \theta_1$. Since the cosine decreases from 1 to 0 as the angle increases from 0 to 90 degrees, it is evident that $\theta_2 < \theta_1$. Thus, when a ray passes from a medium of low sound speed to a medium of high sound speed, it is bent toward the plane of the interface. Conversely, when a ray travels into a medium of lower sound speed, it is bent away from the plane of the interface. Since sound travels faster in water than in air, sound rays traveling from air into water will be bent toward the horizontal, and sound rays traveling from water into air will be bent away from the horizontal. It is interesting to note that the reverse is true for rays of light.

Rewriting Snell's Law in the form

$$\cos \theta_2 = \frac{c_2}{c_1} \cos \theta_1$$

we see that whenever $\cos \theta_1 > c_1/c_2$, $\cos \theta_2 > 1$. Since the cosine of a real angle can never exceed one, the angle θ_2 ceases to exist in this case. The angle $\theta_1 = \theta_c$ in the slower medium, where

$$\cos \theta_c = \frac{c_1}{c_2}, \quad (c_1 < c_2) \tag{442}$$

is called the critical angle. The corresponding angle θ_2 in the faster medium is given by

$$\cos \theta_2 = 1$$

or

$$\theta_2 = 0$$

The physical interpretation of these results is as follows. When a ray in the slower medium (e.g., air) is incident upon the boundary at an angle $\theta_1 > \theta_c$ with the horizontal, the ray enters the faster medium (e.g., water) and is bent toward the horizontal. At the critical angle θ_c the refracted ray travels along the interface ($\theta_2 = 0$). When the incident ray is more nearly horizontal than the critical ray, that is, when $\theta_1 < \theta_c$, the ray does not enter the faster medium at all, but is totally reflected.

When a ray is incident from the faster medium, there is no critical angle. Refraction occurs for all angles of incidence, and the angle between the refracted ray and the horizontal will never be less than θ_c.

As an illustration, let us select typical values for air and water.

$$c_{air} = 1100 \text{ ft/sec}$$

$$c_{water} = 5000 \text{ ft/sec}$$

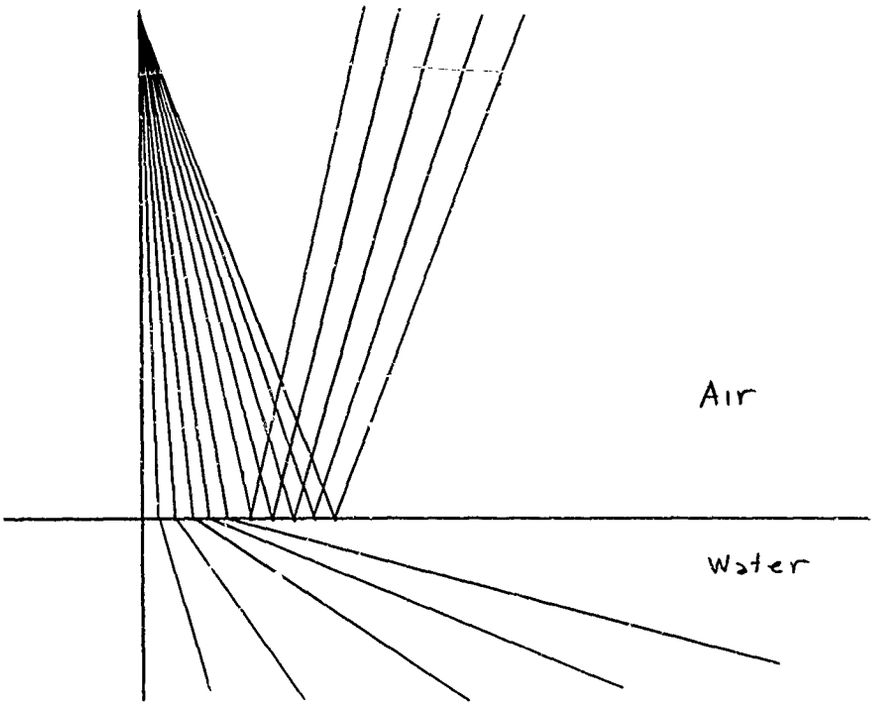
Snell's Law states

$$\cos \theta_{air} = 0.22 \cos \theta_{water} \qquad (443)$$

The critical angle (in air) is

$$\theta_c = \cos^{-1} 0.22 = 77.3 \text{ degrees} \qquad (444)$$

Thus, if a sound source is located in the air above an ideally smooth water surface, as illustrated in the following sketch, only those rays which lie within a narrow cone of 12.7 degrees from the vertical will penetrate into the water. All rays outside this cone are totally reflected back into the air.

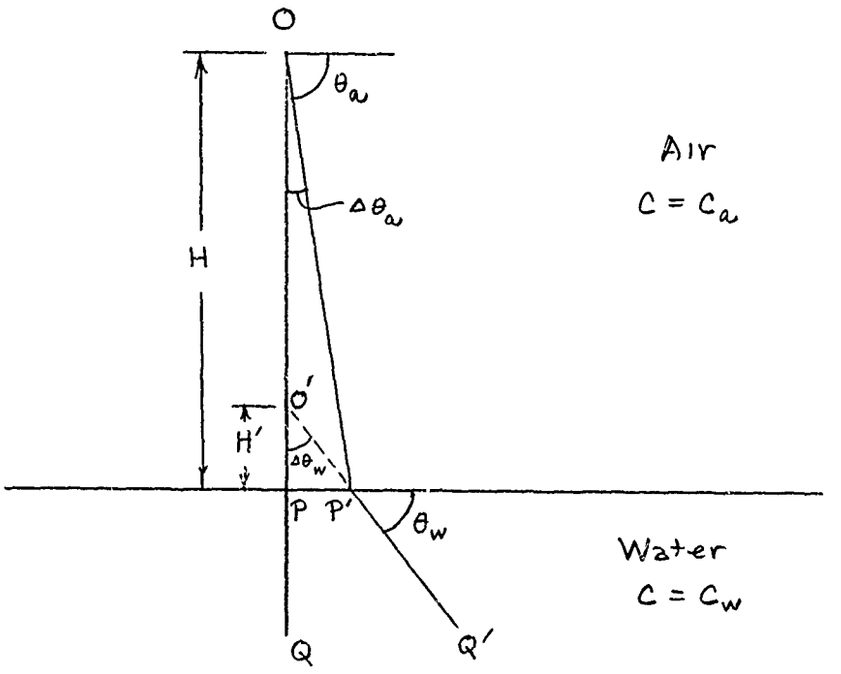


Virtual Sources (Images)

It will be observed that the rays shown in the water in the preceding sketch appear to be spreading out from some point in the air which is closer to the water than the real source. This is the same effect that one finds in optics - the formation of an image or virtual source. Actually, if the rays in the water are projected back beyond the interface, they do not all converge to exactly the same point, but for every small bundle of rays such as would be intercepted by a hydrophone there is an image source somewhere in the air above.

The simplest case to consider is the case of vertical incidence. If a hydrophone is located in the water directly underneath a source in air, where is the image located ? It is obvious that the image must be located somewhere on a vertical line through the source, because a vertical ray (at normal incidence) does not experience a change in direction. To determine where the image is located on this line it is not enough to consider a single ray. We must examine a bundle, or narrow cone, of rays and find the point from which they appear to be spreading after they enter the water. For this purpose the bundle can be represented by two rays separated by an infinitesimally small angle. We shall select the vertical ray as one of the two and shall select for the other a neighboring ray which leaves the source at a small angle $\Delta\theta_a$ with the vertical.

The geometry of the situation is shown in the following sketch. The true source is located at the point O at a height H above the water. The vertical ray is OPQ. The neighboring ray is OP', which upon entering the water is bent away from the vertical and becomes P'Q', making an angle $\Delta\theta_w$ with the vertical. (The angles $\Delta\theta_a$ and $\Delta\theta_w$ have been grossly exaggerated in the sketch.) When P'Q' is extended back into the air, it intersects the vertical ray at the point O' at a height H' above the water. O' is the image. Knowing H, $\Delta\theta_a$, and the sound speeds c_a and c_w in air and water, we must now determine H'.



The angles employed in our formulation of Snell's Law are measured relative to the horizontal. Therefore

$$\theta_a = \frac{\pi}{2} - \Delta\theta_a$$

$$\theta_w = \frac{\pi}{2} - \Delta\theta_w$$

Also, since $\Delta\theta_a$ and $\Delta\theta_w$ are exceedingly small,

$$\cos\theta_a = \sin\Delta\theta_a \approx \Delta\theta_a$$

$$\cos\theta_w = \sin\Delta\theta_w \approx \Delta\theta_w$$

Snell's Law takes the form

$$\frac{\Delta\theta_a}{\Delta\theta_w} \approx \frac{c_a}{c_w}$$

(In the limit as $\Delta\theta_a$ and $\Delta\theta_w$ go to zero, this equation becomes exact.) From the figure it is seen that

$$PP' = H \tan\Delta\theta_a = H' \tan\Delta\theta_w$$

or, since the tangent of a small angle is very nearly equal to the angle itself (in radians),

$$H' \approx \frac{\Delta\theta_a}{\Delta\theta_w} H$$

or

$$H' = \frac{c_a}{c_w} H \tag{445}$$

For the case of air and water $c_a/c_w = 0.22$ and

$$H' = 0.22\, H \tag{445a}$$

At vertical incidence the height of the virtual source above the water is about 1/5 the height of the true source.

We shall not analyze the general case of oblique incidence. However, the results of such an analysis show that as the angle of the incident ray (relative to the horizontal) is decreased from 90 degrees to the critical angle, the image point traces out a smooth curve which begins with infinite slope along the vertical ray and, when the critical angle is reached, ends with zero slope at a point on the air-water interface at a horizontal distance

$$x = \frac{\frac{c_a}{c_w} H}{\left(1 - \frac{c_a^2}{c_w^2}\right)^{1/2}} \tag{446}$$

from the vertical ray. For air and water the denominator in (446) is almost equal to one, so that at the critical angle the image is displaced laterally at the surface by a distance approximately equal to the height above the surface in the case of vertical incidence.

In addition to knowing the direction in which rays are refracted at a boundary we are also interested in knowing how much of the incident energy is transmitted into the adjacent medium and how much is reflected. Unfortunately ray theory does not give us answers to this question. Further discussion of this subject will be found in the section on reflection.

7. The Speed of Sound in the Ocean; Velocity Gradients

(Note: There is much confusion about the terms speed and velocity as applied to wave propagation. When one is concerned about not only the speed but also the direction in which a wave is moving through the water, the word "velocity" is appropriate. However, when one is concerned only about the properties of the medium, as we are at present, the direction of propagation is of no concern. If c has a value of, say, 4950 ft/sec at a certain place in the ocean, then a wave at this place will propagate at 4950 ft/sec regardless of the direction. The proper word to describe this property of the medium is therefore "speed," although the two terms are used indiscriminately in the literature. Gradients of c are almost universally called velocity gradients, and in spite of the inconsistency, we shall use this term also.)

Extensive measurements have been made of the speed of sound in sea water at various temperatures, pressures, and salinities, and highly accurate empirical formulas have been derived. The most accurate results have been obtained by Wayne Wilson of the Naval Ordnance Laboratory. However, since Wilson's formulas are exceedingly complicated, we shall use an earlier formula which is far simpler and sufficiently accurate for the purpose of these notes.

$$c = 4422 + 0.0182y + 11.25T - 0.045T^2 + 4.3\,(sal - 34) \qquad (447)$$

where

c = sound speed in ft/sec

T = temperature in °F

y = depth in feet

sal = salinity in parts per thousand

Except in areas where mixing of fresh water and salt water occurs, the salinity term in (447) is relatively unimportant, and we shall neglect it in these notes. We shall also make the assumption that variations in temperature in the horizontal direction are negligible, so

that T is assumed to be a function of y only. Under these assumptions the speed of sound is also a function only of the depth. The gradient of c therefore has only a vertical component. On the basis of (447), neglecting the salinity term, the velocity gradient is

$$g = 0.0182 + (11.25 - 0.09T)\frac{dT}{dy} \qquad (448)$$

where

$$g = \frac{dc}{dy} = \text{velocity gradient in } sec^{-1}$$

$$\frac{dT}{dy} = \text{temperature gradient in } °F/ft$$

When c or T increases with increasing depth, the corresponding gradient is positive; if it decreases with depth, the gradient is negative.

The first term in (448) is the pressure term. The pressure increases linearly with depth, and therefore its contribution to the velocity gradient is constant. This means that where the temperature and salinity are constant there is always a positive velocity gradient of $0.0182\ sec^{-1}$.

The second term is the temperature term. It is seen that the temperature contribution is proportional to the temperature gradient dT/dy. Because of the parabolic relationship between sound speed and temperature, the coefficient of dT/dy is itself a function of temperature. It is seen that at high temperatures the temperature gradient has less effect on g than at low temperatures. It is also seen that where there is a constant temperature gradient, the temperature varies linearly with depth and hence, according to (448) the velocity gradient cannot be constant. Actually, however, such non-linearities are usually quite small - smaller than the errors in measurements - and it is generally assumed that a constant temperature gradient produces a constant velocity gradient. Temperature gradients in the ocean, especially in the near surface region (down to 1000 to 2000 feet), exert a strong influence upon sound propagation and thus have a powerful effect upon the performance of underwater acoustic systems.

In isothermal water the velocity gradient is positive, due to the pressure effect. A sufficiently negative temperature gradient will cause the velocity gradient to become negative. The limiting value of the temperature gradient required to balance the pressure effect and produce iso-velocity water ($g = 0$) is

$$\frac{dT}{dy} = - \frac{0.0182}{11.25 - 0.09T} \quad °F/ft$$

Its numerical value at a temperature of 50°F is

$$\frac{dT}{dy} = - \frac{0.0182}{6.75} = - 0.0027 \ °F/ft \qquad (449)$$

We see that an extremely small change in temperature - less than 0.3°F in 100 feet - is sufficient to change the velocity gradient from positive to negative.

Measurements of the Speed of Sound

Oceanographic measurements of the speed of sound have in the past been made indirectly—by measuring the temperature, pressure, and salinity, and computing the sound speed from empirical formulas, such as (447). The classical techniques used by oceanographers for making these measurements are somewhat cumbersome, and for this reason measurements are usually made only at a limited number of depths. The standard depth intervals are frequently too large to give an adequate picture of the velocity structure in the variable near-surface region.

Shortly before World War II a continuous recording temperature instrument known as the bathythermograph was developed at the Wood's Hole Oceanographic Institution for use in the near-surface region down to several hundred feet. The record of temperature vs. depth is inscribed on a smoked glass slide held on a moving carrier in a cylindrical body. As the instrument is lowered, the graph is traced by a stylus moved along the arc of a circle by a radial arm. The angle of the arm is controlled by the temperature, while the position of pivot is controlled by the pressure by means of a Bourdon tube.

The bathythermograph gives a quick and fairly precise record of the temperature profile in the near-surface region and has become standard equipment on ships engaged in undersea warfare.

As a result of the development of the sound velocimeter in recent years, sound speeds may now be measured directly. This instrument measures the time required for a pulse to travel a fixed distance. Direct measurement of the speed of sound not only saves time and labor, but also makes it possible to determine the effects of bubbles and foreign particles.

Velocity Structure of the Oceans

The variability of the speed of sound in the ocean is due chiefly to temperature variations. Conditions are most variable in the first thousand feet or so, as a result of surface effects, such as the sun, wind, waves, surface currents, etc. In most ocean areas there exists a surface layer, called the isothermal layer, in which the temperature is substantially constant. The isothermal layer arises from a mixing action at the surface and its thickness varies with the season of the year. During the fall and winter when the surface is cooled, the water at the top becomes more dense than the water underneath. This is an unstable condition, and the mixing which results contributes to the formation of the isothermal layer. Other factors such as the wind also contribute. In the North Atlantic the average thickness of the layer varies from about 500 feet in winter to between 0 and 100 feet in summer. The isothermal layer is a region of positive velocity gradient.

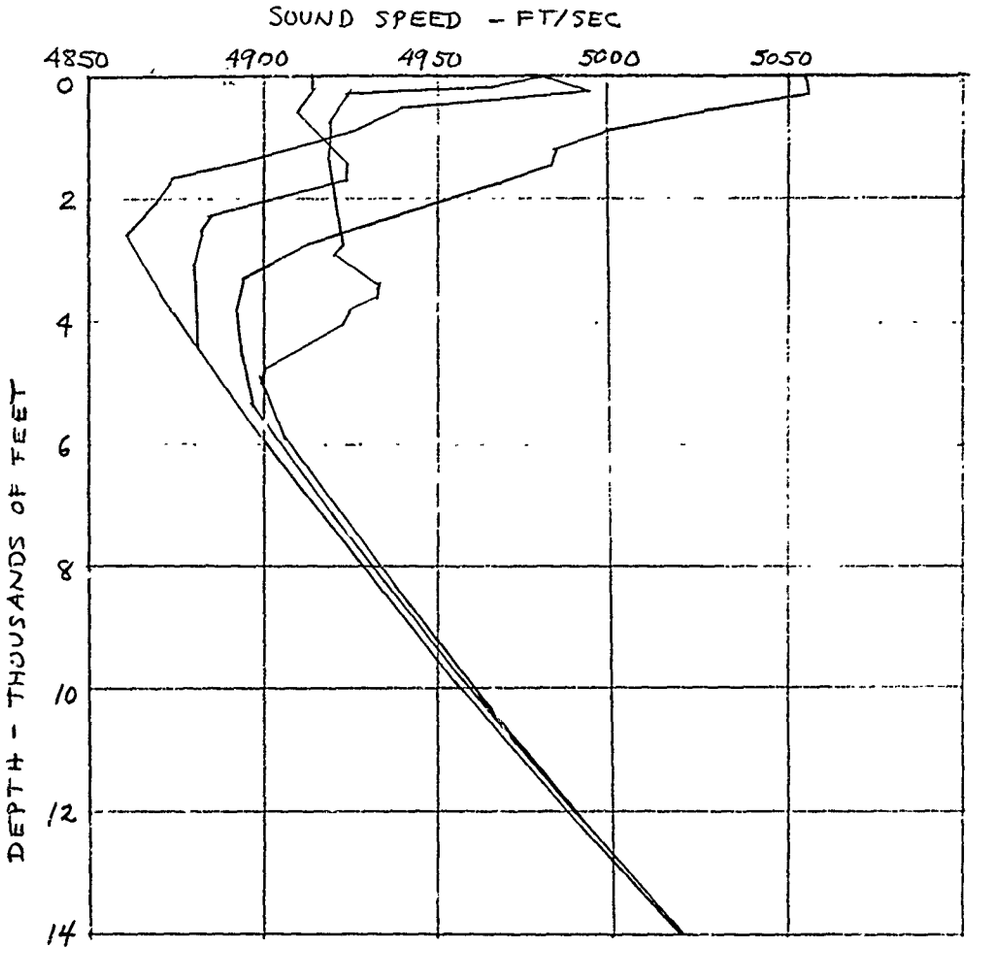
An interesting variation frequently occurs during the daytime. Heating of the surface due to the sun produces a shallow negative gradient layer late in the day, which has been observed to influence the performance of sonars with shallow transducers. It disappears during the night. This is known as the afternoon effect.

Below the isothermal layer there is typically a layer in which the temperature drops rapidly. This is called the thermocline. It is a region of large negative velocity gradient.

Occasionally there exists below the thermocline a layer in which the temperature increases with depth before dropping again. The combination of a positive gradient below a negative gradient produces what is known as a sound channel. In this case it is called a depressed sound channel. Abnormally long detection ranges are obtained when both the sonar and the target are in a channel. By virtue of surface reflections the surface isothermal layer acts as a very efficient sound channel. It is frequently called a surface duct.

At greater depths the temperature continues to drop, but at a diminishing rate, and gradually approaches an isothermal condition toward the bottom. At a depth of several thousand feet (typically 3000 to 4000 feet in the mid-latitudes) there is a point where the thermal effect balances the pressure effect, producing a zero velocity gradient. Here the speed of sound is a minimum; above it the velocity gradient is negative and the sound speed increases (with decreasing depth) due to temperature; below it the velocity gradient is positive and the sound speed increases (with increasing depth) due to pressure. The deep ocean thus behaves as a sound channel - called the deep sound channel or SOFAR channel. The depth of minimum sound speed is the axis of the deep sound channel.

At depths below about 5000 feet most deep ocean areas throughout the world have substantially the same velocity structure (c vs. depth), and there is virtually no change with the season of the year. The accompanying diagram shows several typical velocity profiles for the North Atlantic Ocean. To locate the deep sound channel, draw a vertical line from the point of maximum sound speed near the surface to the point where this line intersects the lower portion of the curve. The interval of depths between these two points is the channel.



The strongest thermoclines generally occur where the surface temperature is the highest - in the lower latitudes and/or during the summer. In polar regions the surface temperature is usually so low that the water is nearly isothermal from top to bottom. In such locations a positive velocity gradient exists at all depths and the entire ocean acts as a surface duct. The presence of an ice cover in the Arctic region produces a peculiar effect. Low frequency sound propagates very well, but high frequency sound suffers high attenuation due to the roughness of the ice.

The Propagation of Underwater Sound

B. Refraction (continued)

8. Ray Paths in a Constant-Gradient Medium

A simple analytic solution of the differential equations can be obtained in the special case where the velocity gradient is constant. In deriving the solution we shall reverse sign of the ray angle and shall use the symbol θ.

$$\theta = -\theta'$$

The angle θ is measured positive upward, although the depth coordinate y is still measured positive downward. With this change the four differential equations (436) to (439) relating x, y, s, and t to θ take the form

$$dx = \frac{c_v}{g} \cos\theta \, d\theta \tag{450a}$$

$$dy = -\frac{c_v}{g} \sin\theta \, d\theta \tag{450b}$$

$$ds = \frac{c_v}{g} \, d\theta \tag{450c}$$

$$dt = \frac{d\theta}{g \cos\theta} \tag{450d}$$

Note also that the slope of the ray is

$$\frac{dy}{dx} = -\tan\theta \tag{451}$$

If g is constant, the solutions are

$$x = \frac{c_v}{g} \sin\theta + \text{const.} \tag{451a}$$

$$y = \frac{c_v}{g} \cos\theta + \text{const.} \tag{451b}$$

$$s = \frac{c_v}{g}\,\theta + \text{const.} \quad (\theta \text{ in radians}) \tag{451c}$$

$$t = \frac{1}{2g}\,\log_e \frac{1 + \sin\theta}{1 - \sin\theta} + \text{const.} \tag{451d}$$

For the moment, for illustrative purposes, let us translate the origin of x and y to such a point that the constants of integration in (451a) and (451b) are zero. We shall call the translated variables x' and y'.

Then (451a) and (451b) become

$$x' = \frac{c_v}{g}\,\sin\theta \tag{452a}$$

$$y' = \frac{c_v}{g}\,\cos\theta \tag{452b}$$

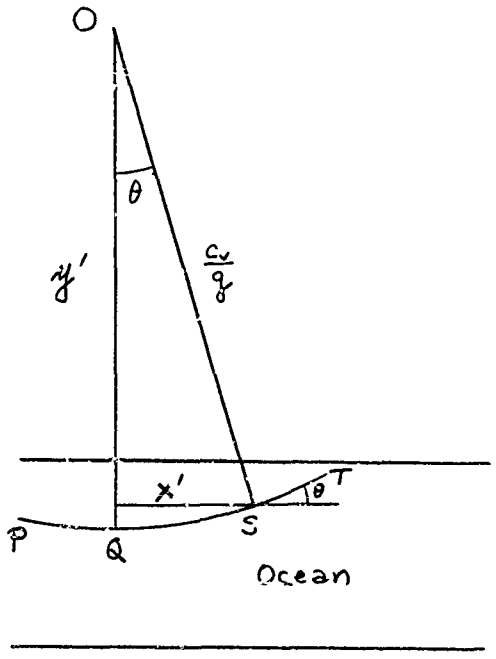
It is quite clear that the ray is an arc of a circle of radius $\frac{c_v}{g}$. Since g may be either positive or negative, the radius may likewise be either positive or negative. Let us consider for the moment an isothermal medium, in which g has a positive value 0.0182 sec^{-1}. Reasonable values of c_v would be in the neighborhood of 4800 to 5000 ft/sec. To make the arithmetic simple, suppose c_v = 4914 ft/sec. Then

$$\frac{c_v}{g} = \frac{4914 \text{ ft/sec}}{.0182 \text{ sec}} = 270{,}000 \text{ ft.}$$

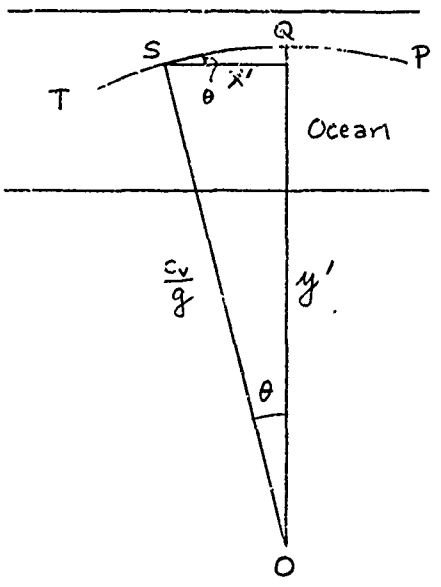
$$= 90{,}000 \text{ yd} = 90 \text{ kyd}$$

$$= 45 \text{ n. mi.}$$

The radius of a ray in an isothermal medium has the tremendous value of about 90 kyd. When the ray is horizontal, that is, when $\theta = 0$, y' has its maximum value of c_v/g. The ray arc is thus part of a huge circle whose center is about 45 miles up in the sky, as illustrated in the sketch below.



POSITIVE GRADIENT



NEGATIVE GRADIENT

The origin of the coordinates x' and y' is at 0, the center of the circle. The arc PQ ST is a portion of the ray. It curves upward.

Conversely, if g, and hence $\frac{c_v}{g}$, is negative, then y' is also negative. In this case the center of the circle is below the bottom of the ocean and the ray curves downward.

Let us now assume that the point S is a sound source and that the speed of sound at this depth is c_o. For convenience assume also that the velocity gradient is positive. The angle of inclination of a ray at the source will be called the initial angle, θ_o. If we consider a family of rays leaving the source in any one vertical plane, we see that each individual member of the family is identified by its particular value of θ_o. Therefore to investigate the characteristics of the ray family we shall express (452a) and (452b) in terms of θ_o by applying Snell's Law which, at the source, takes the form

$$c_v = \frac{c_o}{\cos \theta_o} \tag{453}$$

The coordinates of any point on a ray are then

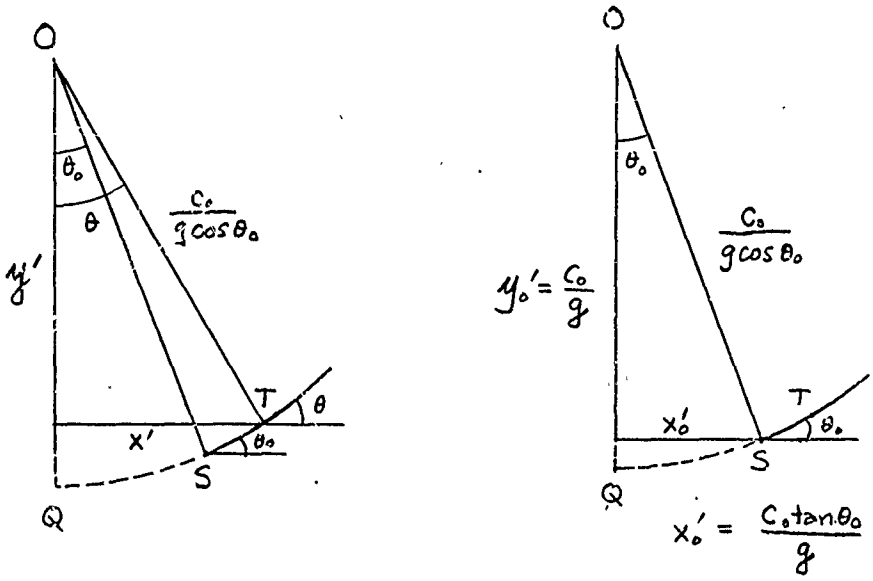
$$x' = \frac{c_o}{g \cos \theta_o} \sin \theta \tag{454a}$$

$$y' = \frac{c_o}{g \cos \theta_o} \cos \theta \tag{454b}$$

where x' and y' are measured from the center of curvature O of the circular ray arc ST. The coordinates (x'_o , y'_o) of the source S are the values obtained when $\theta = \theta_o$, i.e.,

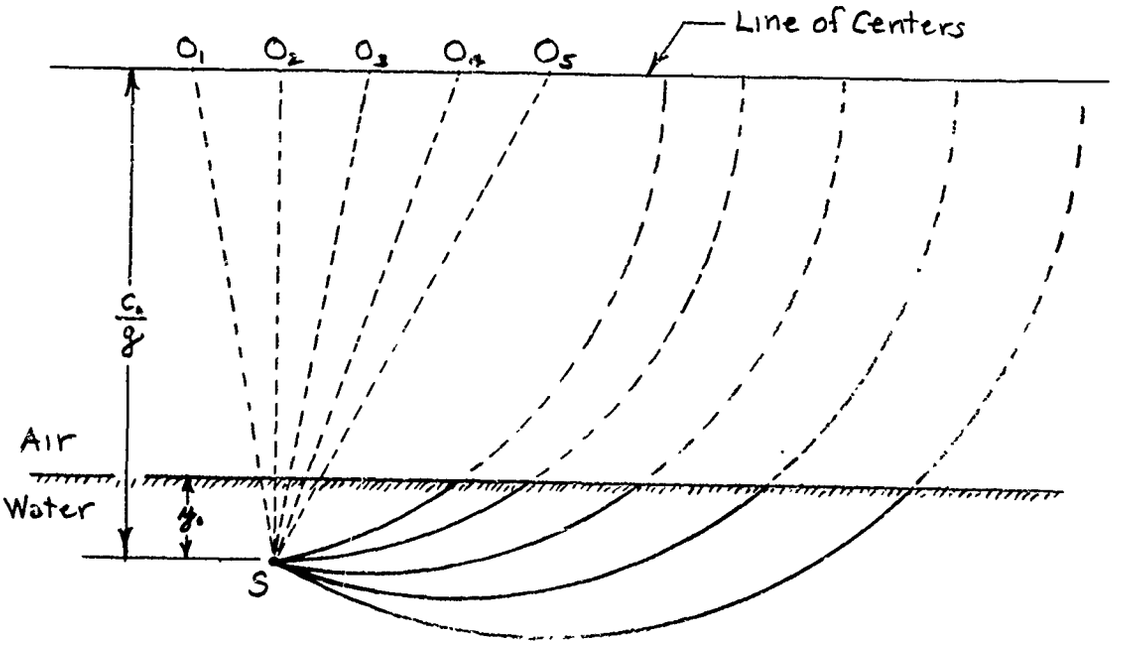
$$x'_o = \frac{c_o}{g} \tan \theta_o \tag{455a}$$

$$y'_o = \frac{c_o}{g} \tag{455b}$$



If we now focus our attention on the source we see that for each different initial angle there is a different ray arc whose center is located in accordance with (455a) and (455b). Note, however, that the y' coordinate is independent of θ_o. This means that the centers of all the circles lie on a horizontal line

at a distance c_o/g above the source. This line is called the line of centers. When the velocity gradient is negative, the line of centers is below the source.



The radius of curvature, $c_o/g \cos \theta_o$, is a minimum for the ray which leaves the source horizontally, and increases as the rays become steeper. The radius of curvature of a vertical ray is infinite, that is, a vertical ray is a straight line.

We shall now express the ray equations in terms of the original coordinates x and y, where, it will be recalled, x is the horizontal coordinate measured from the source and y is the vertical coordinate measured downward from the surface of the ocean. Let the source be placed at a depth y_o where the sound speed has the value c_o. The constants in (451a) to (451d) are evaluated from the following initial conditions:

$$\theta = \theta_o$$
$$x = 0$$
$$y = y_o$$
$$s = 0$$
$$t = 0$$

The resulting equations for x, y, s, and t are

$$x = \frac{c_o}{g \cos \theta_o} (\sin \theta - \sin \theta_o) \tag{456a}$$

$$y - y_o = \frac{c_o}{g \cos \theta_o} (\cos \theta - \cos \theta_o) \tag{456b}$$

$$s = \frac{c_o}{g \cos \theta_o} (\theta - \theta_o) \tag{456c}$$

$$t = \frac{1}{2g} \left[\log_e \frac{1 + \sin \theta}{1 - \sin \theta} - \log_e \frac{1 + \sin \theta_o}{1 - \sin \theta_o} \right] \tag{456d}$$

It is interesting to observe that the y-equation (456b) is in reality only an alternate way of stating Snell's Law. Let c denote the speed of sound at the depth y. Then, because the gradient is constant, it follows that

$$c - c_o = g(y - y_o) \tag{457}$$

Substitution of (457) into (456b) yields

$$y - y_o = \frac{c - c_o}{g} = \frac{c_o}{g \cos \theta_o} (\cos \theta - \cos \theta_o)$$

which reduces to

$$\frac{c}{\cos \theta} = \frac{c_o}{\cos \theta_o} \tag{457a}$$

Practical computations of ray paths in a constant-gradient medium are most conveniently made by treating y as the independent variable and proceeding as follows:

(1) Select a value for y.

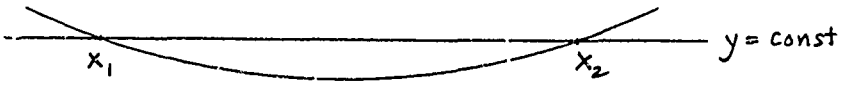
(2) Solve (456b), or the combination of (457) and (457a), for cos θ.

(3) Determine θ and sin θ from cos θ.

(4) Compute x, s, and t from (456a), (456c) and (456d).

It will be noted that in determining θ and sin θ from cos θ there is an ambiguity in algebraic sign, since both positive and negative values of θ lead to the same value of cos θ. The physical significance of this ambiguity

is evident from the fact that the horizontal line y = constant intersects the ray circle at two points. As indicated in the sketch below, there are two ranges, x_1 and x_2 corresponding to the same depth y. One of these ranges occurs



before the vertex of the ray and the other beyond. If one is computing a ray path by manual computation, it is easy to see from a rough sketch which range to use. However, all modern ray tracing is done on digital computers which are not smart enough to make decisions by looking at rough sketches. In writing a computer program, one establishes a set of simple ground rules by which the machine is enabled to make all its decisions.

One might logically inquire why we did not start with x instead of y. In this case we would obtain sin θ directly and avoid the ambiguity. Such a procedure would be excellent if the medium had no boundaries. In practice, however, we are always working in layers of finite thickness, and the boundaries between layers occur at specified values of y.

Small Angle Approximation

Occasions sometimes arise where a quick estimate of the horizontal range from a source to the vertex of a ray is needed. In most practical applications of conventional sonars the ray angles are quite small, and in such cases an extremely useful formula may be derived on the basis of small angle approximations. For this purpose we shall place the origin of x at the vertex, and shall denote the depth of the vertex by y_m. The coordinates of any other point (x, y) on the ray are obtained by inserting

$$\theta_o = 0$$

$$y_o = y_m$$

$$\sin\theta \approx \theta$$

$$\cos\theta \approx 1 - \tfrac{1}{2}\theta^2$$

into (456a) and (456b). The result is

$$x = \frac{c_o \theta}{g}$$

and

$$y - y_m = -\frac{c_o \theta^2}{2g}$$

Elimination of θ between these two equations yields

$$y - y_m = -\frac{g}{2c} x^2 \qquad (458)$$

or

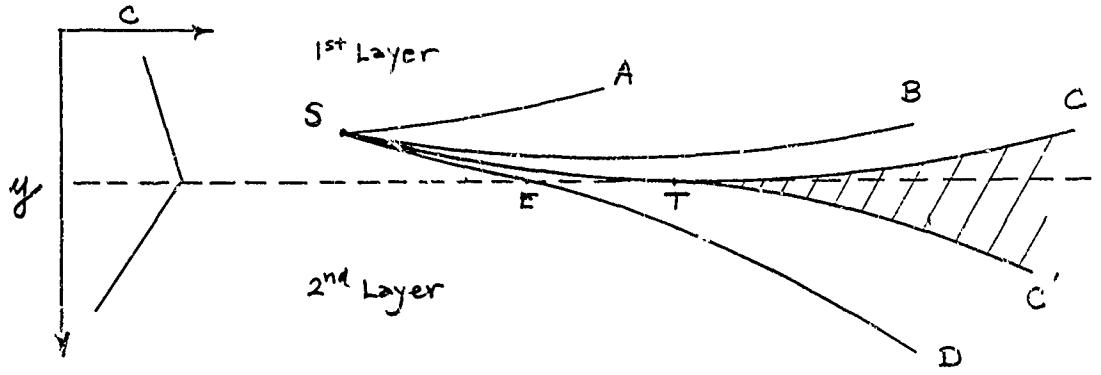
$$x = \sqrt{\frac{2 c_o}{g} (y_m - y)} \qquad (458a)$$

It is seen that the arc of the circle in the vicinity of the vertex has been approximated by a parabola. This approximation is quite good when the ray angle does not exceed about 10 degrees.

9. Multiple Constant-Gradient Layers; Boundary Effects

Having learned that rays in a constant-gradient layer are arcs of circles and that they curve upward where the gradient is positive and downward where it is negative, we are now in a position to examine the phenomena associated with the existence of multiple layers and of the ocean boundaries.

Consider first the simple case of two constant-gradient layers having different gradients. Suppose that the upper layer, which we shall call the first layer, has a positive gradient, while the lower or second layer has a negative gradient. As indicated in the sketch below, the boundary between the two layers is a point of maximum sound speed. Let the source S be located in the first layer. Any ray such as SA, which leaves the source with a positive (upward) initial angle will continue upward. A ray such as SB,

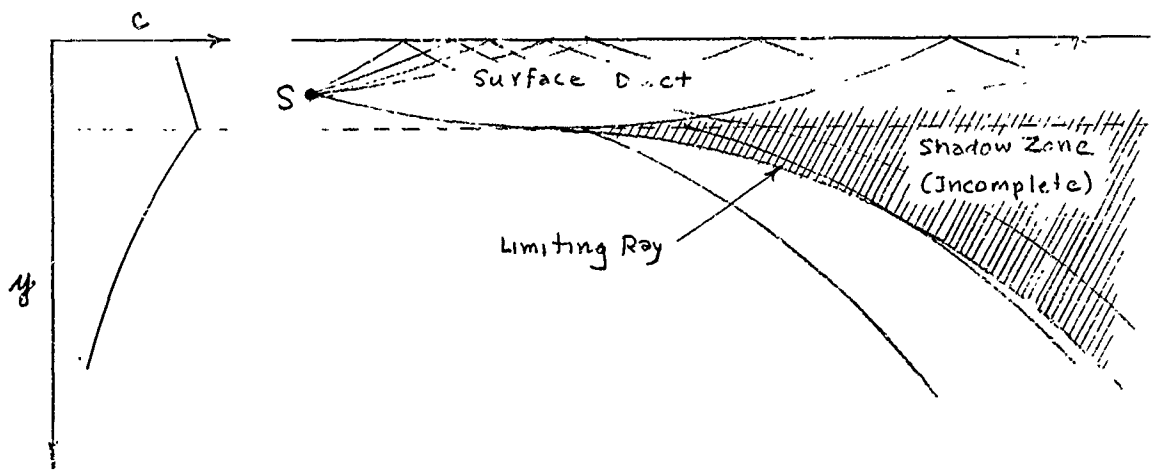


whose initial angle is slightly negative, will descend to a vertex in the first layer and then rise again due to the upward curvature. A ray such as SD, having a strongly negative initial angle, will cross the layer boundary with a finite slope at E and penetrate into the second layer where it will curve downward. Somewhere between SB and SD there is a special ray whose vertex occurs at a point of tangency T on the boundary. This ray is clearly the limiting case between the type SB which diverges upward from the boundary and the type SD which diverges downward. The ray ST is called a limiting ray. Beyond the vertex T it splits into two branches, the near branch TC and the far branch TC'. Between TC and TC' is a region into which (according to ray theory) no sound enters. This region is called a shadow zone. Shadow zones occur in the vicinity of a region of maximum sound speed because on either side of the maximum the rays curve away. A point of maximum sound speed tends to "repel" the rays.

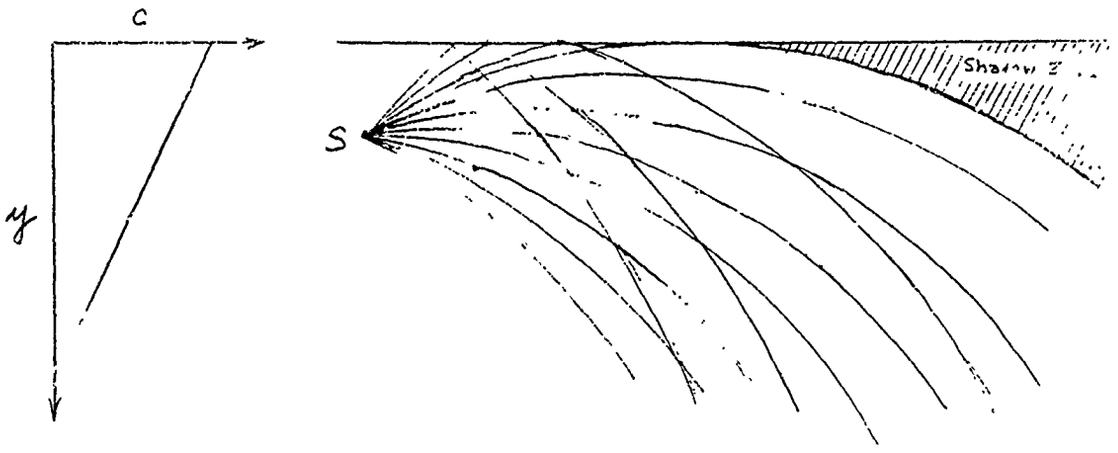
It should be noted that shadow zones do not exist in the strict sense in the real ocean. There are two reasons. In the first place, sharp, discontinuous changes in velocity gradients do not occur in nature. There is always a region, however small, of gradual transition, and this results in a small number of rays entering the idealized shadow zone. Secondly, we have learned that ray theory breaks down wherever the velocity gradient changes rapidly with depth. Therefore, even if the gradient changed discontinuously between the two layers, a certain amount of acoustic energy would leak by diffraction into the shadow zone. In spite of these considerations, however, very pronounced shadow zones, in which the sound intensity is exceedingly small, are actually observed to exist in the ocean. The effects described above result in a continuous transition into the shadow zone rather than a mathematically abrupt one, but inside the zone the spreading loss is so high that detections are all but impossible.

A similar pattern exists when the source is located in the lower layer, except, of course, that in this case the limiting ray leaves the source an upward angle.

The two-layer situation described above is typical of the conditions which exist over large areas of the ocean in the near surface region. The upper layer corresponds to the surface isothermal layer and the lower layer corresponds to the thermocline. However, in this case one new element has been added, namely, the surface boundary, which scatters sound energy back into the water. If we assume that surface acts as a smooth reflector—a fairly good assumption in low sea states—we see from the following sketch that a certain amount of reflected energy penetrates into the shadow zone.

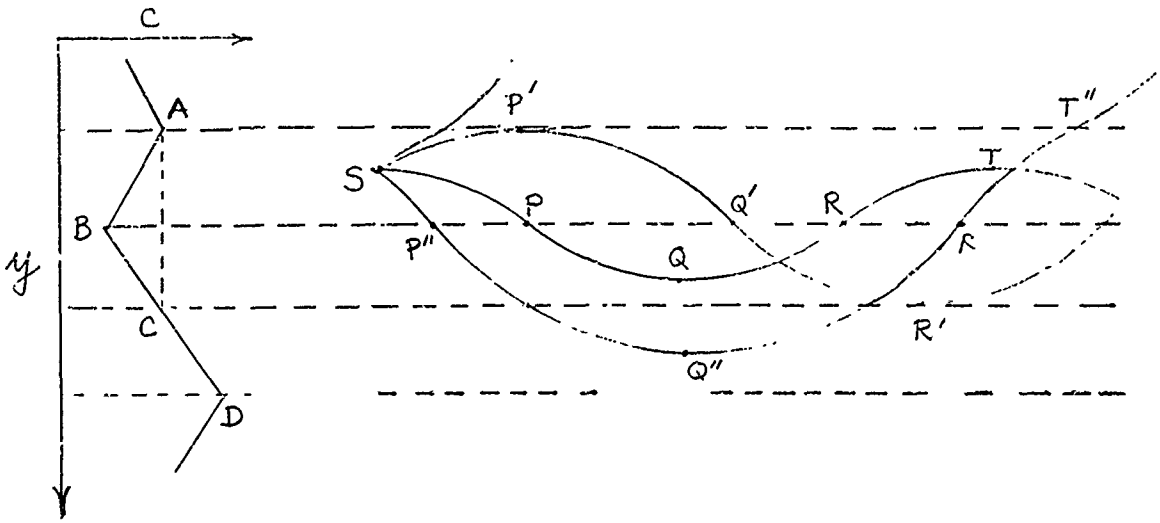


In the limiting case where the surface layer ceases to exist and the negative-gradient layer extends to the surface, the limiting ray generates the maximum range which can be achieved, since surface-reflected rays will always return at smaller ranges, as indicated in the sketch below.



A few general conclusions regarding placement of sonar transducers may be drawn from the preceding remarks. First, if a surface duct exists and both the sonar transducer and the target are located in the duct, abnormally long ranges can be achieved through a combination of surface reflections and upward refraction. Second, if the submarine is below the layer and the transducer is in the layer, or vice versa, poor detection ranges are to be expected. Third, the transducer should never be placed at or slightly below the interface between the two layers. Fourth, if the submarine is below the surface layer, the transducer should be lowered to the maximum practical depth to achieve the maximum detection range.

Let us now consider the case of two layers such that the upper layer has a negative gradient and the lower layer has a positive gradient, as illustrated by the following diagram. The upper layer is shown as extending from A to B on the velocity profile and the lower laye from B to D. The sound speed is a minimum at depth B. It is a maximum at A and D, and is assumed to decrease again beyond these two points. Suppose the sound source is located at S. All rays leaving S will curve downward. The Ray SPQR ..., which leaves S horizontally, crosses the



layer boundary at P and is refracted upward in the lower layer, vertexing at Q and returning to boundary at R. The upper vertex T occurs at the same depth as the source. It is thus seen that the ray oscillates up and down as it propagates forward and remains within the boundaries of the two layers. This is what is meant by a sound channel.

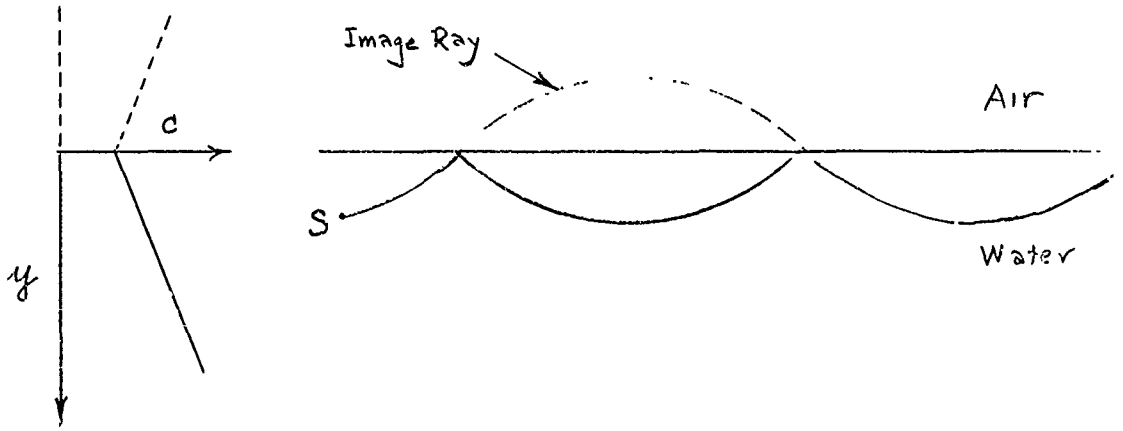
Consider now the ray SP' Q' R' ... which is a limiting ray at the upper boundary A. At P' this ray splits into two branches. The inner branch, with which we are presently concerned, crosses the interface at Q' and vertexes in the lower layer at the point R' at depth C. Since the ray is horizontal both at depth C and at depth A, it is evident from Snell's Law that the speed of sound is the same at both depths. The depth C can be located on the velocity profile by dropping a vertical line from the point A to the intersection with the segment BD. Thus the limiting ray from S is confined to a channel extending from A to C, and this is true for a source located at any depth within the channel.

Any ray which leaves the source at a steeper angle than SP'Q'R'... will cross the upper boundary and leave the channel. Also any ray, such as SP''Q''R'' ..., which leaves the source at a correspondingly steep downward angle will vertex at a depth Q'' below C and upon returning

upward will leave the channel at T". (If the ray is sufficiently steep it may even cross the lower boundary and leave the channel there.)

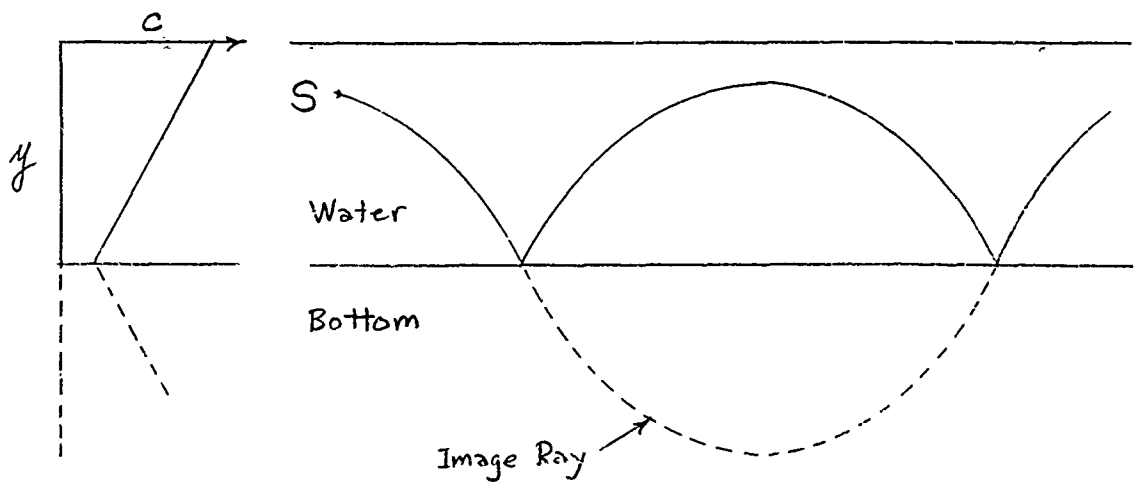
In summary, a region in which the sound speed passes through a minimum gives rise to a sound channel. The depth of the minimum is the axis of the channel. It tends to "attract" rays. To determine the thickness of the channel, one examines the adjacent maxima of the sound speed profile above and below the minimum. The width of the channel is governed by the shallower of the two maxima, i.e., the one which has the smaller value of c_{max}. The channel extends from the depth of this maximum to the depth on the other side of the channel axis where the sound speed has this same value.

As indicated previously, a positive gradient layer at the surface behaves as a sound channel, due to reflections from the surface. This is easy to see if the reflected rays are "re-reflected," so that they travel through an imaginary layer above the surface which is a mirror image of the actual layer below the surface.

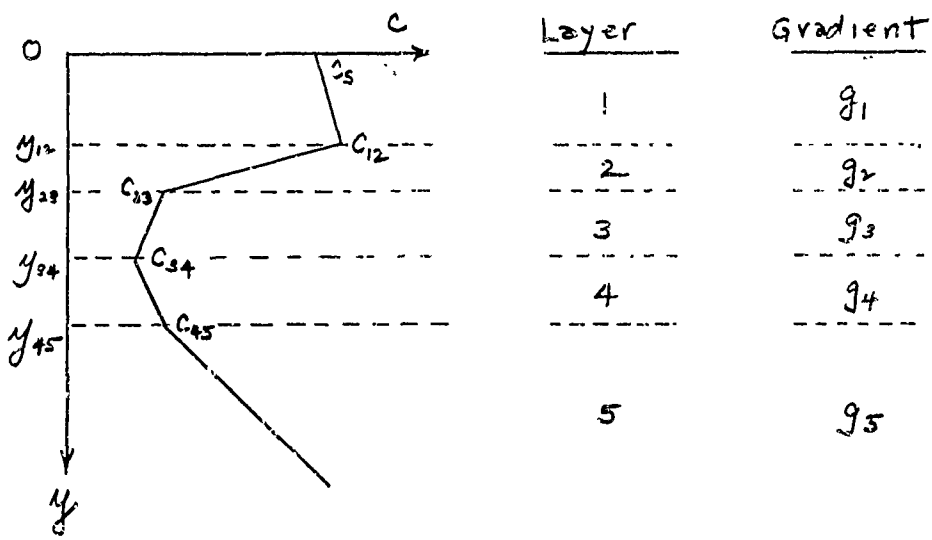


A similar situation can occur in shallow water when a negative gradient exists all the way to the bottom, the combination of downward refraction and bottom reflections giving rise to a sound channel.

In general, however, bottom losses are far larger than surface losses, and the propagation characteristics of such a channel are usually inferior to those of a surface channel.



It will be useful to implement a portion of the preceding qualitative discussion by deriving the appropriate mathematical formulas. For this purpose we shall assume that the velocity profile has been specified in terms of the sound speed at each of the layer boundaries. The value at the surface ($y = 0$) is c_s. The value at the boundary y_{12} between the first and second layers is c_{12}; the value at y_{23} is c_{23}; the value at y_{34} is c_{34}, and so on. The sound speed at any intermediate depth can be computed from the appropriate linear formula. For example,



the sound speed at any depth y in the second layer is obtained from

$$c - c_{12} = g_2 (y - y_{12}) \tag{459}$$

where

$$g_2 = \frac{c_{23} - c_{12}}{y_{23} - y_{12}} \tag{460}$$

We shall suppose that the velocity profile has the characteristics shown in the preceding figure and that the sound source is located at a depth y_o in the first layer. The speed of sound at this depth is

$$c_o = c_s + g_1 y_o \tag{461}$$

Vertex

Consider first the problem of locating a vertex in the first layer. The vertex occurs at the point where $\theta = 0$. The values of x, s, and t at the vertex are obtained by setting $\theta = 0$ in equations (456a), (456c), and (456d). For example, the horizontal range x_m is

$$x_m = - \frac{c_o \sin \theta_o}{g_1 \cos \theta_o}$$

From the sketch below it is seen that g_1 is positive and θ_o is negative, so that x turns out to be positive. The depth y_m of the vertex can be computed by setting $\theta = 0$ in (456b). As an alternative method, we note that the speed of sound at the vertex is

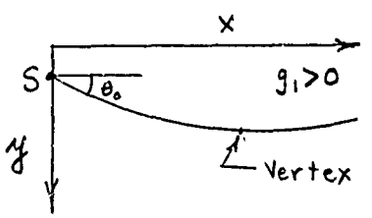
$$c_v = \frac{c_o}{\cos \theta_o}$$

The depth of the vertex may be computed from the equation (461) for the segment of the velocity profile in the first layer. Thus,

$$c_v = c_s + g_1 y_m \tag{461a}$$

or

$$y_m = \frac{c_v - c_s}{g_1} \tag{462}$$



A glance at the velocity profile shown in the above figure indicates that if the ray vertexes in the first layer, its vertex speed c_v is less than the maximum value c_{12} at the boundary, and the ray will (theoretically) never leave the first layer. Any region of the velocity profile where the sound speed exceeds c_v acts as a barrier which the ray cannot cross. (On the other hand, if the source were located below the boundary 12 and a ray were to vertex in, say, the second layer with a vertex speed c_v slightly smaller than c_{12}, this ray would be confined to layers 2, 3, 4, and 5, and would vertex at a depth in layer 5 where the sound speed is c_v.)

Limiting Ray

The limiting ray from a source in layer 1 vertexes at the boundary y_{12}. Its vertex speed is therefore

$$c_v = c_{12}$$

The problem of locating the limiting ray consists of determining its initial angle, which we shall call θ_L. This is found directly from Snell's Law, where

$$c = c_{12};\ \theta = 0;\ \theta_o = \theta_L$$

Thus

$$\cos \theta_L = \frac{c_o}{c_{12}} \tag{463}$$

The values of x, s, and t are determined by setting $\theta_o = \theta_L$ and $\theta = 0$ in the appropriate equations.

Ray Path in Other Layers

If the initial angle is steep enough, the ray will cross into the layers below. The value of x at the boundary between layers 1 and 2 is

$$x_{12} = \frac{c_o}{g_1 \cos \theta_o} (\sin \theta_{12} - \sin \theta_o) \tag{464}$$

To determine the value of x at any depth in the second layer, we return to the original solution (451a). The vertex velocity c_v remains constant in all layers, so that we may retain the factor $c_o/\cos\theta_o$. The initial conditions for the ray in the second layer are

$$x = x_{12} \text{ when } \theta = \theta_{12}$$

Therefore the increment of x in the second layer is

$$x - x_{12} = \frac{c_o}{g_2 \cos\theta_o}(\sin\theta - \sin\theta_{12}) \qquad (465)$$

At the boundary between layers 2 and 3, (465) becomes

$$x_{23} - x_{12} = \frac{c_o}{g_2 \cos\theta_o}(\sin\theta_{23} - \theta_{12}) \qquad (465a)$$

The increment in the third layer is then

$$x - x_{23} = \frac{c_o}{g_3 \cos\theta_o}(\sin\theta - \sin\theta_{23})$$

and so on. The cumulative value is the sum of all the increments. For example, in the third layer,

$$x = x_{12} + (x_{23} - x_{12}) + (x - x_{23}) \qquad (466)$$

The corresponding values of s and t are obtained in a similar manner.

In tracing a ray from the source outward, we continue the process indicated above until either a depth is reached at which the sound speed exceeds c_v, in which case the ray vertexes and returns toward the source depth, or else the surface or bottom is reached, in which case the ray is reflected. These are some of the ground rules which one incorporates into a computer program.

10. Spreading Loss

Ray theory enables us not only to trace ray paths but also to compute the spreading loss. It will be recalled that the spreading loss at any point on a ray is 10 log of the ratio of the intensity at the standard 1-yard reference distance to the intensity at the point in question, assuming

that the effects of attenuation and other losses are neglected. To estimate the intensity ratio we take a small bundle of rays and observe how they spread out. Let I_1 be the intensity at the 1-yard point, that is, at the point where

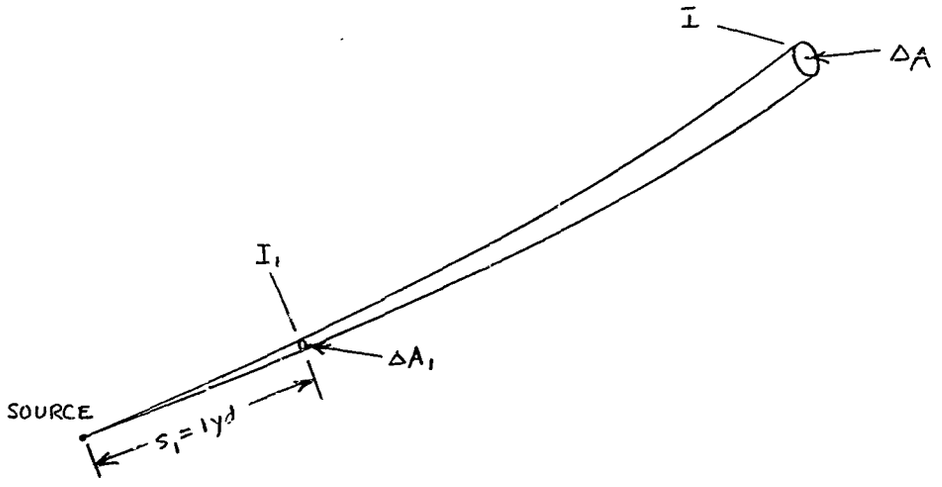
$$s = s_1 = 1 \text{ yd}$$

and I be the intensity at the distant point. Let ΔA_1 be the small area intercepted by the bundle at the 1-yard point and ΔA the corresponding area at the distant point. If no energy is lost by absorption or scattering, the total energy in the bundle remains the same at all distances from the source, that is,

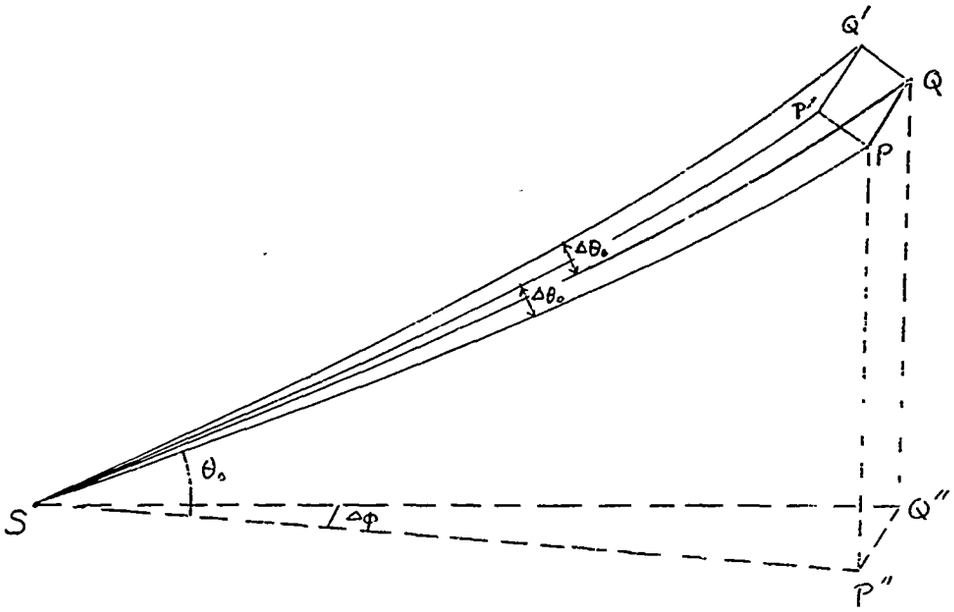
$$I_1 \, \Delta A_1 = I \Delta A$$

or

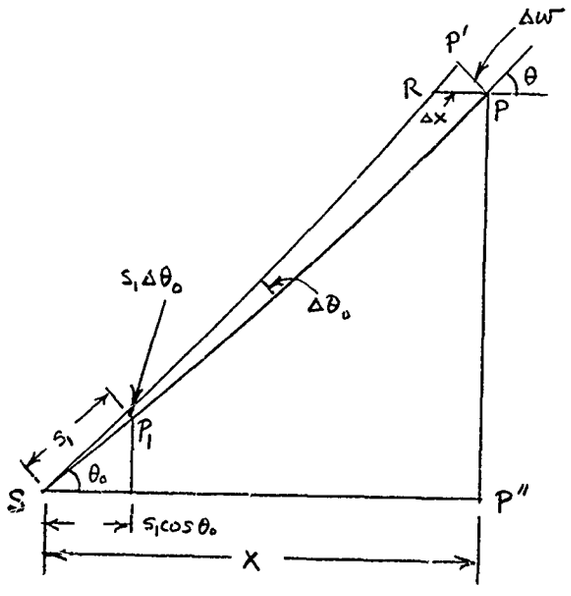
$$\frac{I_1}{I} = \frac{\Delta A}{\Delta A_1} \tag{467}$$

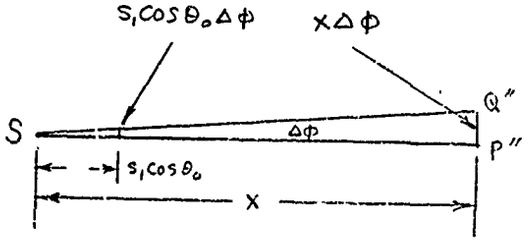


To evaluate this ratio we assume that there are no horizontal gradients and that the rays spread out equally in all directions in azimuth. We select a bundle having a rectangular cross section defined by the four rays SP, SQ, SP' and SQ', as indicated in the sketch below. When viewed in elevation, the rays SP and SQ make an angle θ_0 with the horizontal, the angular width of the bundle at the source S being $\Delta\theta_0$.



When projected onto the horizontal plane, the bundle is enclosed within the two straight lines SP" and SQ" separated by an angle $\Delta\phi$. The one-yard reference distance s_1, which for all practical purposes may be regarded as a straight line segment, is shown as SP_1. The horizontal





coordinate of P_1 is the projection of SP_1 on the horizontal plane, namely, $s_1 \cos \theta_o$. The dimensions of the little rectangular area ΔA_1 are therefore $s_1 \cos \theta_o \Delta\phi$ and $s_1 \Delta\theta_o$; hence,

$$\Delta A_1 = s_1^2 \cos \theta_o \Delta\theta_o \Delta\phi$$

If we denote the width of the bundle at the far point by Δw, the area ΔA is

$$\Delta A = x \Delta w \Delta\phi$$

and the intensity ratio is

$$\frac{I_1}{I} = \frac{x\Delta w}{s_1^2 \cos \theta_o \Delta\theta_o}$$

We must now express Δw in terms of things we know. If we consider the change in x when y is held constant and θ_o increases by $\Delta\theta_o$, we see that x decreases by an amount equal to the length of the horizontal segment PR. Furthermore, if the ray angle at P (and also at P', since $\Delta\theta_o$ is exceedingly small) is θ, then in the right triangle PP'R we find that

$$\Delta w = - \Delta x \sin \theta$$

and the intensity ratio is

$$\frac{I_1}{I} = - \frac{x \sin \theta}{s_1^2 \cos \theta_o} \left. \frac{\Delta x}{\Delta\theta_o} \right|_{y = \text{const.}}$$

In the limit as $\Delta\theta_o$ goes to zero this becomes

$$\frac{I_1}{I} = - \frac{x \sin \theta}{s_1^2 \cos \theta_o} \frac{\partial x}{\partial \theta_o} \tag{468}$$

The spreading loss is

$$N_{spr} = 10 \log \frac{I_1}{I} \tag{469}$$

Equation (468) is a general result whose validity is limited only by the requirement that the sound speed must not vary in a horizontal direction. We shall now apply it to the case of constant-gradient layers.

Single Layer

When the ray path lies entirely within a single layer of constant velocity gradient, the horizontal range is expressed by (456a). In taking the derivative of x with respect to θ_o we must recognize that the angle θ at the far end of the ray is also a function of θ_o. Thus,

$$\frac{\partial x}{\partial \theta_o} = \frac{c_o}{g}\left[\frac{\sin\theta_o}{\cos^2\theta_o}(\sin\theta - \sin\theta_o) + \frac{1}{\cos\theta_o}\left(\cos\theta\frac{\partial\theta}{\partial\theta_o} - \cos\theta_o\right)\right] \qquad (470)$$

The derivative $\partial\theta/\partial\theta_o$ is readily evaluated from Snell's Law,

$$\cos\theta = \frac{c}{c_o}\cos\theta_o$$

Since we are interested in the rate of change at constant depth, c is a constant.

$$\sin\theta\frac{\partial\theta}{\partial\theta_o} = \frac{c}{c_o}\sin\theta_o = \frac{\cos\theta\sin\theta_o}{\cos\theta_o}$$

or

$$\frac{\partial\theta}{\partial\theta_o} = \frac{\sin\theta_o\cos\theta}{\cos\theta_o\sin\theta} \qquad (471)$$

Insertion of (471) into (470) yields

$$\frac{\partial x}{\partial\theta_o} = -\frac{c_o}{g\cos^2\theta_o\sin\theta}(\sin\theta - \sin\theta_o) \qquad (472)$$

or, from (465a)

$$\frac{\partial x}{\partial\theta_o} = -\frac{x}{\cos\theta_o\sin\theta} \qquad (472a)$$

The desired result is obtained by substituting (472a) into (468)

$$\frac{I_1}{I_o} = \frac{x^2}{s_1^2\cos^2\theta_o} \qquad (473)$$

From this result we see that in a single constant-gradient layer the intensity is inversely proportional to the square of the horizontal range. Incidentally, in the limiting case where the velocity gradient is zero and the rays become straight lines, the distance s along the arc is

$$s = \frac{x}{\cos\theta_o}$$

and the intensity ratio boils down to the standard inverse square law.

Multiple Layers

We shall illustrate the general case by considering three layers. To obtain $\partial x/\partial\theta_o$, we take the derivatives of each of the three terms of (466). The first term has already been derived in (472a), where all we need to do is substitute x_{12} and θ_{12} for x and θ. Thus

$$\frac{\partial x_{12}}{\partial \theta_o} = - \frac{x_{12}}{\cos \theta_o \sin \theta_{12}} \tag{474}$$

The derivative of the second term (465a) is

$$\frac{\partial(x_{23} - x_{12})}{\partial \theta_o} = \frac{c_o}{g_2} \left[\frac{\sin \theta_o}{\cos^2 \theta_o} (\sin \theta_{23} - \sin \theta_{12}) \right.$$

$$\left. + \frac{1}{\cos \theta_o} \left(\cos \theta_{23} \frac{\partial \theta_{23}}{\partial \theta_o} - \cos \theta_{12} \frac{\partial \theta_{12}}{\partial \theta_o} \right) \right]$$

The derivatives $\partial\theta_{12}/\partial\theta_o$ and $\partial\theta_{23}/\partial\theta_o$ are obtained by substituting the appropriate values into (471), namely,

$$\frac{\partial \theta_{12}}{\partial \theta_o} = \frac{\sin \theta_o \cos \theta_{12}}{\cos \theta_o \sin \theta_{12}} \; ; \; \frac{\partial \theta_{23}}{\partial \theta_o} = \frac{\sin \theta_o \cos \theta_{23}}{\cos \theta_o \sin \theta_{23}}$$

The result is

$$\frac{\partial(x_{23} - x_{12})}{\partial \theta_o} = - \frac{c_o \sin \theta_o}{g_2 \cos^2 \theta_o \sin \theta_{12} \sin \theta_{23}} (\sin \theta_{23} - \sin \theta_{12})$$

or

$$\frac{\partial(x_{23} - x_{12})}{\partial \theta_o} = - \frac{(x_{23} - x_{12}) \sin \theta_o}{\cos \theta_o \sin \theta_{23} \sin \theta_{12}} \tag{474a}$$

Similarly, the third term yields

$$\frac{\partial(x - x_{23})}{\partial \theta_o} = - \frac{(x - x_{23}) \sin \theta_o}{\cos \theta_o \sin \theta \sin \theta_{23}} \tag{474b}$$

The derivative of (466) is the sum of (474), (474a), and (474b),

$$\frac{\partial x}{\partial \theta_o} = - \frac{\sin \theta_o}{\cos \theta_o} \left[\frac{x_{12}}{\sin \theta_{12} \sin \theta_o} + \frac{x_{23} - x_{12}}{\sin \theta_{23} \sin \theta_{12}} + \frac{x - x_{23}}{\sin \theta_{23} \sin \theta} \right]$$

and the intensity ratio (468) becomes

$$\frac{I_1}{I} = \frac{x \sin\theta_o \sin\theta}{s_1^2 \cos^2\theta_o}\left[\frac{x_{12}}{\sin\theta_{12}\sin\theta_o} + \frac{x_{23} - x_{12}}{\sin\theta_{23}\sin\theta_{12}} + \frac{x - x_{23}}{\sin\theta_{23}\sin\theta}\right] \quad (475)$$

The generalization of (475) to the case of more than three layers should now be evident. For two layers (475) takes the form

$$\frac{I_1}{I} = \frac{x}{s_1^2 \cos^2\theta_o}\left[x_{12}\frac{\sin\theta}{\sin\theta_{12}} + (x - x_{12})\frac{\sin\theta_o}{\sin\theta_{12}}\right] \quad (475a)$$

Here again we see that in the limiting case of straight line propagation, where $g_1 = g_2 = g_3 = 0$, all the angles are equal to θ_o and equation (475) boils down to the classical inverse square law. The presence of the refractive layers modifies the inverse square law by multiplying the increment of x in each layer by a factor involving the sines of the angles. For example, in the second layer the increment $(x_{23} - x_{12})$ is multiplied by the factor

$$\frac{\sin\theta_o \sin\theta}{\sin\theta_{23}\sin\theta_{12}}$$

These factors, however, are capable of producing drastic effects. Let us examine a few of these effects.

Focusing Effect in a Sound Channel

Consider the case of two layers where the upper (first) layer has a negative velocity gradient and the lower (second) layer has a positive gradient. We have already seen that this is a sound channel. Let the source be placed in the first layer and consider the ray which leaves the source horizontally, i.e., $\theta_o = 0$. The intensity ratio (475a) becomes in this case

$$\frac{I_1}{I} = \frac{x\, x_{12} \sin\theta}{s_1^2 \sin\theta_{12} \cos^2\theta_o}$$

After this ray crosses the boundary it will vertex in the second layer. At the vertex the angle θ is zero. Therefore

$$\frac{I_1}{I} = 0, \quad \text{when } \theta_o = \theta = 0$$

Here we see that if the ray leaves the source horizontally, the intensity will be infinite at every vertex in the channel. These are focal points where the rays converge after initially diverging from the source.

Although the mathematics becomes somewhat messy, it can be shown that focal points occur also for rays which leave the source at other angles. The locus of such focal points is called a caustic. The intensity at a caustic is very high and signals can be expected to be very strong.

Limiting Rays

Consider now the opposite two-layer case in which the positive gradient occurs in the upper (first) layer. The limiting ray from the source will graze the layer boundary, so that θ_{12} will be zero. Let θ_L denote the initial angle of the limiting ray. (θ_L is negative in this case.) The intensity ratio at the point of tangency with the boundary is obtained by setting $\theta_o = \theta_L$, $\theta = \theta_{12}$, and $x = x_{12}$ in (473) or (475a).

$$\frac{I_1}{I} = \frac{x^2_{12}}{s_1^2 \cos^2 \theta_L}$$

At this point the ray splits into two branches. Since the near branch remains within the first layer, the single-layer formula (473) holds throughout. To determine the intensity in the far branch beyond the point of tangency, we see that both x and θ increase, while θ_{12} remains zero. Clearly the intensity ratio (475a) is infinite at all values of x beyond x_{12}. This analysis shows that as we approach the shadow zone from the near side, the intensity remains finite all the way to the limiting ray. On the other hand, as we approach the shadow zone from the far side, the intensity approaches zero.

Spurious Caustics and Shadow Zones Due to Constant-Gradient Approximation

The procedure of approximating a continuously changing gradient by a number of constant-gradient layers leads in certain instances

to focusing effects and in other instances to small shadow zones which are not actually present in the ocean. We have already mentioned the behavior in the vicinity of a sound speed maximum. A discussion of other limitations of the constant-gradient approximation will be given in a later section.

11. The Naval Air Development Center Ray Tracing Program

A comprehensive ray tracing program, based on the constant-gradient approximation, is currently available on the IBM 650 digital computer of the Aeronautical Computer Laboratory. The program can handle up to 50 different constant-gradient layers. The input data for the velocity profile may be specified in terms of either the temperature or the sound speed at each of the layer boundaries, including the surface and the bottom, and may be expressed in any of the following ways:

(a) Temperature in °F; depth in feet.

(b) Temperature in °C; depth in meters.

(c) Sound speed in ft/sec; depth in feet.

(d) Sound speed in meters/sec; depth in meters.

If the input data are expressed in metric units, the machine automatically converts to English units. When the inputs are temperatures, the machine assumes a nominal salinity of 34 parts/1000 and computes the sound speed from equation (447). The computer will also accept salinity inputs for use with equation (447). Other input data include the following:

(a) Source depth.

(b) Acoustic frequency. The computer can handle up to four frequencies at a time.

(c) Depths at which outputs are desired.

(d) Instructions for termination of rays.

The program incorporates provision for both bottom and surface reflections. The bottom loss is programmed as a function of the acoustic frequency and the grazing angle of the ray, in accordance with a

set of equations representing average bottom conditions. No comparable provision has been included for surface loss, but any desired constant value may be inserted. When a ray strikes the surface or the bottom, it is reflected and continues on, or if it passes through a maximum or a minimum, it likewise continues on, cycle after cycle, until the machine is told to stop.

The attenuation coefficient is programmed as a function frequency and temperature in accordance with (316), (317) and (318). The theoretical attenuation loss is evaluated from the integral

$$N_{att} = \int_0^s \alpha ds$$

where the attenuation coefficient must be expressed as a function of the distance s along the ray arc. The machine program approximates the integral by a sum

$$\sum_i \alpha_i \Delta s_i$$

where each α_i is a constant whose value is the average of the values at the beginning and end of each interval of computation.

The following combinations of quantities may be computed at the option of the customer:

(a) y, x, θ

(b) The quantities in (a) plus N_{spr}

(c) The quantities in (b) plus s and t

(d) The quantities in (c) plus N_w

The outputs are automatically tabulated at all layer boundaries and at all vertices. Reflections at the surface and bottom (in addition to maxima and minima) are considered to be vertices. Outputs may also be computed and tabulated in equal depth increments between any two desired limits. Any desired increment may be specified. In addition, provision is

made for specifying up to 10 arbitrary "special" depths. It is also possible to delay computation at the incremental and special depths until after a specified number of vertices have been passed.

Two different modes of computation are available: families of rays and limiting rays.

(a) In the computation of ray families, the initial angle of each ray is specified as an input. The procedure is to specify the initial angle of the first ray, the increment by which the initial angle is to be increased for successive rays, and the value for the final ray.

(b) There are several options in the computation of limiting rays. A velocity profile may have more than one maximum point. In addition to maxima occurring within the volume of the ocean, the surface will be a maximum point if the surface layer has a negative gradient, and the bottom likewise if the lowest layer has a positive gradient. The customer may specify the boundary at which he wishes the limiting ray computed. If no such specification is made, the machine program selects the depth at which the largest value of c in the entire profile occurs.

Computations may be made for only the near branch of the limiting ray, or only the far branch, or both.

The following options are available for terminating the computation of a ray:

(a) After a specified number of vertices (including surface and bottom reflections) have been passed.

(b) When a given depth is reached after a specified number of vertices have been passed.

(c) When a specified value of horizontal range x is reached.

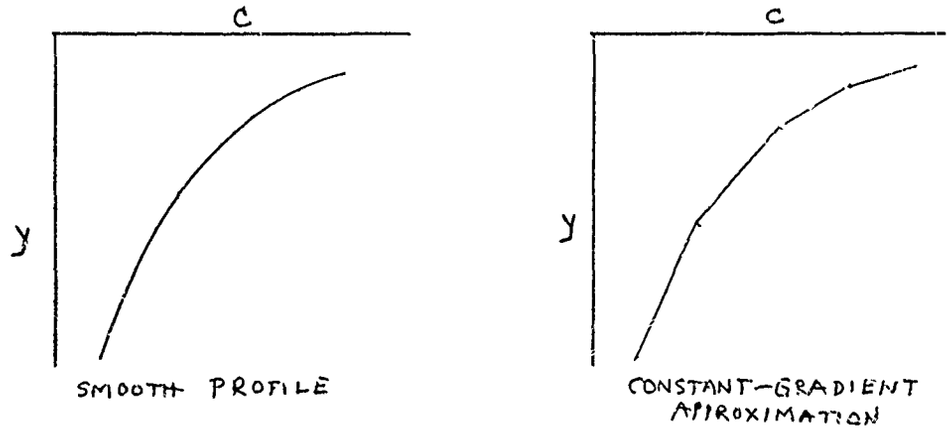
(d) When a specified value of propagation loss at a specified frequency has been attained, or, if the propagation loss is not being computed, the spreading loss may be used as the criterion for stopping.

12. Limitations of the Constant-Gradient Approximation

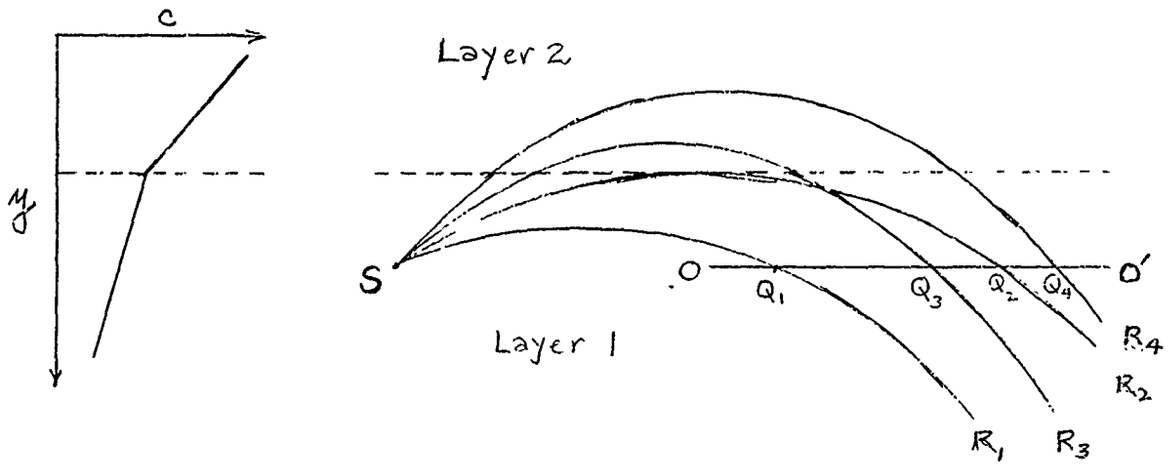
We have already seen that a point of maximum sound speed between two adjacent constant-gradient layers leads to the prediction of shadow zones in which the spreading loss is theoretically infinite. Although such a concept is a mathematical fiction, since ideally sharp boundaries between adjacent layers never occur in the real ocean, it is nevertheless a reasonably good approximation in many instances.

There are other situations, however, in which the constant-gradient layer approximation predicts some very strange intensity patterns which do not exist at all in the ocean. These occur when a smoothly curving velocity profile is approximated by a number of straight-line segments. If the sound speed in this portion of the profile exceeds the sound speed at the source, those rays which vertex at depths in the vicinity of a layer boundary will exhibit some wild fluctuations in spreading loss, and the strange thing about it is that the usual limiting process of approximating the smooth curve by taking more and more straight-line segments of shorter and shorter length, does not work. It merely results in more and more wild fluctuations.

Two different types of behavior occur, depending on whether the curvature of the velocity profile is concave or convex. A typical example of the concave case is the situation in which a strongly negative gradient gradually tapers off with increasing depth, as illustrated in the left-hand diagram below. On the right is shown a straight-line segment approximation.



Let us investigate what happens at one of the artificial layer boundaries. Consider the case of two negative-gradient layers illustrated below. The lower layer, containing the source S, has the



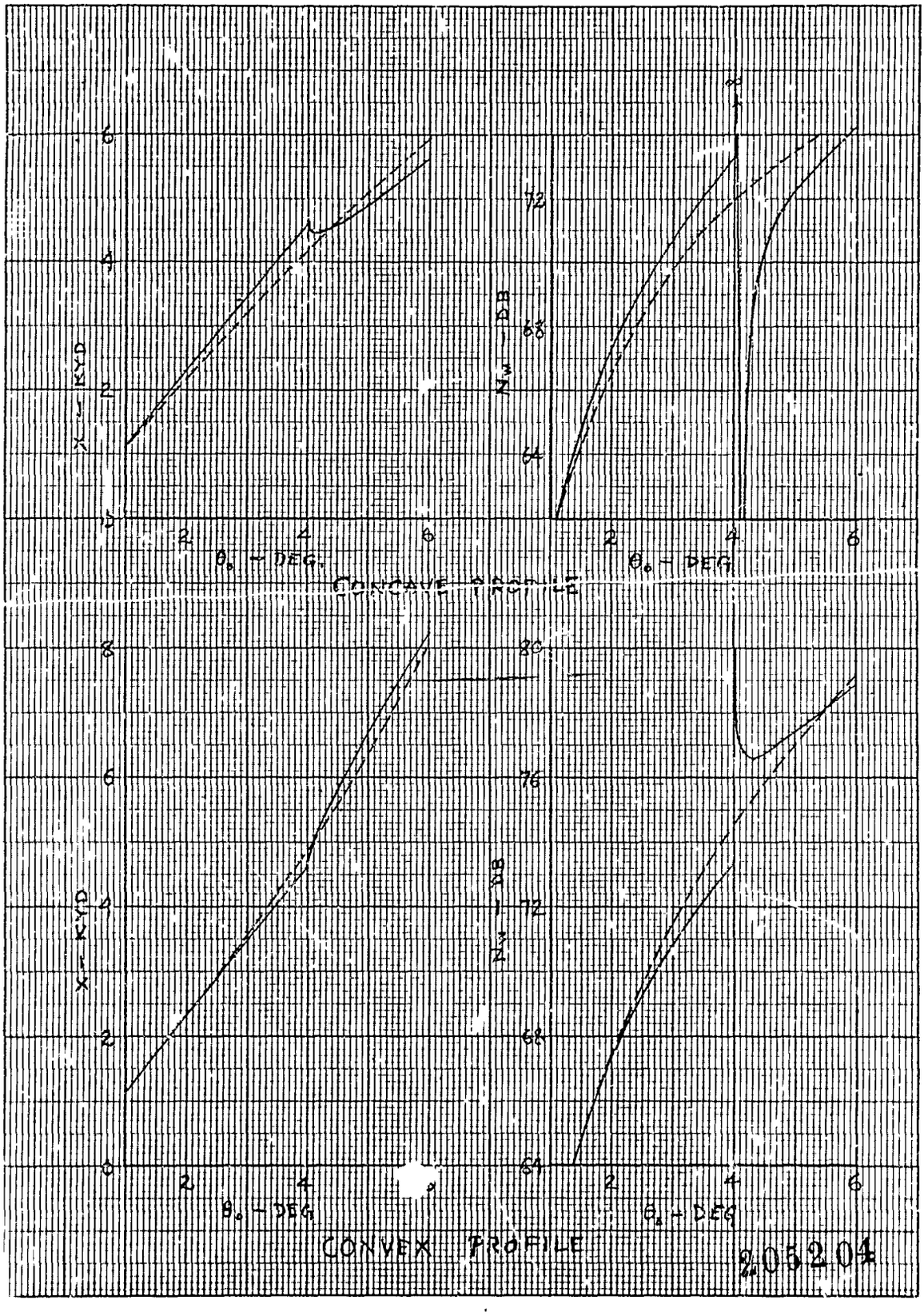
weaker gradient. We shall call this the first layer. The upper layer is the second. Rays from the source may be grouped into two types, depending upon whether they lie below or above the ray SR_2 which grazes the boundary between the two layers. We shall call this ray a limiting ray (although it is not a limiting ray in the strict sense). All rays below the limiting ray remain entirely in the first layer and exhibit no unusual characteristics. SR_1 is such a ray. Rays of the second type, which lie above the limiting ray, penetrate into the second layer where they are refracted more sharply because of the stronger gradient. They therefore return to the first layer at a somewhat shortened range and cut across some of the rays of the SR_1 type. SR_3 and SR_4 are typical examples.

Let us now select some fixed depth below the boundary, such as OO', and investigate the behavior of the horizontal range x at this depth when the initial angle θ_o is varied. For all rays below the limiting ray the horizontal range increases smoothly with θ_o, as indicated by the crossing points Q_1 and Q_2. As soon, however, as the rays begin to penetrate into the second layer, the horizontal range takes a sudden drop,

as indicated by the crossing Q_3. As θ_o is further increased, the range passes through a minimum and begins to increase again, as indicated by Q_4. The behavior is shown in the accompanying graph. The infinite slope of the curve immediately beyond the limiting angle signifies that the intensity is zero and the spreading loss infinite. The minimum point of the curve is a point of infinite intensity, since here the horizontal range does not vary at all with θ_o. This is a focal point and the spreading loss is negatively infinite. In the small sector of θ_o between the limiting ray and the "focal ray" the spreading loss varies all the way from $+\infty$ to $-\infty$. Furthermore, it will be noted that every point in the water in this horizontal range interval is reached by three rays - one below the limiting ray and two above. No shadow zone exists, even though we computed zero intensity at one point for one of the three rays. Beyond the minimum point the horizontal range increases again and the spreading loss ultimately settles down to reasonable values. The smooth dashed curves show the range and spreading loss computed from the continuous velocity profile.

At different depths the focus occurs at different ranges. The constant-gradient approximation thus generates a caustic.

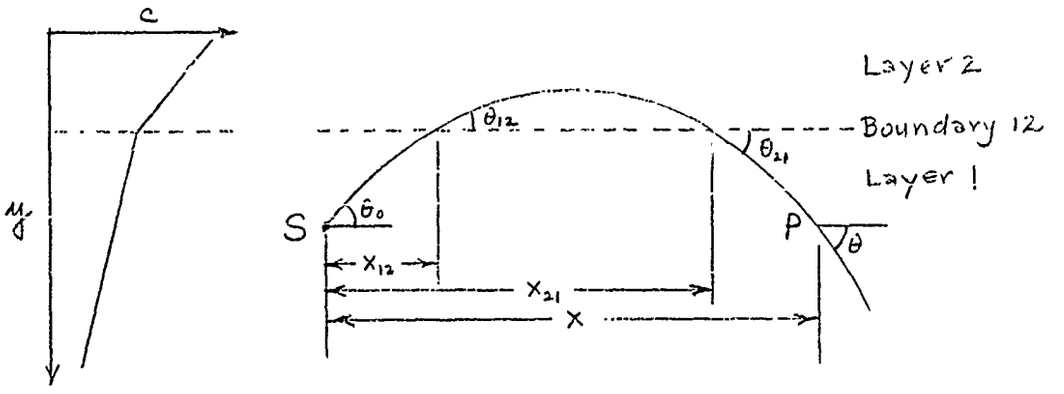
In the reverse case, where the negative gradient becomes stronger as the depth increases, the velocity profile has a convex curvature. The corresponding two-layer approximation for this case shows just the opposite effect. Here the upper layer has the weaker gradient. Rays which penetrate above the layer boundary suffer less refraction and return to the boundary at longer ranges. Plots of horizontal range and spreading loss vs. θ_o for this case are shown on an accompanying graph. Here it is seen that beyond the limiting ray the x curve takes a jump beginning with infinite slope, and gradually settles down. Meanwhile the spreading loss jumps to infinity, then decreases to a minimum value, increases again, and gradually settles down. The infinite loss in this case corresponds to an artificial



205204

shadow zone. The rays do not overlap. The two dashed curves show the range and spreading loss computed from the continuous profile.

These effects may be shown mathematically as follows. Let the gradients in the two layers be g_1 and g_2. The source S is in the first (lower) layer. Consider a ray which penetrates into the second layer. Let the sound speed at the source be c_o. As indicated in the sketch, θ_o is the initial angle of the ray, θ_{12} and X_{12} are angle and horizontal range



at the first crossing of the boundary 12, θ_{21} and X_{21} at the second crossing, and θ and x at the point P in the first layer.

The range X is the sum of three increments in the manner of equation (466). Inserting expressions of the type (456a) and (464), we obtain

$$x = \frac{c_o}{\cos \theta_o}\left[\frac{1}{g_1}(\sin \theta_{12} - \sin \theta_o) + \frac{1}{g_2}(\sin \theta_{21} - \sin \theta_{12}) + \frac{1}{g_1}(\sin \theta - \sin\theta_{21})\right]$$

From Snell's Law it is evident that

$$\theta_{21} = -\theta_{12}$$

Substituting this value into the above equation and combining terms, we obtain

$$x = \frac{c_o}{\cos \theta_o} \left[\frac{1}{g_1} (\sin \theta - \sin \theta_o) + 2\left(\frac{1}{g_1} - \frac{1}{g_2} \right) \sin \theta_{12} \right] \qquad (476)$$

The first term in (476) gives the value of x which would be obtained if the second layer had the same gradient as the first. The second term shows the effect of the discontinuity in the gradient.

To simplify matters, let us assume the point P to be at the same depth as the source, so that

$$\theta = - \theta_o$$

Also, since the gradients are both negative, let us introduce the positive values

$$g_1' = - g_1 \text{ and } g_2' = - g_2$$

Equation (476) then becomes

$$x = \frac{2 c_o \tan \theta_o}{g_1'} \left[1 - \left(1 - \frac{g_1'}{g_2'} \right) \frac{\sin \theta_{12}}{\sin \theta_o} \right] \qquad (476a)$$

The corresponding expression for the intensity ratio can be obtained by inserting the appropriate expressions for the increments of x into (475). Thus

$$\frac{I_1}{I} = \frac{x \sin \theta_o \sin \theta}{s_1^2 \cos^2 \theta_o} \left[\frac{c_o}{g_1 \cos \theta_o} \frac{\sin \theta_{12} - \sin \theta_o}{\sin \theta_{12} \sin \theta_o} + \frac{c_o}{g_2 \cos \theta_o} \frac{\sin \theta_{21} - \sin \theta_{12}}{\sin \theta_{21} \sin \theta_{12}} \right.$$

$$\left. + \frac{c_o}{g_1 \cos \theta_o} \frac{\sin \theta - \sin \theta_{21}}{\sin \theta \sin \theta_{21}} \right]$$

$$= \frac{x c_o \sin \theta_o \sin \theta}{s_1^2 \cos^3 \theta_o} \left[\frac{1}{g_1} \left(\frac{1}{\sin \theta_o} - \frac{1}{\sin \theta} \right) - \left(\frac{1}{g_1} - \frac{1}{g_2} \right) \frac{2}{\sin \theta_{12}} \right]$$

And if $\theta = - \theta_o$, we obtain

$$\frac{I_1}{I} = \frac{2 x c_o \sin \theta_o}{s_1^2 g_1' \cos^3 \theta_o} \left[1 - \left(1 - \frac{g_1'}{g_2'} \right) \frac{\sin \theta_o}{\sin \theta_{12}} \right] \qquad (477)$$

The first term in (477) gives the intensity ratio which would exist if there were only one layer. The second term shows the effect of the discontinuity.

The critical factor is the ratio of the two sines in (476a) and (477), the one being the reciprocal of the other.

To investigate this ratio, let

$$\theta_o = \theta_L + \Delta\theta \tag{478}$$

where θ_L is the initial angle of the limiting ray. Since the limiting ray is tangent to the boundary 12, where we shall assume the sound speed to be c_{12}, we see from Snell's Law,

$$\frac{c_{12}}{c_o} = \frac{1}{\cos\theta_L} = \frac{\cos\theta_{12}}{\cos\theta_o}$$

so that

$$\sin\theta_{12} = \sqrt{1 - \frac{\cos^2\theta_o}{\cos^2\theta_L}} = \frac{\sqrt{\cos^2\theta_L - \cos^2\theta_o}}{\cos\theta_L}$$

which may be transformed to

$$\sin\theta_{12} = \frac{\sqrt{\sin(2\theta_o + \theta_L)\sin(\theta_o - \theta_L)}}{\cos\theta_L}$$

Insertion of (478) yields the ratio

$$\frac{\sin\theta_{12}}{\sin\theta_o} = \sqrt{\frac{\sin(2\theta_L + \Delta\theta)\sin\Delta\theta}{\cos\theta_L \sin(\theta_L + \Delta\theta)}} \tag{479}$$

Here we see that as θ_o approaches the limiting ray angle θ_L, i.e., as $\Delta\theta$ approaches zero, the ratio of sines (479) also approaches zero. The reciprocal of this ratio, which appears in the formula (477) for the intensity ratio, becomes infinite. The horizontal range is thus changing infinitely fast with respect to θ_o.

In the concave case discussed earlier, the second layer has the stronger gradient. Hence

$$g_2' > g_1'$$

or

$$1 - \frac{g_1'}{g_2'} > 0$$

In this case x takes a sudden drop, as may be seen from (476a), and the sound rays double back on themselves. While this is occurring the second term in (477) is numerically larger than the first and the intensity ratio is negative.† As θ_o increases still further, the second term in (477) becomes numerically smaller until a point is reached where the two terms cancel. At this point x becomes stationary and the intensity ratio goes to zero; the spreading loss is $-\infty$. This is the focal point. As θ_o continues to increase, x begins to increase again and the spreading loss begins to approach its normal value.

In the convex case the upper layer has a weaker gradient and g_2' is less than g_1'. In this case the factor

$$1 - \frac{g_1'}{g_2'} < 0$$

is negative and x experiences a sudden increase. The spreading loss shoots up to infinity. But now a further increase in θ_o continues to increase the range. The second term in (477) decreases rapidly (from infinity) at first, then more slowly. The intensity ratio thus decreases at first, but later increases again, due to the increase in range.

Mr. M. A. Pedersen of the Navy Electronics Laboratory has investigated this problem extensively and has shown * that when a continuous concave gradient is approximated by more and more constant-gradient layers, the approximate horizontal range approaches the true value in a very peculiar fashion. The curve acquires more and more scallops. To be sure, the scallops become smaller and smaller in size but, however small, they nevertheless remain, and each scallop gives rise to infinite values of spreading loss. In the limit, therefore, the horizontal range is accurately computed, but the spreading loss computation becomes worthless.

†The negative sign may be ignored in computing spreading loss. Use $N_{spr} = 10 \log \left|\frac{I_1}{I}\right|$.

* J.A.S.A., Vol 33, p. 465 (1961)

As a solution to this problem Pedersen has suggested a quite different approach whereby the profile is approximated by curvilinear segments which are fitted together with no discontinuities in slope.

13. Method of Curvilinear Gradients

On first thought the simplest form of curvilinear segment would appear to be a parabola, such as

$$c = A + By + Cy^2$$

However, this function does not lead to a practical analytic solution. It turns out that a reasonably simple solution can be obtained from a function of the form

$$c^2 = \frac{c_a^2}{1 - k_a (y - y_a)^2} \tag{480}$$

This function gives three different types of curves, depending upon the algebraic signs of the constants. These are:

Type I: $c_a^2 > 0, \quad k_a > 0$

Type II: $c_a^2 > 0, \quad k_a < 0$

Type III: $c_a^2 < 0, \quad k_a > 0$

Each of the three types leads to a different form of solution.

The basic differential equation employed in this method is (451), which may be rewritten in the form

$$dx = - \cot \theta \, dy \tag{481}$$

The angle θ may be expressed as a function of y by Snell's Law

$$\cos \theta = \frac{c}{c_v}$$

Substitution of (480) yields

$$\cos \theta = \sqrt{\frac{c_a^2}{c_v^2 [1 - k_a (y - y_a)^2]}} \tag{482}$$

$$\sin \theta = \pm \sqrt{\frac{c_v^2 - c_a^2 - c_v^2 k_a (y - y_a)^2}{c_v^2 [1 - k_a (y - y_a)^2]}} \tag{482a}$$

The algebraic sign of sin θ depends upon whether the ray is ascending or descending and is handled in the same manner as in the constant-gradient method. Insertion of these values into (481) yields

$$dx = \pm \sqrt{\frac{c_a^2}{c_v^2 - c_a^2 - c_v^2 k_c (y - y_a)^2}}\, dy \qquad (483)$$

Equation (483) takes on a different form for each of the three types.

Type I:

Let
$$w = \frac{c_v\sqrt{k_a}}{\sqrt{c_v^2 - c_a^2}} (y - y_a) \qquad (484)$$

Then
$$dx = \pm \frac{c_a}{c_v\sqrt{k_a}} \frac{dw}{\sqrt{1 - w^2}}$$

The solution is
$$x = \pm \frac{c_a}{c_v\sqrt{k}} \sin^{-1} w + \text{constant} \qquad (485)$$

Type II:

Since k_a is negative, let
$$w = \frac{c_v\sqrt{-k_a}}{\sqrt{c_v^2 - c_a^2}} (y - y_a) \qquad (486)$$

Then
$$dx = \pm \frac{c_a}{c_v\sqrt{-k_a}} \frac{dw}{\sqrt{1 + w^2}}$$

and
$$x = \pm \frac{c_a}{c_v\sqrt{-k_a}} \log (w + \sqrt{w^2 + 1}) + \text{constant} \qquad (487)$$

Type III:

Here we define w as in (484). Since c_a^2 is negative, the differential equation is
$$dx = \pm \frac{\sqrt{-c_a^2}}{c_v\sqrt{k_a}} \frac{dw}{\sqrt{w^2 - 1}}$$

and the solution is

$$x = \pm \frac{\sqrt{-c_a^2}}{c_v\sqrt{k_a}} \log (w + \sqrt{w^2 - 1}) + \text{constant} \qquad (488)$$

The constants of integration in (485), (487), and (488) are evaluated in terms of the initial values of x and y in the particular segment to which they apply. The velocity profile is divided into a number of segments, each of which is fitted by a function of the form (480). As a ray is traced through one segment after another, the x, y coordinates at the end of one segment become the initial conditions for the next.

One price which must be paid for the curvilinear segment method is the problem of curve fitting. There are a number of possible methods of fitting curves of the form (480) to the velocity profile data, which are usually available in the form of a table of sound speed data at discrete depths. In order to avoid the difficulties of the constant gradient approach, it is necessary that there be no kinks in the curve. Two adjacent segments must meet with the same slope.

Where practicable it is desirable to fit a single segment to a considerable number of points by the method of least squares. When this is done, adjacent segments will not in general meet with the same slope. In this case a gap is placed between the two segments and is filled by a pair of "bridging curves." Each curve has three constants making a total of six, which may be adjusted to meet the following six requirements:

Point and slope at each end (4)

Bridging curves meet at same point with same slope (2)

We shall not discuss the curve fitting problem in further detail, except to point out that when a profile is built up by successively adding segments, two of the three constants for any one segment are required to fit the point and slope at the junction with the previous segment. It is therefore possible to fit either the point or the slope at the other end, but not both.

The curvilinear segment method is a significant improvement over the constant-gradient method in those situations where the constant-gradient approach leads to false caustics. It should be noted, however, that in a wide variety of situations there is little difference between the two methods, particularly with respect to ranges and travel times. Furthermore, when data are available only at discrete depths, there is no guarantee that the curvilinear segments represent the actual profile at intermediate points any more accurately than do the straight-line segments. Finally, the large amount of arithmetic required for curve fitting practically restricts the curvilinear method to digital computers.

14. Differential Equation Method

A few years ago an IBM 650 ray-tracing program was set up at the Aeronautical Computer Laboratory, based on the numerical solution of the differential equations. This approach has two advantages over the curvilinear-segment method.

(1) The selection of the horizontal range x as the independent variable avoids the ambiguity in the sign of θ, which introduces complications in other methods.

(2) The method is exceedingly flexible with respect to curve fitting. The only requirement is that the profile have no discontinuities (this requirement is of no importance, since discontinuities occur only at the boundaries between different media). Any reasonable type of curve may be used, and kinks such as occur between straight-line segments, can be handled by a minor modification.

The solutions for θ, y, s, and t are straightforward. The four differential equations are obtained by rearranging (450a) to (450d) and (451). Expressed in terms of c_0 and $\cos\theta_0$, they are

$$\frac{d(\sin\theta)}{dx} = \frac{g\cos\theta_o}{c_o} \tag{489a}$$

$$\frac{dy}{dx} = -\tan\theta \tag{489b}$$

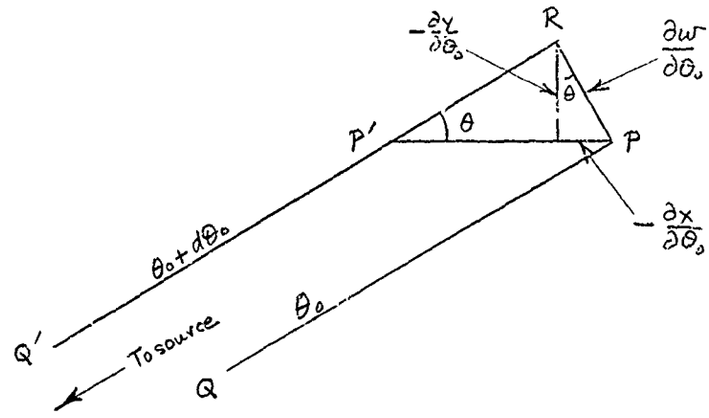
$$\frac{ds}{dx} = \frac{1}{\cos\theta} \tag{489c}$$

$$\frac{dt}{dx} = \frac{\cos\theta_o}{c_o\cos^2\theta} \tag{489d}$$

The intensity ratio is expressed in terms of the variable w (introduced in section 10).

$$\frac{I_1}{I} = \frac{x}{s_1^2\cos\theta_o}\left.\frac{\partial w}{\partial\theta_o}\right|_{t=\text{const}} \tag{490}$$

The partial derivative $\frac{\partial w}{\partial\theta_o}$ is taken along a wave front at right angles to the ray. It is thus a derivative with the travel time held constant. In the sketch



QP is the ray with initial angle θ_o, Q'P' the ray with initial angle $\theta_o + d\theta_o$, and PR is the constant - t wave front. The derivatives $\frac{\partial w}{\partial\theta_o}, \frac{\partial x}{\partial\theta_o}, \frac{\partial y}{\partial\theta_o}$ are all taken at constant t. From the figure,

$$\frac{\partial w}{\partial \theta_o} = -\frac{\partial x}{\partial \theta_o} \cdot \sin\theta - \frac{\partial y}{\partial \theta_o} \cdot \cos\theta \qquad (491a)$$

$$0 = -\frac{\partial x}{\partial \theta_o} \cdot \cos\theta + \frac{\partial y}{\partial \theta_o} \sin\theta \qquad (491b)$$

The minus signs in (491a) indicate that both x and y decrease with θ_o.
The time derivative of (491a) is

$$\frac{d}{dt}\left(\frac{\partial w}{\partial \theta_o}\right) = -\sin\theta \frac{d}{dt}\left(\frac{\partial x}{\partial \theta_o}\right) - \cos\theta \frac{d}{dt}\left(\frac{\partial y}{\partial \theta_o}\right) + \left(-\frac{\partial x}{\partial \theta_o}\cos\theta + \frac{\partial y}{\partial \theta_o}\sin\theta\right)\frac{d\theta}{dt}$$

By (491b) the last term is zero, leaving

$$\frac{d}{dt}\left(\frac{\partial w}{\partial \theta_o}\right) = -\sin\theta \frac{d}{dt}\left(\frac{\partial x}{\partial \theta_o}\right) - \cos\theta \frac{d}{dt}\left(\frac{\partial y}{\partial \theta_o}\right) \qquad (492)$$

To evaluate the derivatives on the right, we start with

$$\frac{dx}{dt} = \frac{c_o \cos^2\theta}{\cos\theta_o}$$

$$\frac{dy}{dt} = -\frac{c_o \sin\theta\cos\theta}{\cos\theta_o}$$

Taking derivatives with respect to θ_o at constant t, we obtain

$$\frac{d}{dt}\left(\frac{\partial x}{\partial \theta_o}\right) = \frac{c_o}{\cos\theta_o}\left[\frac{\sin\theta_o}{\cos\theta_o}\cos^2\theta - 2\sin\theta\cos\theta\frac{\partial\theta}{\partial\theta_o}\right]$$

$$\frac{d}{dt}\left(\frac{\partial y}{\partial \theta_o}\right) = \frac{c_o}{\cos\theta_o}\left[-\frac{\sin\theta_o}{\cos\theta_o}\sin\theta\cos\theta - (\cos^2\theta - \sin^2\theta)\frac{\partial\theta}{\partial\theta_o}\right]$$

Substitution of these equations into (492) gives

$$\frac{d}{dt}\left(\frac{\partial w}{\partial \theta_o}\right) = \frac{c_o\cos\theta}{\cos\theta_o}\frac{\partial\theta}{\partial\theta_o} \qquad (493)$$

We must now obtain a differential equation for $\frac{\partial\theta}{\partial\theta_o}$ which appears in (493).
This is done by taking the derivative of the equation

$$\frac{d\theta}{dt} = g\cos\theta$$

$$\frac{d}{dt}\left(\frac{\partial\theta}{\partial\theta_o}\right) = \frac{dg}{dy}\frac{\partial y}{\partial\theta_o}\cos\theta - g\sin\theta\frac{\partial\theta}{\partial\theta_o}$$

But

$$\frac{\partial y}{\partial \theta_o} = -\frac{\partial w}{\partial \theta_o} \cos \theta$$

Therefore

$$\frac{d}{dt}\left(\frac{\partial \theta}{\partial \theta_o}\right) = -\frac{dg}{dy} \frac{\partial w}{\partial \theta_o} \cos^2 \theta - g \sin \frac{\partial \theta}{\partial \theta_o}$$

Finally, we must convert to derivatives with respect to x. According to (489d) the conversion factor is

$$\frac{\cos \theta_o}{c_o \cos^2 \theta}$$

To obtain the derivative $\frac{\partial w}{\partial \theta_o}$ for the intensity ratio (490) we must integrate two differential equations

$$\frac{d}{dx}\left(\frac{\partial w}{\partial \theta_o}\right) = \frac{1}{\cos \theta} \frac{\partial \theta}{\partial \theta_o} \tag{494}$$

$$\frac{d}{dx}\left(\frac{\partial \theta}{\partial \theta_o}\right) = -\frac{\cos \theta_o}{c_o}\left[\frac{dg}{dy} \frac{\partial w}{\partial \theta_o} + \frac{g \sin \theta}{\cos^2 \theta} \frac{\partial \theta}{\partial \theta_o}\right] \tag{495}$$

It is seen that the derivative of the gradient, which appears in (495), becomes infinite at a boundary between two constant-gradient layers, or at any point where the slope of the velocity profile changes discontinuously. An impulse in $\frac{dg}{dy}$ produces a step in $\frac{\partial \theta}{\partial \theta_o}$. To evaluate the step, we assume a thin intermediate layer in which g changes continuously between the two values, integrate over x through this layer, and then let the thickness of the layer go to zero. In this process we may assume that all functions except $\frac{dg}{dy}$ are constant, and shall designate them with the subscript $_{12}$, referring to the boundary. The integral is

$$\int_{x_1}^{x_2} \frac{dg}{dy}\, dx = \int_{y_1}^{y_2} \frac{dg}{dy} \frac{dx}{dy}\, dy = -\cot \theta_{12} \int_{y_1}^{y_2} \frac{dg}{dy}\, dy$$

$$= -(g_2 - g_1) \cot \theta_{12}$$

The increment in $\frac{\partial \theta}{\partial \theta_o}$ at the boundary is

$$\Delta\left(\frac{\partial \theta}{\partial \theta_o}\right) = \frac{(g_2 - g_1) \cos \theta_o \cos \theta_{12}}{c_o \sin \theta_{12}} \left(\frac{\partial w}{\partial \theta_o}\right)_{12} \tag{496}$$

The complete solution consists of (489a), (489b), (489c), (489d), (490), (494), (495), and (496).

The chief disadvantage of this method is the long running time required on the IBM 650 computer. It should prove more practical on a faster computer.

The Propagation of Underwater Sound (cont'd)

C. Reflection

1. Introduction

When sound waves strike the surface or the bottom of the ocean, a combination of processes occurs, whereby some of the energy is reflected, some is scattered in all directions, and some is transmitted into the adjoining medium. The scattering depends on a number of factors such as the roughness of the bottom and the waves at the surface, and is beyond the scope of the present discussion. In this section we shall assume an ideal surface between two adjacent media and shall describe the reflection and transmission characteristics in terms of plane waves.

2. Reflected and Transmitted Pressure and Intensity

The solution of the wave equation for two-dimensional plane waves has been derived in an earlier section. The pressure and the two components of the particle velocity are given by equations (132), (133), and (134). The pressure equation is reproduced below

$$p = p_m e^{j\omega\left[t - \frac{1}{c}(x \cos\theta + y \sin\theta)\right]} \tag{132}$$

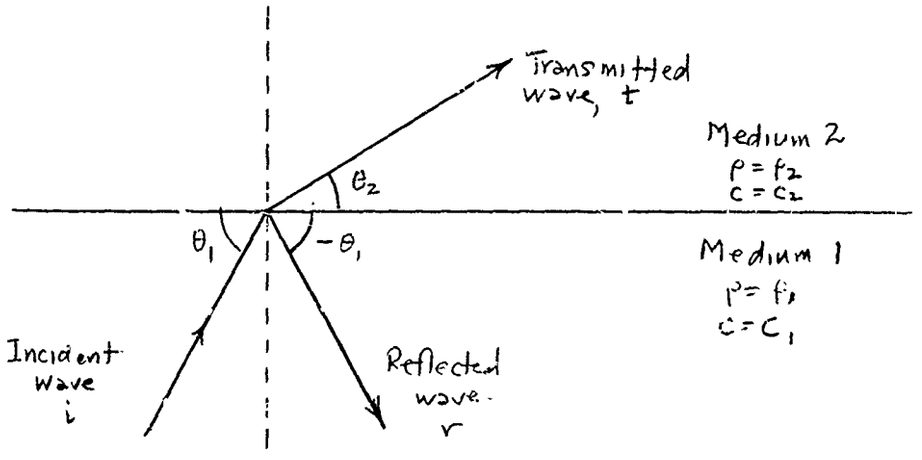
The x- and y-components, u and v, of the particle velocity vector $\vec{u}$ can be expressed in terms of the instantaneous pressure

$$u = \frac{p \cos\theta}{\rho c} \tag{133a}$$

$$v = \frac{p \sin\theta}{\rho c} \tag{134a}$$

In these equations the angle θ is measured from the x-axis toward the positive y-direction and is therefore opposite in sign to the angle employed above in the description of ray paths.

We shall assume that the incident ray in the first medium is traveling toward the boundary between the first and second media, the direction of propagation making an angle θ_1 with the boundary. The density and sound speed have constant values ρ_1 and c_1 in the first medium and different constant values ρ_2 and c_2 in the second. The abrupt change in parameters



at the boundary gives rise (in general) to both a reflected wave and a transmitted wave, whose directions of propagation make angles of $-\theta_1$ and θ_2 respectively with the boundary. We shall identify the three waves by the subscripts i, r, and t. The respective pressures and particle velocity components are

Incident:

$$p_i = p_{mi}\, e^{j\omega\left[t - \frac{1}{c_1}(x\cos\theta_1 + y\sin\theta_1)\right]} \tag{501}$$

$$u_i = \frac{p_i \cos\theta_1}{\rho_1 c_1} \tag{502}$$

$$v_i = \frac{p_i \sin\theta_1}{\rho_1 c_1} \tag{503}$$

Reflected:

$$p_r = p_{mr} e^{j\omega\left[t - \frac{1}{c_1}(x \cos \theta_1 - y \sin \theta_1)\right]} \tag{501a}$$

$$u_r = \frac{p_r \cos \theta_1}{\rho_1 c_1} \tag{502a}$$

$$v_r = -\frac{p_r \sin \theta_1}{\rho_1 c_1} \tag{503a}$$

Transmitted:

$$p_t = p_{mt} e^{j\omega\left[t - \frac{1}{c_2}(x \cos \theta_2 + y \sin \theta_2)\right]} \tag{501b}$$

$$u_t = \frac{p_t \cos \theta_2}{\rho_2 c_2} \tag{502b}$$

$$v_t = \frac{p_t \sin \theta_2}{\rho_2 c_2} \tag{503b}$$

Two conditions must be met at the boundary:

(1) The pressure is continuous, that is, the resultant pressure is the same on either side of the boundary.

$$p_i + p_r = p_t \tag{504}$$

This condition results from the fact that there is no physical mechanism at the boundary to support a difference in pressure.

(2) The normal component, v, of the particle velocity is continuous at the boundary, i.e.,

$$v_i + v_r = v_t \tag{505}$$

This condition expresses the fact that particles of both media cannot occupy the same space or leave void spaces at the boundary. Equations (504) and (505) must be identically true, that is, must be true for all values of t and for all values of x. Without loss of generality in the present analysis we may place the origin of y at the boundary, that is

$$y = 0 \text{ at the boundary}$$

Substitution of (501), (501a), and (501b) into (504) yields at the boundary

$$(p_{mi} + p_{mr})\, e^{j\omega\left(t - \frac{x \cos \theta_1}{c_1}\right)} \equiv p_{mt}\, e^{j\omega\left(t - \frac{x \cos \theta_2}{c_2}\right)} \tag{506}$$

If (506) is to be identically true in t and x, the exponents must be identical. Thus,

$$\omega\left(t - \frac{x \cos \theta_1}{c_1}\right) \equiv \omega\left(t - \frac{x \cos \theta_2}{c_2}\right)$$

It is seen that we have already guaranteed identity in t by selecting the same frequency for all three waves. The condition for identity in x is then

$$\frac{\cos \theta_1}{c_1} = \frac{\cos \theta_2}{c_2}$$

which, of course, is Snell's Law.

Since the exponents of all three waves are identical at the boundary, equations (504) and (505) hold for both instantaneous and rms values of the variables.

Equation (505) may be written in terms of the pressures by insertion of (503), (503a), and (503b)

$$\frac{p_i \sin \theta_1}{\rho_1 c_1} - \frac{p_r \sin \theta_1}{\rho_1 c_1} = \frac{p_t \sin \theta_2}{\rho_2 c_2} \tag{507}$$

We now have two simultaneous equations (504) and (507) which may be solved for the pressure ratios p_r/p_i and p_t/p_i. We shall call these ratios the reflection and transmission coefficients and shall designate them by the symbols R and T respectively. The results are

$$R = \frac{p_r}{p_i} = \frac{\frac{\sin \theta_1}{\rho_1 c_1} - \frac{\sin \theta_2}{\rho_2 c_2}}{\frac{\sin \theta_1}{\rho_1 c_1} + \frac{\sin \theta_2}{\rho_2 c_2}} \tag{508}$$

and

$$T = \frac{p_t}{p_i} = \frac{\frac{2 \sin \theta_1}{\rho_1 c_1}}{\frac{\sin \theta_1}{\rho_1 c_1} + \frac{\sin \theta_2}{\rho_2 c_2}} \tag{509}$$

The angie θ_2 is related to θ_1 by Snell's Law.

$$\sin \theta_2 = \sqrt{1 - \frac{c_2^2}{c_1^2} \cos^2 \theta_1} \tag{510}$$

The reflection and transmission coefficients may therefore be written

$$R = \frac{\sin \theta_1 - \frac{\rho_1}{\rho_2} \sqrt{\frac{c_i^2}{c_2^2} - \cos^2 \theta_1}}{\sin \theta + \frac{\rho_1}{\rho_2} \sqrt{\frac{c_1^2}{c_2^2} - \cos^2 \theta_1}} \tag{508a}$$

$$T = \frac{2 \sin \theta_1}{\sin \theta_1 + \frac{\rho_1}{\rho_2} \sqrt{\frac{c_1^2}{c_2^2} - \cos^2 \theta_1}} \tag{509a}$$

At normal incidence both θ_1 and θ_2 are 90 degrees and the coefficients become

$$R = \frac{\rho_2 c_2 - \rho_1 c_1}{\rho_2 c_2 + \rho_1 c_1} \tag{511}$$

$$T = \frac{2 \rho_2 c_2}{\rho_2 c_2 + \rho_1 c_1} \tag{512}$$

Intensity

Since the incident and reflected waves are both in the first medium and the transmitted wave is in the second medium, the respective intensities are

$$I_i = \frac{p_1^2}{\rho_1 c_1} \tag{513a}$$

$$I_r = \frac{p_r^2}{\rho_1 c_1} \tag{513b}$$

$$I_t = \frac{p_t^2}{\rho_2 c_2} \tag{513c}$$

The intensity ratios are therefore

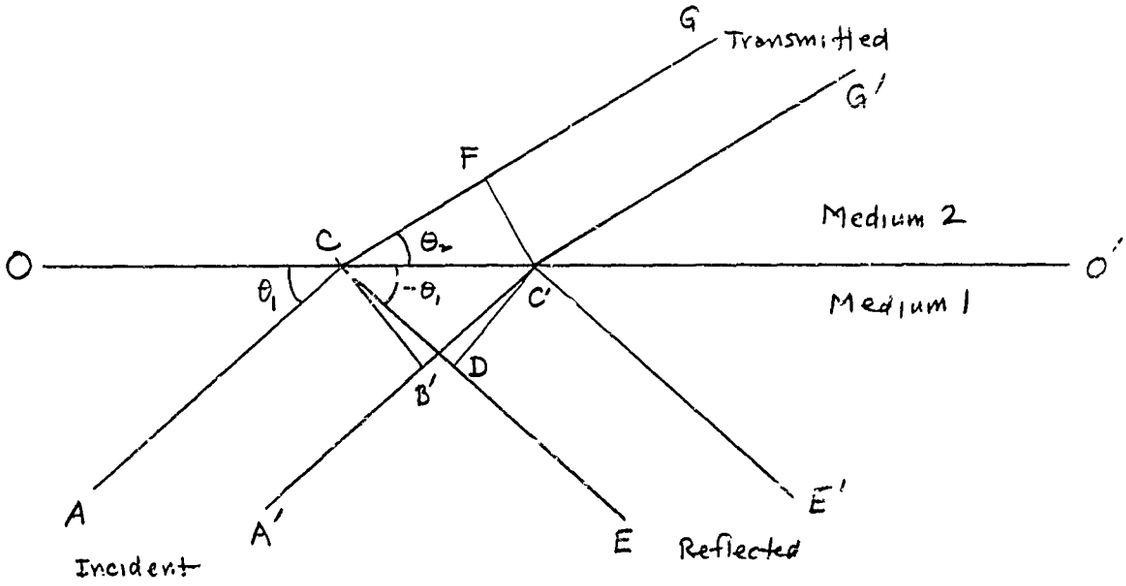
$$\frac{I_r}{I_i} = \left(\frac{p_r}{p_i}\right)^2 = \left[\frac{\dfrac{\sin\theta_1}{\rho_1 c_1} - \dfrac{\sin\theta_2}{\rho_2 \rho_2}}{\dfrac{\sin\theta_1}{\rho_1 c_1} + \dfrac{\sin\theta_2}{\rho_2 c_2}}\right]^2 \tag{514a}$$

$$\frac{I_t}{I_i} = \frac{\rho_1 c_1}{\rho_2 c_2}\left(\frac{p_t}{p_i}\right)^2 = \frac{\dfrac{4\sin^2\theta_1}{\rho_1 c_1 \rho_2 c_2}}{\left(\dfrac{\sin\theta_1}{\rho_1 c_1} + \dfrac{\sin\theta_2}{\rho_2 c_2}\right)^2} \tag{514b}$$

It is interesting to note that the sum of the two intensity ratios (514a) and (514b) is not equal to 1, that is,

$$I_r + I_t \neq I_i$$

as may be seen by carrying out the algebraic operations. At first thought this would appear to violate the principle of conservation of energy. A closer examination, however, will resolve the difficulty. The rate of energy propagation across a surface at right angles to the direction of propagation is



equal to the product of the intensity and the area of the surface. Consider the energy falling on a small rectangular area of the interface 00' in the sketch. Let the dimensions of this area be

$$\Delta x = CC'$$

in the 00' direction and one unit in the direction normal to the paper. Then the area normal to the incident wave is B'C, the area normal to the reflecte wave is C'D, and the area normal to the transmitted wave is C'F, where

$$B'C = \Delta x \sin \theta_1$$

$$C'D = \Delta x \sin \theta_1$$

$$C'F = \Delta x \sin \theta_2$$

By the principle of conservation of energy we should expect the power of the incident wave to be equal to the combined power of the reflected and transmitted waves,

$$I_i \Delta x \sin \theta_1 = I_r \Delta x \sin \theta_1 + I_t \Delta x \sin \theta_2$$

or

$$\frac{I_r}{I_i} + \frac{I_t \sin \theta_2}{I_i \sin \theta_1} = 1 \tag{515}$$

This result is easily verified by substituting (514a) and (514b).

At normal incidence the intensity ratios are

$$\frac{I_r}{I_i} = \frac{(\rho_2 c_2 - \rho_1 c_1)^2}{(\rho_2 c_2 + \rho_1 c_1)^2} \tag{516a}$$

and

$$\frac{I_t}{I_i} = \frac{4 \rho_1 c_1 \rho_2 c_2}{(\rho_2 c_2 + \rho_1 c_1)^2} \tag{516b}$$

It will be noted that in this case the intensity ratios add to give 1, since

$$\sin \theta_1 = \sin \theta_2 = 1$$

3. Effects of Boundary Conditions

Rigid Boundary

Consider first the extreme case of reflection of the sound waves from an infinitely dense and rigid solid, where

$$\rho_2 c_2 = \infty$$

Here we find

$$\left.\begin{aligned} R &= 1 \ ; & \frac{I_r}{I_i} &= 1 \\ T &= 2 \ ; & \frac{I_t}{I_i} &= 0 \end{aligned}\right\} \qquad (517)$$

The wave is perfectly reflected with no phase change. Since the incident and reflected pressures add in phase, the pressure in the solid wall is doubled. This ideal situation is approximated in nature when sound waves in air are reflected from a water surface. Here we have

$$\rho_1 c_1 = 42 \text{ sp. ac. ohms (air)}$$

$$\rho_2 c_2 = 154{,}000 \text{ sp. ac. ohms (water)}$$

At normal incidence the ratios are

$$R = \frac{1 - \dfrac{\rho_1 c_1}{\rho_2 c_2}}{1 + \dfrac{\rho_1 c_1}{\rho_2 c_2}} = \frac{1 - .00027}{1 + .00027} = 0.99945$$

$$T = \frac{2}{1 + \dfrac{\rho_1 c_1}{\rho_2 c_2}} = 1.99945$$

$$\frac{I_r}{I_i} = \left(\frac{p_r}{p_i}\right)^2 = 0.99891$$

$$\frac{I_t}{I_i} = \frac{\rho_1 c_1}{\rho_2 \rho_2} \left(\frac{p_t}{p_i}\right)^2 = 0.001090$$

Thus, the reflected and incident intensities are almost equal and the transmitted intensity is about 30 db down.

Free Boundary

Consider next the opposite extreme of a vacuum instead of a solid wall. If

$$\rho_2 c_2 = 0$$

we have

$$\left.\begin{aligned} R &= -1 \ ; & \frac{I_r}{I_i} &= 1 \\ T &= 0 \ ; & \frac{I_t}{I_i} &= 0 \end{aligned}\right\} \qquad (518)$$

In this case the adjoining medium cannot support any pressure at all, so that the incident and reflected pressures must add to zero; hence the ratio is -1. The reflection therefore involves a phase reversal, that is, a phase shift of 180 degrees. This case is approximated in nature by the reflection of waves in water from an air surface. The following table lists the various ratios as a function of the angle of incidence θ_1 in water. The angle of refraction in air is computed from Snell's Law.

θ_1 deg.	θ_2 deg.	$\frac{p_r}{p_i}$	$\frac{p_t}{p_i}$	$\frac{I_r}{I_i}$	$\frac{I_t}{I_i}$
0	77.3	-1.00000	0	1.00000	0
15	77.7	-0.99986	0.00014	0.99971	0.00008
30	79.0	-0.99972	0.00028	0.99944	0.00028
45	81.1	-0.99961	0.00039	0.99922	0.00056
60	83.7	-0.99952	0.00048	0.99905	0.00083
75	86.7	-0.99947	0.00053	0.99895	0.00105
90	90.0	-0.99945	0.00055	0.99891	0.00109

It will be noted that the intensity loss at normal incidence is the same whether the sound waves are traveling from air into water or vice versa, as may be seen from the symmetric form of (516b).

Total Reflection

When the speed of sound is larger in the second medium, i.e., when $c_2 > c_1$, the critical angle of incidence is given by

$$\cos \theta_c = \frac{c_1}{c_2} \quad \text{or} \quad \sin \theta_c = \sqrt{1 - \left(\frac{c_1}{c_2}\right)^2} \tag{519}$$

Whenever the angle of incidence is less than the critical angle, i.e., when $\theta_1 < \theta_c$, the angle of refraction ceases to exist. In this case

$$\cos \theta_2 = \frac{c_2}{c_1} \cos \theta_1 > 1$$

When we try to compute the sine of θ_2, we obtain an imaginary number.

$$\sin \theta_2 = \pm j\sqrt{\left(\frac{c_2}{c_1}\right)^2 \cos^2 \theta_1 - 1} \tag{520}$$

To investigate the significance of the imaginary sine, let us substitute this value into the formula (501b) for the transmitted pressure. The result is

$$p_t = p_{mt}\, e^{j\omega\left[t - \frac{1}{c_2}\left(x \cos \theta_2 \pm jy\sqrt{\left(\frac{c_2}{c_1}\right)^2 \cos^2 \theta - 1}\right)\right]}$$

It is seen that the coefficient of y in the exponent is now <u>real</u>. Thus, in the second medium the pressure varies exponentially with the distance from the boundary. It is clear physically that the pressure must decay rather than build up, and hence (assuming y is positive in the second medium) the lower sign in (520) must apply. The transmitted pressure is then

$$p_t = p_{mt}\, e^{-\frac{2\pi y}{\lambda_2}\sqrt{\left(\frac{c_2}{c_1}\right)^2 \cos^2 \theta_1 - 1}}\, e^{j\omega\left(t - \frac{x \cos \theta_2}{c_2}\right)} \tag{521}$$

where, of course, $\frac{2\pi}{\lambda_2} = \frac{\omega}{c_2}$

Upon substitution of (520) into (508a) the reflection coefficient becomes

$$R = \frac{\sin\theta_1 + j\frac{\rho_1}{\rho_2}\sqrt{\cos^2\theta_1 - (\frac{c_1}{c_2})^2}}{\sin\theta_1 - j\frac{\rho_1}{\rho_2}\sqrt{\cos^2\theta_1 - (\frac{c_1}{c_2})^2}} \tag{522}$$

To simplify the writing, let

$$a = \sin\theta_1 \tag{523}$$

$$b = \frac{\rho_1}{\rho_2}\sqrt{\cos^2\theta_1 - (\frac{c_1}{c_2})^2} \tag{523a}$$

so that

$$R = \frac{a + jb}{a - jb} = \frac{(a + jb)^2}{a^2 + b^2} \tag{522a}$$

The reflection coefficient is thus a complex number whose absolute value is

$$|R| = \left|\frac{a + jb}{a - jb}\right| = 1$$

At angles less than the critical angle, the magnitude of the reflected pressure is equal to the magnitude of the incident pressure, and the power is therefore totally reflected. There is, however, a phase change, whose value can be readily determined if the reflection coefficient is expressed in the exponential form

$$R = \left(\frac{a + jb}{\sqrt{a^2 + b^2}}\right)^2 = e^{j\epsilon} \tag{522b}$$

from which we see that

$$\tan\frac{\epsilon}{2} = \frac{b}{a} = \frac{\rho_1}{\rho_2}\frac{\sqrt{\cos^2\theta_1 - (\frac{c_1}{c_2})^2}}{\sin\theta_1} \tag{524}$$

At the critical angle the numerator of (524) is zero and hence the phase change is likewise zero. As θ_1 decreases from θ_c to 0, the numerator of (524) increases and the denominator decreases. Thus ϵ increases from 0 at the critical angle to 180 degrees at grazing incidence.

It is of interest to note that although the incident intensity is totally reflected from the boundary, nevertheless a certain amount of acoustic energy is present beyond the boundary. From (521) we see that the transmitted pressure is a wave which travels parallel to the boundary and whose amplitude decreases exponentially with the penetration distance in the second medium.

Angle of Intromission

In the preceding discussion it has been tacitly assumed that if one medium has a greater density than another, it will also have a higher sound speed. This is not necessarily true. Oceanographic measurements have demonstrated that there are areas of the ocean where the bottom sediments (which of course are denser than water) have a lower sound speed. Such areas exhibit very peculiar reflection characteristics, as we shall now show.

Assume the following:

$$\rho_2 > \rho_1$$

$$c_2 < c_1$$

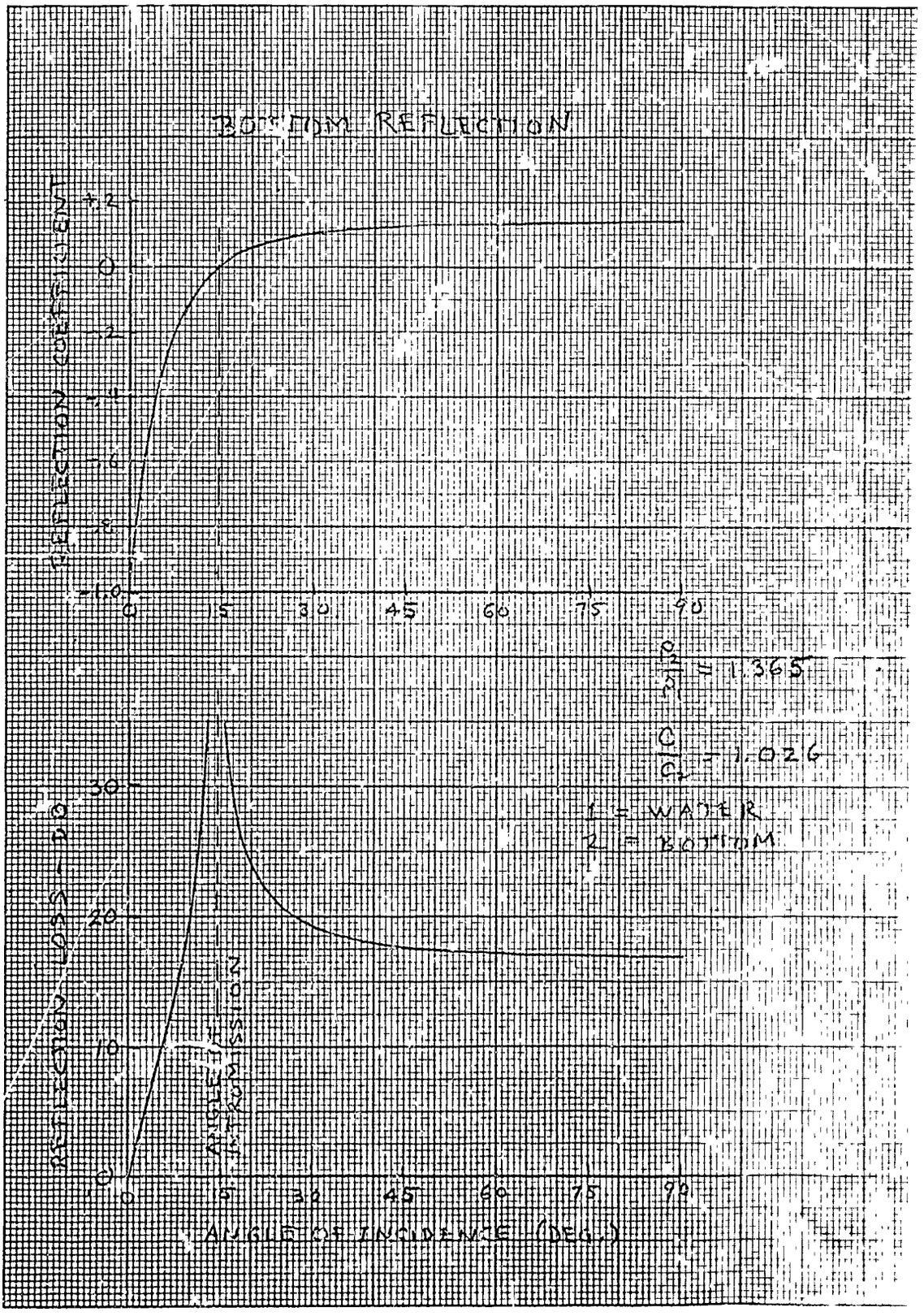
$$\rho_2 c_2 > \rho_1 c_1$$

The reflection coefficient (508a) is

$$R = \frac{\sin\theta_1 - \frac{\rho_1}{\rho_2}\sqrt{\frac{c_1^2}{c_2^2} - 1 + \sin^2\theta_1}}{\sin\theta_1 + \frac{\rho_1}{\rho_2}\sqrt{\frac{c_1^2}{c_2^2} - 1 + \sin^2\theta_1}} \qquad (508b)$$

Since the sound speed in the second medium is less than in the first, there is no critical angle and θ_1 may range all the way from 0 to 90 degrees. It is clear that at grazing incidence, where $\theta_1 = 0$, the coefficient is

$$R = -\frac{\frac{\rho_1}{\rho_2}\sqrt{\frac{c_1^2}{c_2^2} - 1}}{\frac{\rho_1}{\rho_2}\sqrt{\frac{c_1^2}{c_2^2} - 1}} = -1$$


BOTTOM REFLECTION
REFLECTED INTENSITY (DECIBELS)
+2
0
-2
-4
-6
-8
-10
0
15
30
45
60
75
90
n1 = 1.365
n2/n1 = 1.026
1 = WATER
2 = BOTTOM
REFLECTED LOSS (dB)
30
20
10
0
CRITICAL ANGLE
ANGLE OF INCIDENCE (DEG.)

whereas at normal incidence the coefficient is

$$R = \frac{1 - \frac{\rho_1 c_1}{\rho_2 c_2}}{1 + \frac{\rho_1 c_1}{\rho_2 c_2}} > 0$$

At small angles of incidence the coefficient is negative and at large angles it is positive. There exists, therefore, some intermediate angle at which the coefficient is zero. Setting the numerator of (508b) equal to 0 and solving for θ_1, we obtain

$$\sin \theta_1 = \sqrt{\frac{\frac{c_1^2}{c_2^2} - 1}{\frac{\rho_2^2}{\rho_1^2} - 1}} \qquad (525)$$

This angle is called the angle of intromission. A sound wave incident on the boundary at the angle of intromission will produce no reflected intensity; it is totally transmitted. The accompanying graph contains a plot of the reflection coefficient and the reflection loss in db calculated from data obtained by Fry and Raitt* at a location in the Pacific Ocean. The density and sound speed ratios are

$$\frac{\rho_2}{\rho_1} = 1.365; \quad \frac{c_1}{c_2} = 1.026$$

It is seen that at a zero angle of incidence in the water (ray horizontal) the reflection coefficient is -1, indicating perfect reflection with a 180 degree phase change. As the angle of incidence increases, the

— — — — — — — — —

*J. C. Fry and R. W. Raitt, "Sound Velocities at the Surface of Deep Sea Sediments," Journal of Geophysical Research, Vol 66, No. 2, p. 589, 1961.

ıde of the reflection coefficient drops rapidly to zero, the angle of ıssion being approximately 15 degrees. As the angle of incidence ses further, the reflection coefficient becomes slightly positive but exceeds a value of about 0.14. The corresponding value of the ion loss starts at zero at grazing incidence and increases rapidly to y at the angle of intromission. Above this angle it decreases, reaching nptotic value of 17 db at normal incidence. A bottom of this type is poor reflector, the loss being 17 db or higher at all angles of nce except in the immediate vicinity of the horizontal.

Note on Pressure Levels

We have seen that so long as we restrict our attention to ı in the water of the ocean, we may ignore the variations in ρc and the intensity as being proportional to the square of the pressure. Under conditions it makes sense to refer intensity levels to pressure units. ver, when we are concerned with the passage of waves between media ıg widely different values of ρc, we must be very cautious. The level esponding to a given intensity has very different numerical values when essed in db//μb^2 in the two media. For the same intensity the pressures ıd p_2 in two media having specific acoustic impedances ρ_1 c_1 and $_2$ are related by the basic equation

$$\frac{p_2^2}{\rho_2 c_2} = \frac{p_1^2}{\rho_1 c_1}$$

$_1$ is the pressure level in the first medium, expressed in db//μb^2, and ıs the corresponding level in the second medium, also expressed in μb^2, then

$$L_2 = L_1 + 10 \log \frac{\rho_2 c_2}{\rho_1 c_1} \qquad (526)$$

he first medium is air and the second is sea water, then

$$L_2 = L_1 + 10 \log \frac{154,000}{42}$$

$$= L_1 + 35.6 \text{ db}$$

4. Image Sources

In the discussion of refraction we learned that when waves from a source in one medium are refracted into a second medium, they appear to be diverging from an image source at a different location. In particular, if the source is located at a perpendicular distance H_1 from the boundary, then for normal incidence the waves in the second medium will appear to diverge from an image located in the first medium at a distance H_2 from the boundary, where

$$H_2 = \frac{c_1}{c_2} H_1 \tag{527}$$

and the source and its image lie on the same line normal to the boundary.

Our problem now is to determine the relative source levels of the source and its image. Let P_1 denote the power of the source. Assuming inverse square spreading, the intensity of the incident wave at the point where it first reaches the boundary is

$$I_1 = \frac{P}{4\pi H_1^2} \tag{528}$$

The ratio of the transmitted to the incident intensity (516b) is

$$\frac{I_2}{I_1} = \frac{4 \rho_1 c_1 \rho_2 c_2}{(\rho_2 c_2 + \rho_1 c_1)^2} \tag{529}$$

In order to generate an intensity I_2 at a distance H_2, the power P_2 of the image source must be

$$P_2 = 4\pi H_2^2 I_2 \tag{530}$$

Combining (527), (528), (529) and (530) we obtain

$$P_2 = \frac{4 \rho_1 c_1 \rho_2 c_2}{(\rho_2 c_2 + \rho_1 c_1)^2} \left(\frac{c_1}{c_2}\right)^2 P_1 \tag{531}$$

If the source level expressed in db//watt/cm^2 is S_0, then the equivalent source level of the image is

$$S_I = S_o + 10 \log \left[\frac{4 \rho_1 c_1 \rho_2 c_2}{(\rho_2 c_2 + \rho_1 c_1)^2} \left(\frac{c_1}{c_2} \right)^2 \right] \qquad (532)$$

where S_I is also in db//watt/cm^2. If both levels are expressed in db//μb^2, an appropriate correction must be made according to (526). All distances in the second medium must of course be measured from the location of the image.

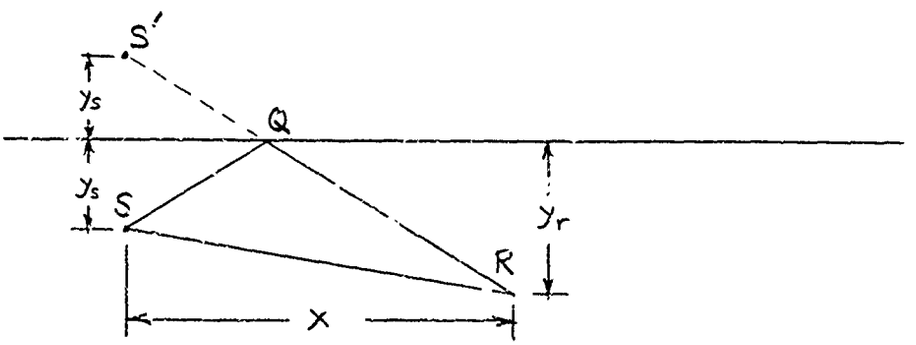
As an example, consider a source in air above the ocean, where $\rho_1 c_1 = 42$ sp. ac. ohms, $\rho_2 c_2 = 154,000$ sp. ac. ohms, and $c_1/c_2 = 0.22$. In terms of db//watt/cm^2 the difference in source levels is

$$S_I - S_o = 10 \log \left[\frac{4\left(\frac{42}{154,000}\right)}{\left(1 + \frac{42}{154,000}\right)^2} (0.22)^2 \right] = -42.8 \text{ db}$$

Two factors contribute to the reduction in equivalent source level: (1) the transmission loss at the boundary (29.6 db), which was discussed earlier, and (2) the increased divergence of the waves (13.2 db) due to the refraction.

5. The Lloyd Mirror Effect

When a source and receiver are located at shallow depths in the ocean, sound transmission takes place over two paths — the direct path and the surface-reflected path. We shall simplify the present discussion by assuming that the waves travel in straight lines and that the surface is perfectly smooth. The geometry is shown in the sketch below. The source S



is located at a depth y_s and the receiver R at a depth y_r, the horizontal separation between S and R being x. The surface-reflected wave, after reflection at Q, appears to have originated at the image source S', located at a distance y_s above the surface. Since the difference in length between the two paths SR and SQR varies with the geometry (depths and horizontal separation), it is to be expected that interference phenomena will be observed.

Let r_1 represent the length of the direct path SR and r_2 the length of the reflected path SQR (which is the same as that of S'QR). Then

$$r_1 = \sqrt{x^2 + (y_r - y_s)^2} \tag{533}$$

$$r_2 = \sqrt{x^2 + (y_r + y_s)^2} \tag{533a}$$

Recalling that a phase shift of 180 degrees (π radians) is encountered when the wave is reflected from the surface, we see that the difference in phase between the two waves arriving at R is

$$\Delta\phi = \frac{2\pi}{\lambda}(r_2 - r_1) + \pi \tag{534}$$

where λ is the wavelength. The two waves will reinforce each other and hence the resultant pressure will be a maximum when the phase difference is an integral number of cycles, that is, when

$$\text{Max:} \quad \Delta\phi = 2n\pi \qquad n = 1, 2, 3, \ldots \tag{535}$$

Destructive interference will occur and hence the resultant pressure will be a minimum when the phase difference is an odd number of half cycles, that is, when

$$\text{Min:} \quad \Delta\phi = (2n + 1)\pi \qquad n = 1, 2, 3, \ldots \tag{535a}$$

The two expressions may be combined into a single expression

$$\Delta\phi = (2n' + 1)\pi \tag{535b}$$

if n' is defined as follows

$$n' = 1, 2, 3, \ldots \quad \text{for minima}$$

and

$$n' = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots \quad \text{for maxima}$$

The locus of receiver positions at which n' has a constant value is a curve along which

$$r_2 - r_1 = \frac{\lambda}{2\pi}(\Delta\phi - \pi) = \text{constant}$$

or

$$r_2 - r_i = n'\lambda = \text{constant} \tag{536}$$

Such a curve is a member of a family of hyperbolas having the source and image as foci. The equation of the hyperbolas is obtained by inserting (533) and (533a) into (536) and squaring twice to eliminate the radicals.

$$\frac{y_r^2}{(\frac{n'\lambda}{2})^2} - \frac{x^2}{y_s^2 - (\frac{n'\lambda}{2})^2} = 1 \tag{537}$$

The horizontal range corresponding to a fixed receiver depth is found by solving (537) for x

$$x = \frac{2}{n'\lambda}\sqrt{\left[y_s^2 - (\frac{n'\lambda}{2})^2\right]\left[y_r^2 - (\frac{n'\lambda}{2})^2\right]} \tag{538}$$

When the source and receiver depths are large compared with the wavelength, the above expression simplifies approximately to

$$x \approx \frac{2\, y_s\, y_r}{n'\lambda} \tag{538a}$$

If the amplitudes of the two pressures were exactly equal at the receiver, the interference would be complete and the resultant pressure would be doubled at the maxima and zero at the minima. Actually, however, the amplitude of the surface-reflected wave is in general less than that of the direct wave, due in part to the reflection loss and in part to the longer path. Refraction effects due to velocity gradients also have a significant effect. As a result, the interference is not complete. As an illustration, let us assume that the amplitude ratio of the surface-reflected wave to the direct wave is K. Let p_1 and p_2 denote the instantaneous values of the direct and surface-reflected waves, and p denote the rms pressure of the direct wave. The real values of the pressures are

$$p_1 = \sqrt{2}\phi \cos\left(\omega t - \frac{2\pi r_1}{\lambda}\right)$$

$$p_2 = -\sqrt{2}Kp \cos\left(\omega t - \frac{2\pi r_2}{\lambda}\right)$$

where the minus sign represents the phase shift due to the reflection. The resultant instantaneous pressure is

$$p_1 + p_2 = \sqrt{2}\, p \cos\left[\left(\omega t - \frac{2\pi r_1}{\lambda}\right) - K \cos\left(\omega t - \frac{2\pi r_2}{\lambda}\right)\right]$$

and its square is

$$(p_1 + p_2)^2 = 2p^2\left[\cos^2\left(\omega t - \frac{2\pi r_1}{\lambda}\right) - 2K \cos\left(\omega t - \frac{2\pi r_1}{\lambda}\right) \cos\left(\omega t - \frac{2\pi r_2}{\lambda}\right) + K^2 \cos^2\left(\omega t - \frac{2\pi r_2}{\lambda}\right)\right]$$

The cross term may be expressed in terms of the sum and difference of the angles as follows

$$\cos\left(\omega t - \frac{2\pi r_1}{\lambda}\right) \cos\left(\omega t - \frac{2\pi r_2}{\lambda}\right) = \frac{1}{2}\left\{\cos\left[2\omega t - \frac{2\pi(r_2 + r_1)}{\lambda}\right] + \cos\frac{2\pi(r_2 - r_1)}{\lambda}\right\}$$

Therefore

$$(p_1 + p_2)^2 = 2p^2\left\{\cos^2\left(\omega t - \frac{2\pi r_1}{\lambda}\right) - K \cos\left[2\omega t - \frac{2\pi r_2 + r_1}{\lambda}\right] - K \cos\frac{2\pi(r_2 - r_1)}{\lambda} + K^2 \cos^2\left(\omega t - \frac{2\pi r_2}{\lambda}\right)\right\}$$

In taking the time average we see that the $\cos^2$ terms average to $\frac{1}{2}$ the second term averages to zero, while the third term is independent of time. The mean square resultant pressure is therefore

$$\overline{(p_1 + p_2)^2} = p^2\left[1 - 2K \cos\frac{2\pi(r_2 - r_1)}{\lambda} + K^2\right] \qquad (539)$$

But, from (536) and (538a)

$$\frac{2\pi(r_2 - r_1)}{\lambda} = 2\pi n' \approx \frac{4\pi\, y_s y_r}{\lambda x} \qquad (540)$$

so that

$$\overline{(p_1 + p_2)^2} = p^2 \left(1 - 2K \cos \frac{4\pi\, y_s\, y_r}{\lambda x} + K^2\right) \qquad (539a)$$

The interference pattern results from the last two terms of (539a). If there were no surface reflection, the mean square pressure would be simply p^2. The intensity anomaly is the ratio of the actual intensity to the intensity of the direct wave alone, and the transmission anomaly is 10 log of this ratio.

$$\text{Anomaly} = 10 \log \left(1 - 2K \cos \frac{4\pi\, y_s\, y_r}{\lambda x} + K^2\right) \qquad (541)$$

As an example let

$$K = 0.5$$

$$y_s = y_r = 25 \text{ ft.}$$

$$\lambda = 1.5 \text{ ft. } (f = 3333 \text{ cps})$$

The anomaly is

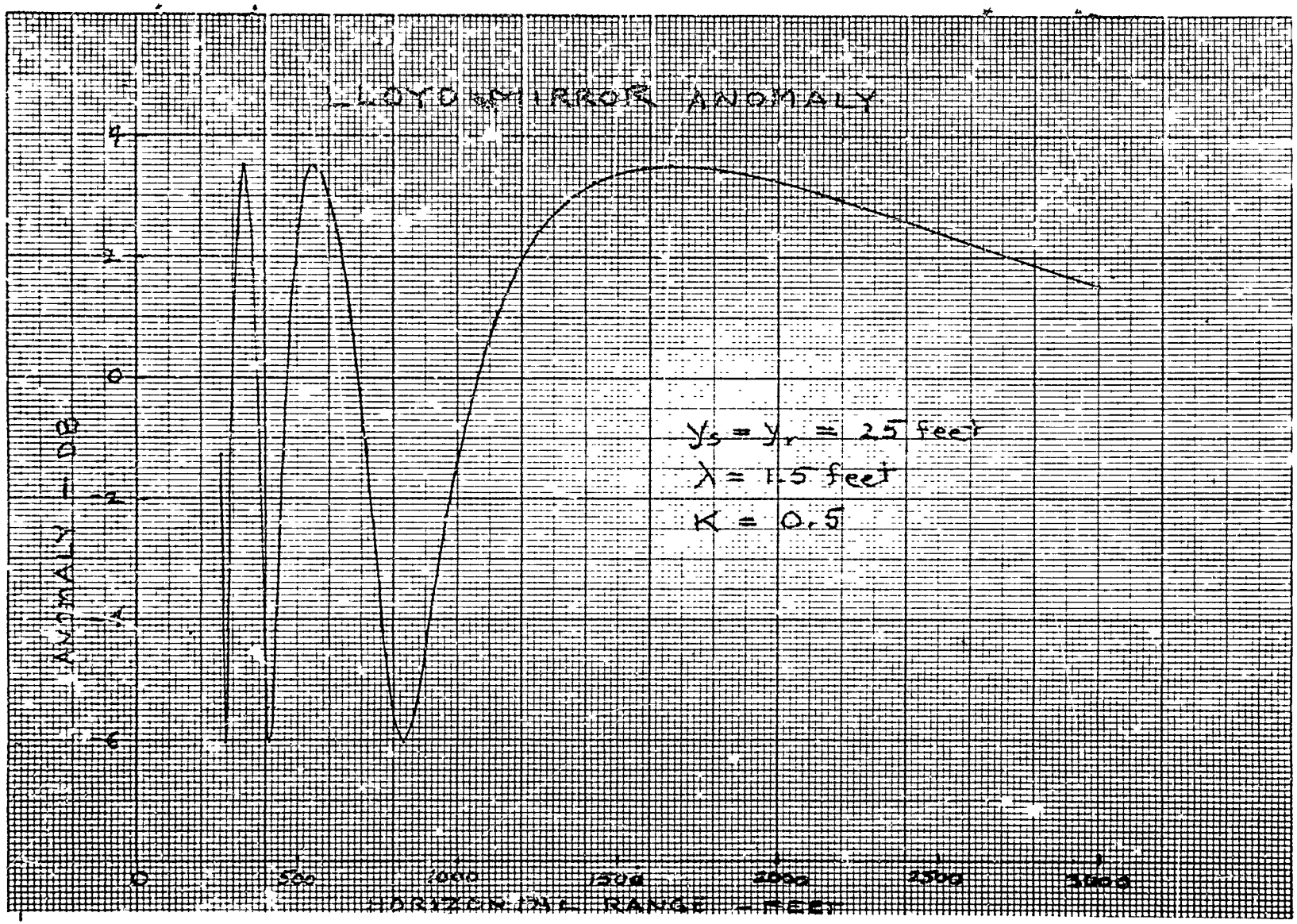
$$\text{Anomaly} = 10 \log \left(1.25 - \cos \frac{5000\pi}{3x}\right)$$

The accompanying graph shows the anomaly plotted as a function of the horizontal range. The first minimum theoretically occurs at infinite range (though at long ranges the assumption of straight-line propagation has little validity). The first maximum occurs at 1667 feet. As the range decreases further, the curve oscillates more and more rapidly.

It is seen that the anomaly varies from -6 db to + 3.5 db. This example illustrates the fact that surface reflections can cause significant errors in acoustic measurements.

6. Normal Mode Propagation

There are situations — notably in the propagation of low frequency sound in shallow water — where interference effects resulting from repeated surface and bottom reflections are a dominant factor in determining the nature of the propagation. In such cases, if the bottom is reasonably flat and level, the ocean acts as a wave guide. If the wave is traveling in certain specific directions relative to the horizontal, it will be reenforced by constructive interference. In all other directions cancellation will occur.


LLOYD'S MIRROR ANOMALY
4
2
0
-2
-4
-6
ANOMALY dB
$y_s = y_r = 2.5$ feet
$\lambda = 1.5$ feet
$K = 0.5$
0
500
1000
1500
2000
2500
3000
HORIZONTAL RANGE - FEET

This scope of this discussion will be limited to the ideal case of straight-line propagation in a homogeneous ocean. The bottom is assumed to be perfectly flat. The density and sound speed of the water are ρ_1 and c_1, and the corresponding quantities for the bottom are ρ_2 and c_2. It is assumed that

$$\rho_2 > \rho_1 \quad \text{and} \quad c_2 > c_1$$

The critical angle in water is θ_c where

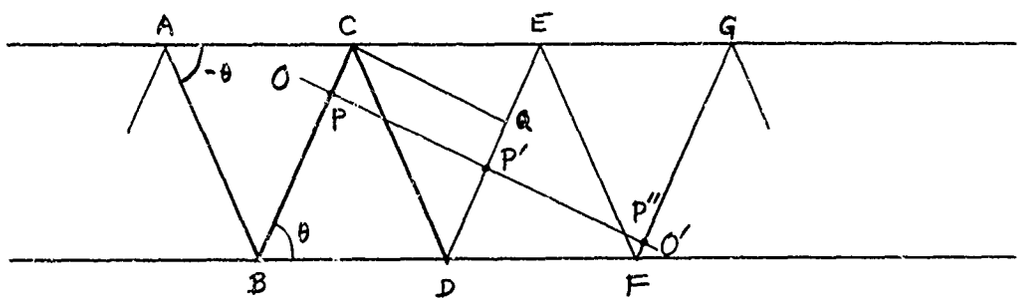
$$\cos \theta_c = \frac{c_1}{c_2}$$

It will be recalled that when θ is greater than θ_c, the bottom reflection coefficient is given by (508a), and no phase shift occurs. On the other hand, when θ is less than θ_c, the magnitude of the reflection coefficient is unity and the phase is advanced by an angle ϵ, where ϵ is given by (524). A certain amount of simplification may be achieved in the following discussion if we define ϵ as follows

$$\epsilon = 2 \tan^{-1} \frac{\rho_1}{\rho_2} \frac{\sqrt{\cos^2 \theta - \left(\frac{c_1}{c_2}\right)^2}}{\sin \theta}, \quad \theta < \theta_c$$

$$\epsilon = 0 \qquad \theta \geq \theta_c \tag{542}$$

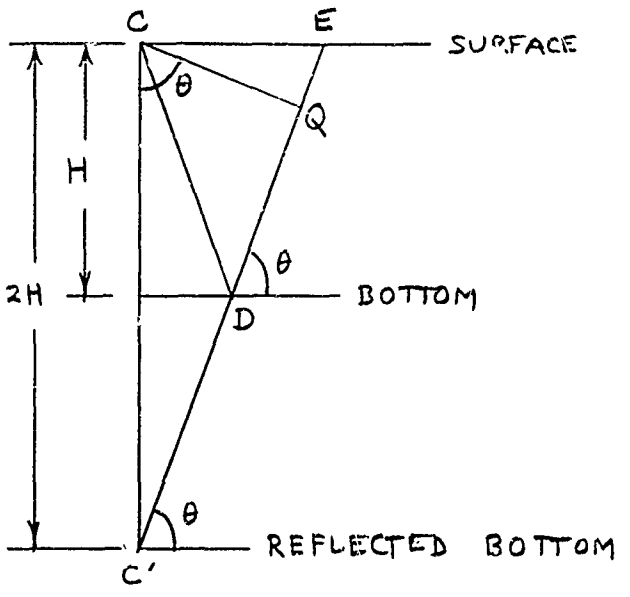
The basic concept of normal mode propagation is illustrated in the following sketch. The waves, traveling at angles $\pm \theta$ relative to the horizontal, undergo repeated reflections from the surface and bottom. One of the rays comprising the wave is shown as ABCDEFG. The space between A and C, B and D, etc., is filled with other rays all traveling at the same



angle. Consider the up-going ray BC. Lines perpendicular to this ray are wave fronts, one of which is shown by the line OO'. Since the wave must have the same phase at all points on a wave front, it is seen that the wave BC must have the same phase at P as DE has at P', FG at P'', etc. The requirement for reinforcement is therefore the following:

> When the wave travels the distance PCDP', the phase must change by an integral number of cycles.

In traveling along this path the wave suffers one reflection from the surface and one reflection from the bottom. The phase change on reflection from the bottom is ϵ. The direction of this change is such as to cause the phase to advance; it is therefore equivalent to a reduction in path length. The phase change at the surface is π radians. Since this is one half cycle, the algebraic sign is immaterial. We shall regard it as a retardation. To compute the path length between P and P' we note that the same result would be obtained if we selected any other wave front between B and C. The geometry is simpler if we select CQ. The length of the path CDQ, which is the same as PCDP', may be found quite simply by reflecting the ray CD in the bottom, as indicated in the following sketch. The reflected segment C'D is equal to CD, so that the



desired path length is C'Q. If the depth of the ocean is H, the length CC' of the hypoteneuse of the right triangle CC'Q is 2H. Hence

$$CD + DQ = C'Q = 2H \sin\theta$$

The phase change from C to Q is then

$$\Delta\phi = \frac{2\pi f_n}{c_1} \; 2H \sin\theta - \epsilon + \pi \tag{543}$$

and the requirement for constructive interference is

$$\Delta\phi = 2n\pi \tag{544}$$

where

$$n = 1, 2, 3, \ldots$$

Combining (543) and (544), we obtain

$$f_n = \frac{c_1}{4H \sin\theta} \left(2n - 1 + \frac{\epsilon}{\pi}\right) \tag{545}$$

where ϵ is defined by (542).

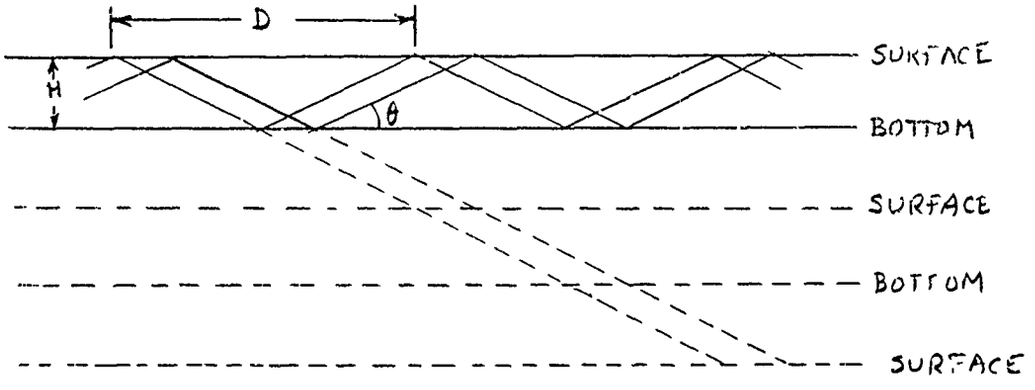
It is seen from (545) that corresponding to each angle θ there is a sequence of frequencies f_n which lead to a wave-guide type of propagation. The number n denotes the number of cycles by which the phase of the wave changes between successive points such as P and P' on the wave front. This type of propagation is called normal-mode propagation. The first mode corresponds to n = 1, the second to n = 2, etc.

On the accompanying graph is a plot of f_n vs. θ for a typical case where the ocean depth is 500 feet and the density and sound speed ratios are

$$\frac{\rho_1}{\rho_2} = 0.75; \quad \frac{c_1}{c_2} = 0.975$$

It will be noted that the curves show an abrupt change in slope at the critical angle, which is about 14 degrees. The change is due to the phase angle ϵ which suddenly begins to increase as θ decreases below the critical angle.

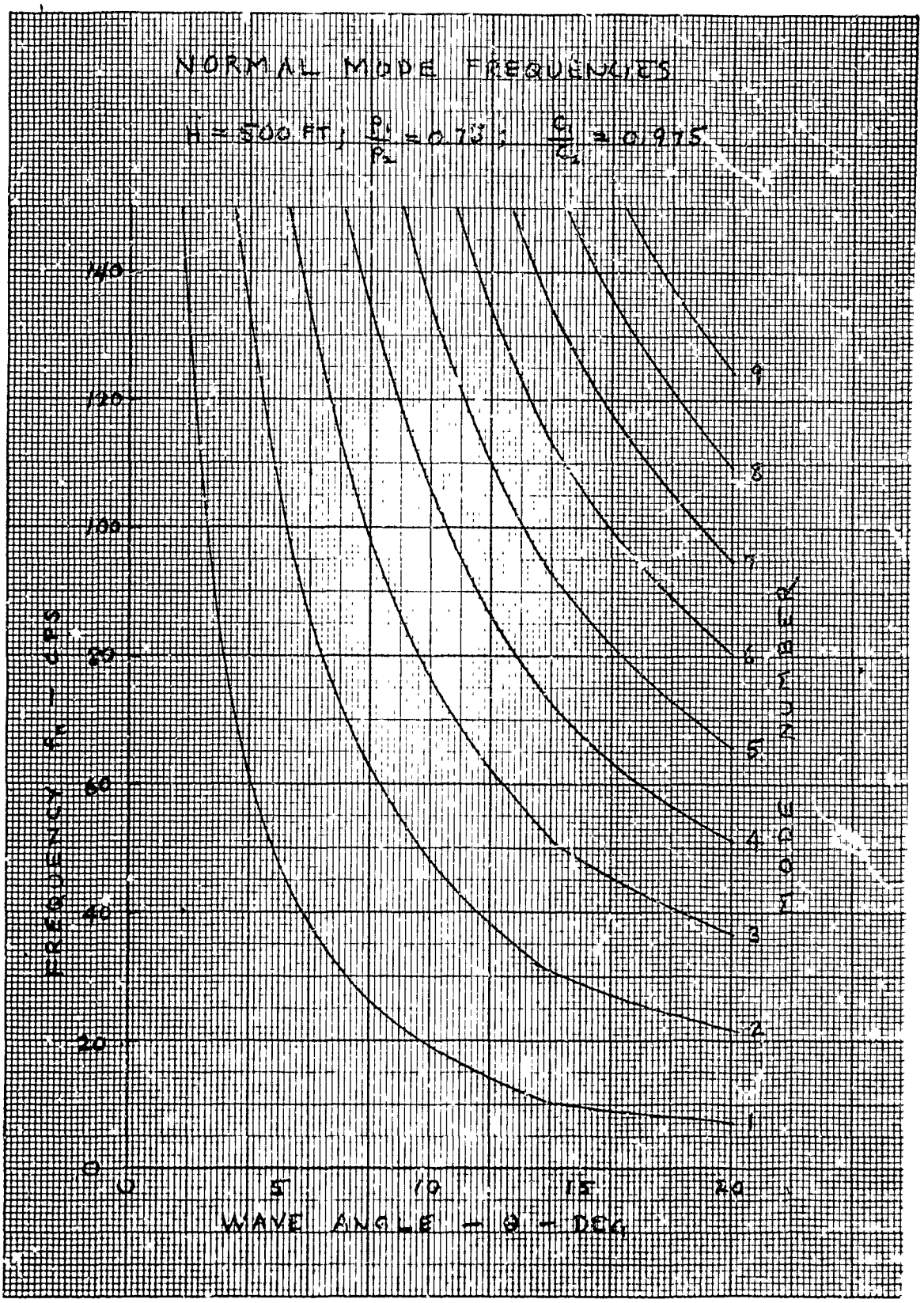
Referring to the above example, suppose a sound source generates single-frequency waves at 60 cps. Reference to the graph will indicate that the first mode will propagate at an angle of 4.1 degrees, the second mode at 8.3 degrees, the third at 12.4 degrees, the fourth at 17 degrees, and so on. If we take a horizontal view of the pattern by looking down from above, we see that the waves are spreading out in all directions from the source. In any vertical plane they are traveling at an angle θ with the horizontal. To visualize the nature of the propagation more clearly, let us reflect the ocean instead of the waves, as illustrated in the sketch below, so that the reflected waves are straightened out as shown by the dotted lines. We see that the waves are spreading out along a cone



in the manner described earlier in the section on cylindrical waves. The propagation loss thus follows the law of cylindrical spreading. It will be noted further that the reflections follow a regular pattern with a skip distance D, where, in our hypothetical example,

$$D = 2H \cot \theta$$

In traversing each skip distance the wave experiences one surface bounce and one bottom bounce. If N_S and N_B represent the reflection loss at the surface and bottom, respectively, the loss occurring in a distance D is $N_S + N_B$. On the average, the attenuation loss per kiloyard (assuming D is expressed in kyd) is


NORMAL MODE FREQUENCIES
h = 500 FT; $\frac{\rho_1}{\rho_2}$ = 0.73; $\frac{c_1}{c_2}$ = 0.975
140
120
100
80
60
40
20
0
FREQUENCY - CYCLES
0
5
10
15
20
WAVE ANGLE - θ - DEG
9
8
7
6
5
4
3
2
1
MODE

$$\alpha_D = \frac{N_S + N_B}{D}$$

if we are interested in long range propagation of low-frequency sound, where the individual reflection losses are small and there are many of them, then α_D may be considered as an ordinary attenuation coefficient and may be added to the attenuation coefficient due to other causes such as absorption by the water.

In situations such as the above example, where a critical angle exists, the bottom loss is theoretically zero when θ is less than the critical angle, but increases rapidly when θ exceeds the critical angle. It is thus apparent that those modes which propagate at angles beyond the critical angle will be rapidly attenuated, so that the number of modes which propagate to long distances may be expected to be quite limited.

It will also be noted from (545) that there is a cut-off frequency below which normal mode propagation cannot occur. The minimum frequency is obtained by setting $n = 1$ and $\theta = 90$ degrees in (545). In this case the phase shift ϵ is zero. Hence the cut-off frequency is

$$f_{min} = \frac{c_1}{4H} \qquad (546)$$

In the numerical example the cut-off frequency is 5000/2000 = 2.5 cps. The wavelength λ_{max} corresponding to f_{min} is

$$\lambda_{max} = \frac{c_1}{f_{min}} = 4H \qquad (547)$$

In the limiting case the waves are traveling vertically up and down and the water depth is a quarter wavelength.

In deriving the condition for constructive interference leading to normal mode propagation we considered only the up-going waves. It is clear that if we had performed our analysis on the down-going waves, we should have arrived at the same result. There is, however, another phenomenon present which has not yet been discussed. This is the interference between the up-going and down-going waves, which results in a pattern of standing

waves in the vertical direction and traveling waves in the horizontal direction. If the angle θ is measured positive upward and the vertical coordinate y is measured positive downward, the pressures of the down-going and up-going waves may be written in the following form

$$p_{down} = p_m e^{j\omega\left[t - \frac{1}{c_1}(x \cos\theta - y \sin\theta)\right]}$$

$$p_{up} = - p_m e^{j\omega\left[t - \frac{1}{c_1}(x \cos\theta + y \sin\theta)\right]}$$

where the minus sign on p_{up} is due to the phase change on surface reflection. The resultant pressure is the sum of the two, which is readily shown to be

$$p = 2j\, p_m \sin\frac{\omega y \sin\theta}{c_1} e^{j\omega\left(t - \frac{x \cos\theta}{c_1}\right)} \qquad (548)$$

The sine factor involving y shows the standing wave pattern in which loops and nodes occur in horizontal planes. The complex exponential shows the traveling wave, which propagates along the x direction with speed V such that

$$V = \frac{c_1}{\cos\theta} \qquad (549)$$

It will be noted that (except where $\theta = 0$) the speed V is always greater than the sound speed c_1. V is called the phase velocity. It must be pointed out, however, that V does not represent the speed at which the energy is being propagated. The energy of the up-going and down-going waves is propagated with speed c_1 in directions making angles $\pm\theta$ with the horizontal. The two waves are carrying equal amounts of energy up and down in the vertical direction, so that the net transfer is zero. The net flow of energy is therefore in the horizontal direction and the speed with which it is propagated is

$$U = c_1 \cos\theta \qquad (550)$$

U is called the group velocity.

Let us now suppose that a number of different discrete frequencies are present in the wave. From the preceding discussion, and in particular from (545), it is seen that the different frequencies will propagate at different angles, i.e., with different values of θ. Therefore the different frequencies will have different phase velocities V and different group velocities U. Suppose that in the numerical example above there were two frequencies, say, 40 and 60 cps. The 40 cps wave propagates at an angle of 6.8 degrees and the 60 cps wave at 4.1 degrees. The respective phase and group velocities are listed in the following table.

f	40 cps	60 cps
θ	6.8 deg.	4.1 deg.
cos θ	.9930	.9974
V	1.0070 c_1	1.0026 c_1
U	.9930 c_1	.9974 c_1

From the table it is seen that the phase of the 40 cps wave propagates (in the horizontal direction) faster than the phase of the 60 cps wave, whereas the energy propagates more slowly. If the two waves are emitted together as a pulse,* it is seen that the pulse will change shape as it travels outward. The energy of the higher frequency component travels faster, so that theoretically if the pulse traveled far enough, the two components would separate completely.

It will be seen that the phase velocity of each wave is larger than the group velocity. The significance of this fact is that as a

* Theoretically a pulse of finite duration always has a finite bandwidth. We here assume that the pulse is sufficiently long that the bandwidth is very narrow.

single-frequency pulse travels outward in a horizontal direction, the wave fronts will continually progress forward from the trailing edge to the leading edge.

The next logical step in the argument is to consider the propagation of a pulse containing a continuous band of frequencies. We shall not analyze this problem mathematically, but a qualitative extension of the preceding argument shows that there is now a continuous spread of phase velocities and group velocities. The method of analysis is to work with the spectrum of the pulse. Suppose, for example, that the amplitude spectrum of a pulse $\psi(t)$ is $\Psi(f)$. This means that the amplitude of that portion of the wave in the infinitesimal frequency band from f to f + df is $\Psi(f)df$. The phase of this portion of the pulse travels in the x direction with the phase velocity V which for any given mode is a function of the frequency. The wave form of the infinitesimal wave is therefore

$$\Psi(f)\, e^{2\pi jf\left[t - \frac{x}{V(f)}\right]} df$$

and the complete pulse, as it travels out is

$$\psi(x,t) = \int_{-\infty}^{\infty} \Psi(f) e^{2\pi jf\left[t - \frac{x}{V(f)}\right]} df$$

We shall not carry the mathematics any further. However, the concept of breaking the pulse down into infinitesimal frequency components provides some qualitative insight into the problem. Each infinitesimal component may be regarded as a wave of a fixed frequency such as was described earlier. Each such component travels with a different group velocity from every other component. In the idealized ocean which we have been considering, the higher frequency components will tend to concentrate toward the leading edge of the pulse and the lower frequency components toward the trailing edge. The pulse thus tends to change shape and spread out as it propagates.

The above phenomenon in which the speed of propagation of a wave varies with the frequency, is called dispersion.

D. Long-Range Propagation Paths in the Ocean

1. Introduction

The performance of sonars which employ the conventional near-surface propagation path is strongly dependent upon the variable thermal conditions which prevail in upper layers of the ocean. When a reasonably deep surface duct exists and the target is within the duct, fairly good performance can be expected. If, on the other hand, the duct is shallow, or the target hides in the thermocline below the duct, performance is poor. The near-surface region is in general not conducive to reliable long-range detection because of the prevalence of negative gradients which cause sound rays to curve downward, creating shadow zones. For this reason there has been increasing interest in the investigation of the various long-range propagation paths provided by nature in the deep ocean. In the following paragraphs we shall investigate briefly the nature of some of these paths.

2. Bottom-Bounce Path

Low frequency sound is observed to propagate over extremely long distances. One of the significant paths responsible for this propagation involves repeated reflections from the surface and bottom. The normal mode propagation discussed previously is an example. Consideration of the bottom-bounce path for active sonar detection involves many factors which cannot be discussed in these notes. However, a few general remarks are in order.

The principal reason for considering such a path is to overcome the difficulties resulting from downward refraction of sound rays due to thermal gradients. If a beam is projected downward at an angle greater than about 10 or 15 degrees, refraction has a relatively minor effect.

In principle, therefore, complete coverage of the surface is possible at all horizontal ranges out to roughly 7 times the water depth. Coverage at even greater ranges can frequently be obtained using shallower tilt angles, but in this region the variable effects of refraction begin to become important.

The principal difficulty of the bottom-bounce path is the high two-way propagation loss. The loss each way consists of the propagation loss in the water plus the bottom reflection loss. The longest ranges are obtained with relatively shallow tilt angles. In this case the propagation loss in the water is very high. As the tilt angle is increased to give shorter ranges, the loss in the water decreases, but at the same time the bottom loss increases due to the larger angles of incidence. These two effects tend to compensate each other, with the result that the overall propagation loss is relatively insensitive to the horizontal range.

For any given type of bottom the propagation loss depends more strongly upon the water depth than upon the horizontal range. The requirements therefore become increasingly severe as the water depth becomes greater. Conversely, as the depth decreases the loss also decreases but at the same time area of coverage also decreases.

The propagation loss also depends strongly upon the reflection characteristics of the bottom, which vary greatly from one area to another.

In summary, the maximum available detection range via the bottom-bounce path is determined primarily by the depth of the ocean. The propagation loss is determined principally by the depth of the ocean and the bottom reflection characteristics and is relatively independent of the horizontal range. For this reason if in a given ocean area a target is not detectable at relatively long ranges, the probability of detection will in general not improve significantly at shorter ranges. Expressed in another way, as the depth of the ocean increases, the area of coverage by the bottom-bounce propagation path also increases, but the probability of detection

decreases over the whole area. The behavior in this respect differs from that of conventional sonars which normally have an area of high-probability coverage beyond which the detection probability is relatively low.

The shape of the ocean bottom also has an effect upon sonar performance. The nature of the effect depends upon the scale of the irregularities. Small scale roughness, such as occurs in rocky areas, tends to scatter sound waves out of the beam, thereby increasing the effective bottom loss. Larger irregularities of the order of the beamwidth may cause focusing or divergence effects. More commonly encountered large irregularities, however, are of such a large scale as to be considered a sloping plane.

A sloping bottom gives rise to range and bearing errors, the relative magnitudes of which depend upon the direction in which the slope occurs. Let ϕ_B denote the direction of the slope. Looking in the direction of the sonar beam, let $\phi_B = 0$ indicate a fore-and-aft tilt and $\phi_B = 90°$ indicate a lateral tilt. Let θ_B denote the amount of the tilt. When $\phi_B = 0$, a positive θ_B signifies a down-in-front tilt. When $\phi_B = 90°$, a positive θ_B denotes a down-to-the-right tilt.

It is evident that the range error will be a maximum and the bearing error zero when $\phi_B = 0$, and the reverse will be true when $\phi_B = 90°$. The slant range to the target is obtained from the observed travel time. Let us call this range r. If refraction is neglected, r consists of two straight-line segments, one from the source (assumed at the surface) to the bottom and the other from the bottom to the target (also at the surface). Let H be the water depth and θ the beam tilt angle at the source. If ϕ_B and θ_B are known, the horizontal range x and the target bearing error ϵ may be computed from simple geometry. Assuming θ_B to be a small angle, the resulting expressions may be simplified to the following

$$x = \frac{2H \cot \theta \, (1 + \theta_B \cos \phi_B \tan \theta)}{1 - 2\,\theta_B \cos \phi_B \cot \theta} \qquad (551)$$

$$\epsilon = \theta_B \sin \phi_B \tan \theta \qquad \text{(radians)} \qquad (552)$$

If the bottom is assumed to be level* at depth H, the horizontal range x' computed from the same slant range r will be slightly different. To the same approximation it is

$$x' = \frac{2H \cot \theta \, (1 + 2\,\theta_B \cos \phi_B \tan \theta)}{1 - 2\,\theta_B \cos \phi_B \cot \theta} \qquad (553)$$

The bearing angle is zero in this case, of course. Combining (551) and (553), we find the fractional error in range to be

$$\frac{x' - x}{x} = \theta_B \cos \phi_B \tan \theta \qquad (554)$$

From (552) and (554) we see that for small tilt angles the maximum range and bearing errors are proportional to the tangent of the beam tilt angle. As a numerical example, suppose θ_B is 5 degrees and θ is 45°. In this case the bearing error for a lateral slope is 5 degrees and the range error is 8.7 percent. Although this is a rather extreme example, it is clear that significant errors can be introduced if the slope of the bottom is not accurately known.

3. SOFAR Path

The SOFAR path, referred to previously, is the propagation path which employs the deep sound channel, whose axis is normally located at depths of 3000 to 4000 feet. When a source and receiver are both located in the channel, propagation conditions are excellent. A small explosive charge may be heard at distances of as much as 2000 miles. Although the depths involved are not of current interest for submarine detection, the SOFAR path has a useful function in determining the location of distant sound sources at sea. If, for example, a downed aviator or a vessel in distress at sea drops an explosive charge which is set to detonate near the axis of the channel, and the time of arrival at three different locations is

* This assumption must be made if the bottom slope is unknown.

accurately measured, the geographical position of the explosion can be determined in a manner similar to LORAN.

It is of interest to note in this connection that since the channel is very wide there exist a number of paths between the source and receiver, each characterized by a different number of refraction cycles, or "orders." Two such paths are illustrated in the sketch below. The path



of lowest order leaves the source with the largest initial angle and executes the largest excursions away from the axis. As the order increases, the range of excursion decreases and the rays converge toward the axis. To investigate the significance of this effect, let us consider the simple case where the source and receiver are located on the axis and the velocity gradient has a constant positive value g below the axis and a constant negative value -g above it. Let c_o be the sound speed on the axis.

A ray whose initial angle is θ_1 will, according to (456a), traverse a horizontal distance

$$\frac{c_o}{-g \cos \theta_1}(-\sin \theta_1 - \sin \theta_1) = \frac{2c_o}{g} \tan \theta_1$$

between the source and the first crossing of the axis. The horizontal range corresponding to one refraction cycle is twice this distance

$$x_1 = \frac{4c_o}{g} \tan \theta_1 \qquad (555)$$

The corresponding time, according to (456d), is

$$t_1 = \frac{2}{g} \log_e \frac{1 + \sin \theta_1}{1 - \sin \theta_1} \qquad (556)$$

Suppose this ray undergoes n_1 refraction cycles between the source and receiver, while a second ray having an initial angle θ_2 undergoes n_2 cycles in traveling the same distance. Then

$$\frac{4\, n_1\, c_o}{g} \tan \theta_1 = \frac{4\, n_2\, c_o}{g} \tan \theta_2$$

or

$$\frac{n_2}{n_1} = \frac{\tan \theta_1}{\tan \theta_2} \tag{557}$$

The total travel times of the two rays are

$$T_1 = \frac{2n_1}{g} \log_e \frac{1 + \sin \theta_1}{1 - \sin \theta_1}$$

$$T_2 = \frac{2n_2}{g} \log_e \frac{1 + \sin \theta_2}{1 - \sin \theta_2}$$

The ratio of the two travel times is

$$\frac{T_2}{T_1} = \frac{n_2 \left[\log_e (1 + \sin \theta_2) - \log_e (1 - \sin \theta_2)\right]}{n_1 \left[\log_e (1 + \sin \theta_1) - \log_e (1 - \sin \theta_1)\right]}$$

or, by (557),

$$\frac{T_2}{T_1} = \frac{\tan \theta_1 \left[\log_e (1 + \sin \theta_2) - \log_e (1 - \sin \theta_2)\right]}{\tan \theta_2 \left[\log_e (1 + \sin \theta_1) - \log_e (1 - \sin \theta_1)\right]} \tag{558}$$

If we expand the logarithms in infinite series, (558) becomes

$$\frac{T_2}{T_1} = \frac{\tan \theta_1 \left(\sin \theta_2 + \frac{1}{3} \sin^3 \theta_2 + \cdots\right)}{\tan \theta_2 \left(\sin \theta_1 + \frac{1}{3} \sin^3 \theta_1 + \cdots\right)}$$

$$= \frac{\cos \theta_2 \left(1 + \frac{1}{3} \sin^2 \theta_2 + \cdots\right)}{\cos \theta_1 \left(1 + \frac{1}{3} \sin^2 \theta_1 + \cdots\right)} \tag{558a}$$

It can be seen by substituting reasonable numerical values into (558a) that if $\theta_2 < \theta_1$, then $T_2 > T_1$. Since the angles involved are small, we may also show this by making the usual small angle approximations

$$\cos \theta \approx 1 - \tfrac{1}{2}\theta^2 \text{ and } \sin \theta \approx \theta$$

Thus

$$\frac{T_2}{T_1} \approx \frac{1 - \frac{1}{6}\theta_2^2}{1 - \frac{1}{6}\theta_1^2} \tag{558b}$$

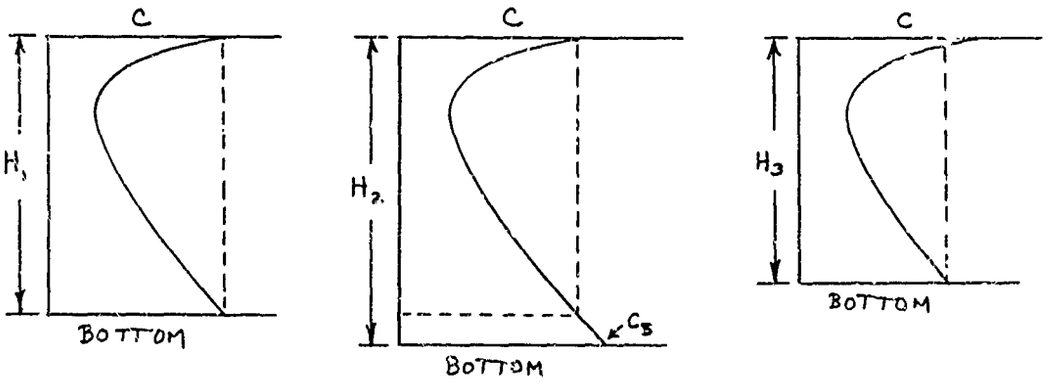
From this result we see that the lowest order SOFAR ray, which leaves the source at the largest angle and travels farthest away from the channel axis, arrives at the receiver first. This may seem strange since the ray actually travels a greater distance than any of the others. The explanation is that it is traveling through a region of higher sound speed.

If we look at the travel times of the rays of increasing order, we see that they decrease and approach a fixed limit, $\frac{X}{c_0}$ (the time to travel in a straight line along the channel axis). Thus, although the SOFAR propagation of a pulse is a multipath phenomenon with different paths arriving at different times, it ends abruptly. The abrupt ending makes accurate timing possible.

4. Convergence Zone Path

The width of the SOFAR channel depends upon the relative values of the sound speed in the vicinity of the surface and at the bottom of the ocean. The maximum near-surface value normally occurs in the first few hundred feet and is determined chiefly by the surface temperature. In the deep ocean the water at the bottom is very nearly isothermal at the temperature of maximum water density, a few degrees above the freezing point. The sound speed at the bottom is therefore determined almost exclusively by the pressure, which is proportional to the ocean depth.

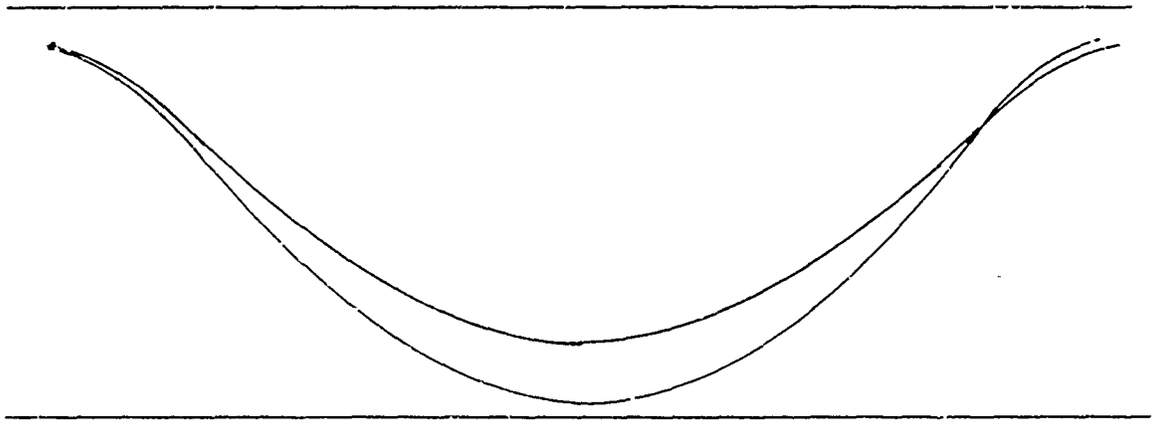
If the ocean depth had the critical value H_1 indicated in the left-hand velocity profile shown below, such that the sound speed at the bottom were equal to the maximum value near the surface, the SOFAR channel would extend all the way to the bottom.



encompassing the region indicated by the dotted line. Where the depth is greater than H_1, as indicated by H_2 in the second profile, the width of the channel is determined by the near-surface maximum. Assuming the two profiles are the same, the width of the channel is the same in both cases, the only difference being that in the second case there is a region of depth excess below the bottom of the channel. In the third case shown at the right the depth H_3 is less than the critical depth and the width of the channel is limited by the bottom, as indicated by the dotted line.

Let us now assume that the ocean is very deep (second case above) and let us place a sound source near the surface in the vicinity of the depth of the maximum sound speed. Let C_B denote the sound speed at the ocean bottom. Then there will be a limiting ray having a vertex velocity equal to C_B, which will just graze the bottom. In fact, there will be two such rays, a direct downward-going ray and a surface-reflected upward-going ray. All rays which leave the source at angles beyond this limiting angle will have vertex velocities greater than C_B and will strike the bottom. On the other hand, those rays which leave the source with initial angles more shallow than the limiting angle will be refracted upward without striking the bottom and will remain within the channel. This type of propagation in the deep sound channel has the remarkable property that all of the rays (within the relatively narrow angular limits) tend to return to the surface at approximately the same horizontal range, as illustrated in the sketch below. The result is a very pronounced focusing or convergence effect. If the sound is radiated in all directions horizontally, the region of convergence is a narrow annular ring, called a convergence zone.

The radius of a convergence zone varies somewhat with geographical location. In the temperate regions the zone radius is generally between 30 and 35 miles. In the colder regions it tends to be somewhat smaller, 25 miles or so being a representative value. The detailed



acoustic structure of a convergence zone is rather complicated, being dependent upon depth excess at the bottom of the ocean, the source and receiver depths, and the local thermal structure in the near-surface region. As a rough rule of thumb, the annular width of a zone is of the order of 10 percent of the range and the gain due to the convergence (the spreading anomaly) is of the order of 10 db - a very significant enhancement.

Convergence zone propagation is not limited to a single zone. As the rays spread out from a source they undergo repeated cycles of refraction, with the result that second, third, fourth, etc. zones occur at regular intervals. The spreading loss from one zone to the next follows approximately the cylindrical spreading law.

In order for convergence zone propagation to exist at all it is necessary that the speed of sound at the bottom be larger than at any depth above the SOFAR channel axis. This condition places a requirement on the ocean depth in relation to the surface temperature. As an example

of such a condition, the speed of sound at the surface is related to the temperature (in °F) by (447) with y = 0,

$$c = 4422 + 11.25T - .045T^2 \text{ ft/sec} \tag{559}$$

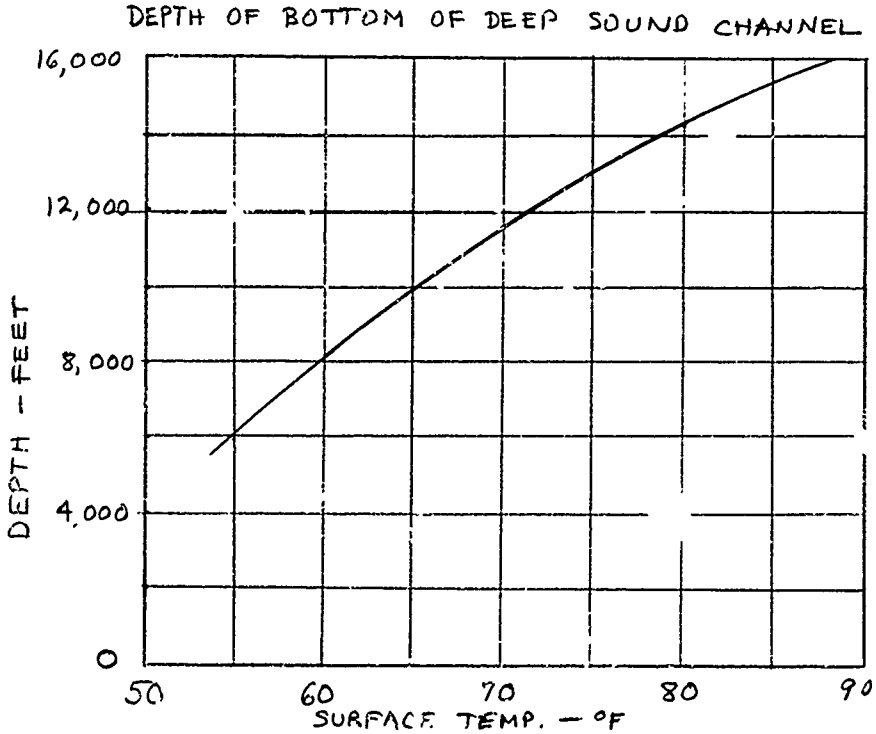
(neglecting the salinity term). In the deep ocean c may be expressed as a function of depth by the approximate formula

$$c = 4768 + 0.0182y + 82.3\, e^{-1.8 \times 10^{-4}} \text{ ft/sec} \tag{560}$$

where y is in feet. The minimum depth

$$y = H_{min}$$

at which convergence zone propagation can occur may be related to the surface temperature by equating (559) and (560). The resulting curve is shown in the accompanying graph. Since surface temperatures vary with



the season of the year, the depth requirements are likewise variable. There are some areas of the ocean where convergence zone propagation occurs the year around; there are other areas where it never occurs; and there are still other areas where it occurs in the winter but not in the summer. The depth requirements are most severe in the tropical

regions and least severe at the higher latitudes. As the polar regions are approached, the surface temperatures become lower until an approximately isothermal condition is reached where a positive velocity gradient exists all the way to the bottom, and the convergence zone propagation changes to a combination of upward refraction and surface reflection.

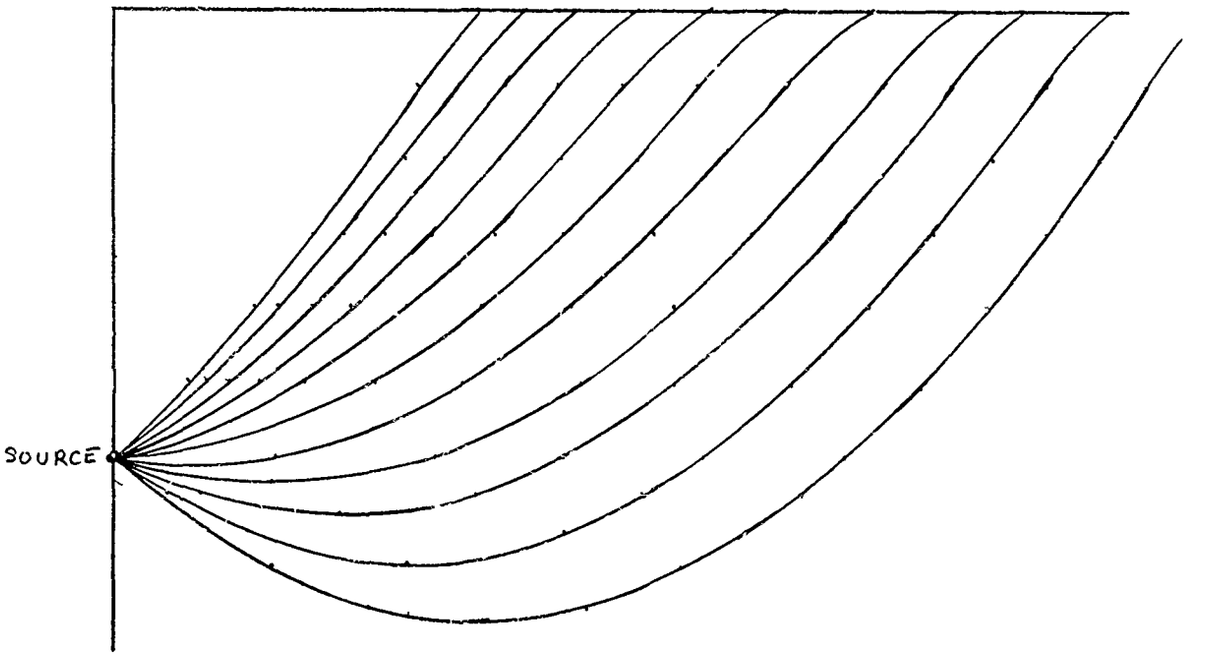
The depth criterion outlined above represents only an absolute minimum requirement that at least one ray be propagated to the convergence zone. The amount of sound propagated depends upon the extent of the depth excess between the minimum required depth and the ocean bottom, since all rays reaching the convergence zone must pass through this region. The same conclusion may be reached by considering the rays as they leave the sound source. The greater the ocean depth the larger will be the initial angle of the limiting ray which grazes the bottom. Therefore a greater fraction of the energy radiated by the source will be retained in the deep sound channel without striking the bottom. It has been found that a depth excess of several hundred fathoms is required for satisfactory operational use of convergence zone propagation.

5. Reliable Acoustic Path (RAP)

In our discussion of convergence zone propagation we have seen the value of the refractive properties of the deep ocean in providing a long-range acoustic path between a source and a distant receiver, both located at the surface. The convergence zone path has one serious limitation, however; the surface coverage is limited to a narrow annular zone. The entire circular area inside the zone, except for a small region of direct coverage in the vicinity of the source, is blank.

Is there any way of making use of refraction in the deep ocean to obtain solid coverage of the surface? The answer to this question is "yes," provided the source is located at a great depth. Consider a location where the ocean depth is in excess of the minimum requirement for convergence zone propagation and let the sound source be placed at this

minimum depth, i.e., at the depth where the sound speed has a value equal to its maximum value near the surface. Then the ray which leaves the source horizontally will traverse a path similar to half of a convergence zone path, reaching the surface (actually, the depth of maximum sound speed near the surface) at a horizontal range equal to half the radius of the first convergence zone. All rays which leave the source at angles above the horizontal will have vertex velocities larger than that of the horizontal ray and will therefore reach the surface at finite angles of incidence. Furthermore, all rays which leave the source at downward angles between the horizontal and the limiting ray at the bottom will also be refracted upward and will reach the surface at finite angles of incidence. A family of such rays is shown in the sketch below. We see therefore that when the ocean is deep enough to provide



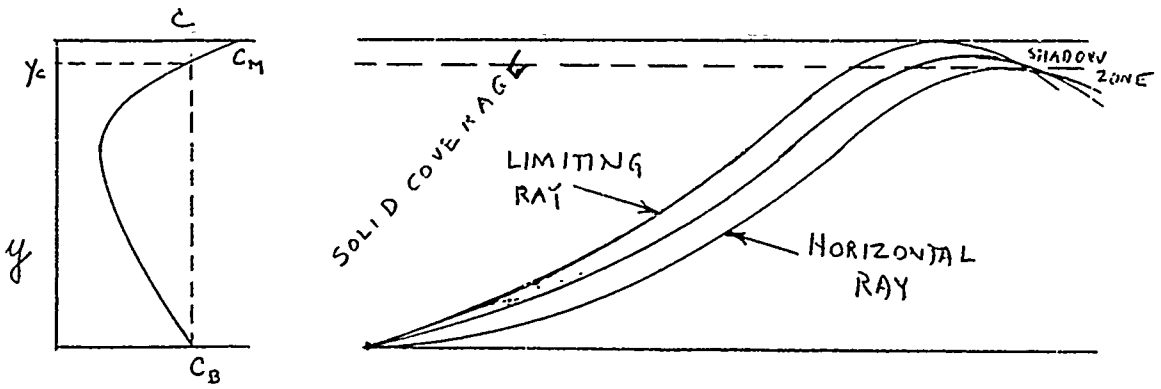
convergence zone propagation, solid coverage of the surface out to horizontal ranges in excess of one-half a convergence zone radius can be obtained by placing the source at the critical depth defined above. (Actually, as a practical operational measure the source should be placed a little deeper to provide a margin of safety.) This path has been given the name reliable acoustic path and is usually designated by the initials RAP.

In addition to providing solid coverage of the surface the RAP has the advantage of providing a certain amount of ray convergence. The maximum convergence gain occurs for those rays which leave the source in the vicinity of the horizontal. Calculations based on typical conditions in the Atlantic Ocean show a gain of about 5 db. As the initial angle increases above the horizontal the convergence gain decreases and the spreading loss approaches the spherical type. The reduction of gain in this region is of little importance, however, since the overall propagation loss decreases rapidly because of the reduction in range. In the case of the downward-going rays, on the other hand, the reduction in convergence gain combines with the increase in range to cause a rapid increase in overall propagation loss. Therefore, although RAP ranges in excess of 25 miles may exist in very deep water, the requirements for implementing them become very severe.

The reliable acoustic path is sometimes spoken of as being half of a convergence zone. This statement is misleading for two reasons. The first has already been discussed. Unlike the convergence zone path the RAP provides solid coverage of the surface out to ranges in excess of half a convergence zone, the theoretical limit depending upon the depth excess. The second reason is also a significant one. In areas of insufficient water depth the convergence zone path ceases to exist and a convergence zone sonar becomes utterly useless (except for the small area of conventional direct coverage). This is not true for the RAP. Although the extent of the surface coverage is considerably reduced under such conditions, substantial performance is still obtainable.

When the water depth is insufficient for convergence zone operation, that is, when the sound speed at the bottom is less than the maximum value near the surface, the sound source should be placed as close as practicable to the bottom. A rough velocity profile and ray diagram

illustrating the case under discussion is shown in the sketch below. The



sound speed at the bottom is c_B and the maximum value in the surface region is c_M. The extent of the deep sound channel is determined by the vertical dotted line running from the bottom to the depth y_c. A ray leaving the source horizontally at the bottom will vertex at the depth y_c at a horizontal range determined by the velocity profile. At depths shallower than y_c at this range there will be a shadow zone. A ray which leaves the source at a slight upward angle will vertex at a slightly shallower depth and a somewhat shorter horizontal range. Thus as the horizontal range decreases the shadow zone becomes shallower until the limiting ray to the surface is reached. The initial angle of the limiting ray is

$$\theta_L = \cos^{-1}\left(\frac{c_B}{c_M}\right)$$

All rays leaving the source at angles above θ_L will strike the surface. Hence solid coverage of the surface is provided at all ranges out to the point of tangency of the limiting ray.

In regions where the depth is greater than about 5000 or 6000 feet, the value of θ_L is determined chiefly by the relative values of the water depth and surface temperature, as indicated previously. In shallower water the situation becomes more variable. However, in most ocean areas the limiting angle is less than 10 to 15 degrees. For efficient use of a

vertically directional sound source in areas of insufficient water depth it is necessary to steer the beam upward at an angle sufficient to insure that most of the radiated acoustic energy is directed into the region of surface coverage.

The reliable acoustic path, while extremely attractive from an acoustic point of view, has two obvious disadvantages. First, there is the technical problem of designing equipment to operate at great depths and of supplying power to it. Secondly, if the path is to be considered for application to mobile systems, there is the operational problem of lowering transducers to such depths and, if necessary, of retrieving them.

6. Refracted Surface-Reflected (RSR) Path

The term refracted surface-reflected path is used to describe those propagation paths which involve a sequence of cycles of refraction in the deep ocean and reflection at the surface. It is evident that to a limited extent paths of this type are included in convergence zone propagation. The present discussion will be limited to a consideration of the RSR path as an extension of the RAP.

A necessary condition for RSR propagation to occur is the existence of a substantial depth excess. Let us assume this condition to be met and let the source be placed at a depth slightly below the bottom of the deep sound channel (i.e., slightly below the minimum depth for convergence zone propagation). Let the sound speed be c_o at the source and c_B at the ocean bottom. The cosine of the initial angle of the bottom limiting ray is

$$\cos \theta_L = \frac{c_o}{c_B}$$

All rays which leave the source at downward angles between 0 and θ_L will be refracted to the surface without striking the bottom. After reflection from the surface they will return downward and be refracted again, continuing the process in repeated cycles with about the same range increment as convergence zones. Those rays which leave the source at upward

angles between 0 and θ_L will travel to the surface, be reflected, and proceed in similar RSR cycles. Rays which leave the source at downward angles greater than θ_L will strike the bottom directly and rays which leave the source at upward angles in excess of θ_L will strike the bottom after a surface reflection. The RSR propagation in this case is therefore limited to those rays having initial angles between $\pm \theta_L$.

We see that at the first surface reflection the surface coverage consists of an annulus whose width extends from the horizontal range of the $+ \theta_L$ ray to the horizontal range of the $- \theta_L$ ray. Since the distance traveled per refraction cycle is substantially the same for all rays in this group, the annuli of the zones at the second, third, etc. reflections will have roughly the same width as the first. It is also clear that the zone width we are talking about depends strongly upon the depth excess. However, a relatively moderate depth excess will give substantially broader coverage than can be obtained with convergence zone propagation. For example, under typical Atlantic Ocean conditions depth excess values of about 1000, 1750, and 4000 feet will provide zone widths equal to 1/4, 1/3, and 1/2, respectively, of the spacing between successive zones.

Obviously a price must be paid to achieve such broad coverage as compared to convergence zones, and the price is the absence of the high convergence gain. The requirements placed on equipment for implementing such a long-range path would be very severe.

TRANSDUCERS

A transducer may be defined in general as an element in an energy transmission system, which connects one portion of the system - the source - to a second portion the load. According to this general definition it is not necessary that a transducer convert energy from one form to another. However, in these notes we shall be concerned with transducers in which such a conversion takes place. For transmitting acoustic signals in the water we are interested in converting electrical energy into acoustic energy, and for receiving acoustic signals from the water we are interested in the reverse process.

A transducer which converts electrical energy into acoustic energy, or vice versa, is called an electroacoustic transducer. An electroacoustic transducer which is used to transmit sound waves into the water is called a sonar projector. An electroacoustic transducer which is used to receive sound waves from the water is called a hydrophone. Many transducers can be used interchangeably for both purposes; these will be referred to simply as sonar transducers. It should be noted, however, that the word transducer is sometimes used in a restricted sense to designate only a projector, as distinguished from a hydrophone.

A. The Nature of Sonar Transducers

1. Introduction

Because the specific acoustic impedance of water is several orders of magnitude larger than that of air, the conventional techniques employed in the design of air-audio equipment (microphones and loudspeakers) are not applicable to sonar. Since air is a very light substance, the driving mechanism must produce a large displacement with very little force. In underwater sound, on the other hand, it is necessary to apply a very large

force to generate even a small displacement. In other words, a sonar transducer must have a large mechanical impedance.

Among the physical phenomena which are capable of fulfilling this requirement, two are widely used. These are electrostriction and magnetostriction.

a. Electrostriction. In 1880 Jacques and Pierre Curie discovered that when certain crystals such as quartz, tourmaline, and Rochelle salt, are subjected to pressure, electric charges of opposite sign appear on opposite faces, thus producing a potential difference. Conversely, when a potential difference is applied to opposite faces, a change in physical dimensions occurs. Such a crystal is called a piezoelectric crystal, and transducers made from such crystals are called piezoelectric transducers.

There are other materials such as barium titanate, which, though not electrostrictive in their natural condition, can be made so if they are subjected to a strong electric field under suitable conditions. Such materials are said to be polarized.

b. Magnetostriction. When certain materials such as nickel are placed in a magnetic field, they undergo a change in physical dimensions. Therefore, when a properly designed nickel element is subjected to an oscillating magnetic field, a mechanical oscillation is produced which is capable of generating acoustic waves in the water.

Also, the converse process takes place. When a piece of magnetostrictive material is subjected to a physical pressure, a magnetic field is generated, and this can be detected electrically. The phenomenon of magnetostriction has been known for over a century.

The design of a piezoelectric or magnetostrictive transducer consists of the application of the appropriate oscillating electric or magnetic driving force to the mechanical vibrating element, one face of which is coupled to the water. In the case of crystals which are soluble in water

the coupling is accomplished by means of an intermediate liquid such as castor oil, which has a specific acoustic impedance approximately equal to that of water and which is encased within an acoustically transparent membrane. The other face of the vibrating element is either attached to a heavy mass backing, or else is backed up by pressure-release material which allows the face to vibrate freely without transmitting any appreciable acoustic radiation. (Ideally the element should terminate in a vacuum.)

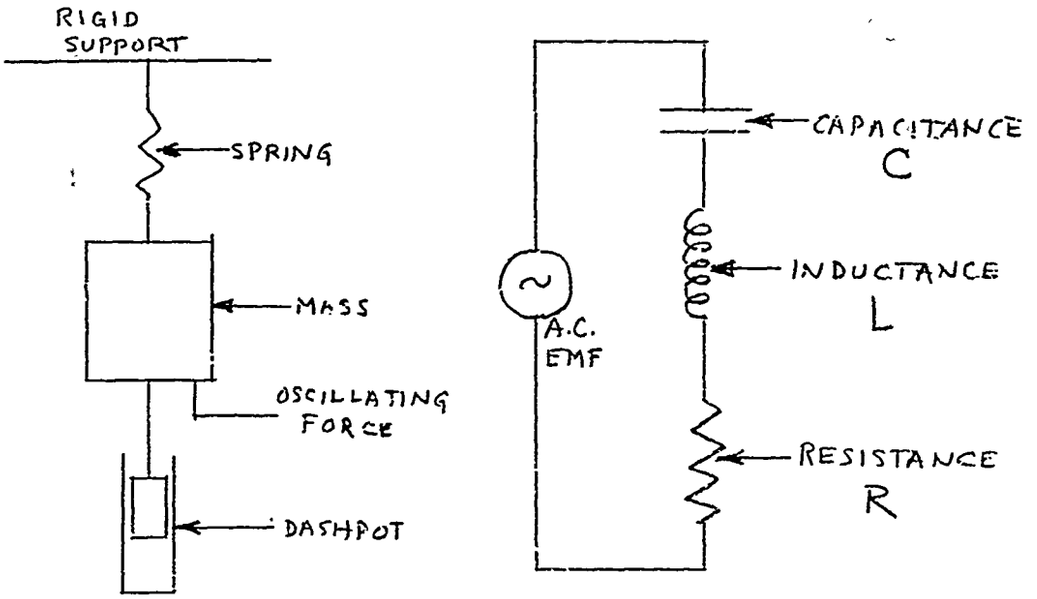
For maximum efficiency the vibrating element should operate at the resonant frequency. This can be done with transducers which are designed to operate at a fixed single frequency. Transducers which transmit or receive over a broad band of frequencies must operate well away from the resonant point. Most sonobuoy hydrophones, for example, must have a flat response over a very wide range of frequencies.

No attempt will be made in these notes to give a complete description of the theory and design of sonar transducers. The scope will be limited to a discussion of the basic concepts together with a rather oversimplified presentation of the theory of operation. The operation of a sonar transducer is most conveniently described in terms of its equivalent electric circuit, in which the mechanical effects are represented in terms of their equivalent electrical characteristics as seen looking into the electrical input terminals. As an introduction to this discussion we shall first consider the mechanical impedance of a simple vibrating system and shall then discuss the effect of the radiation load, that is, of the power transmitted into the water as acoustic radiation.

2. Mechanical Impedance

Consider the mechanical system shown in the sketch below. A mass M is supported by a spring from a rigid support. The mass is driven by a sinusoidally oscillating force $F e^{jwt}$*, causing it to oscillate in a vertical direction. The mass is also attached to a piston in a dashpot which produces a retarding force proportional to the velocity. Let ξ be the displacement of

*w = angular frequency = $2\pi f$



MECHANICAL SYSTEM ELECTRICAL SYSTEM

the mass from its static equilibrium position, and let u be the velocity, that is,

$$u = \dot{\xi} = \frac{d\xi}{dt}$$

The following forces act on the mass m:

Externally applied driving force = $F\, e^{jwt}$

Tension in the spring = $- s\xi$

Resistance of the dashpot = $- R_m\dot{\xi}$

where s and R_m are constants. According to Newton's Second Law, we see that

$$F e^{jwt} - s\xi - R_m\dot{\xi} = M\ddot{\xi}$$

or

$$M\ddot{\xi} + R_m\dot{\xi} + s\xi = Fe^{jwt} \qquad (601)$$

Let us assume a steady-state solution of the form

$$\xi = Ae^{jwt} \qquad (602)$$

The complex constant A is evaluated by substituting (602) into the differential equation (601).

$$A = \frac{F}{jwR_m - w^2 M + s} \tag{603}$$

We are particularly interested in the velocity, u, which is

$$u = \frac{d\xi}{dt} = jwAe^{jwt}$$

If we write u in terms of its complex amplitude U , that is,

$$u = Ue^{jwt} \tag{604}$$

we find the relation between U and F to be

$$U = \frac{F}{R_m + j\,(wM - \frac{s}{w})} \tag{605}$$

In the preceding illustration a simple series RLC electrical circuit is shown at the right of the mechanical diagram. If the alternating emf applied to the circuit is

$$e = Ee^{jwt}$$

and the alternating current flowing in the circuit is

$$i = Ie^{jwt}$$

where I is the complex amplitude of the current, the relation between I and E is

$$I = \frac{E}{R + j\,(wL - \frac{1}{wC})} \tag{606}$$

where R, L, and C are the resistance, inductance, and capacitance of the circuit. According to standard A.C. electrical theory, the denomicator of (606) is the complex impedance $\overline{z}$ of the circuit and is expressed in terms of the resistance and the reactance,

$$\overline{z} = R + j\,X \tag{607}$$

where, for this series circuit

$$X = wL - \frac{1}{wC} \tag{607a}$$

By analogy we may define a complex mechanical impedance

$$\bar{z}_m = R_m + j\, X_m \tag{608}$$

where

$$X_m = wM - \frac{s}{w} \tag{609}$$

The analogy between the mechanical and electrical circuits is shown in the following table:

Mechanical	Electrical
Applied force, Fe^{jwt}	Applied EMF, Ee^{jwt}
Velocity, Ue^{jwt}	Current, Ie^{jwt}
Damping constant, R_m	Resistance R
Mass, M	Inductance, L
Stiffness coefficient, s	$\frac{1}{\text{Capacitance}} \cdot \frac{1}{C}$
Complex mechanical impedance, $\bar{z}_m$	Complex impedance, $\bar{z}$
$Z_m = \lvert\bar{z}_m\rvert$	$Z = \lvert\bar{z}\rvert$
wM	Inductive reactance, wL
$\frac{K}{w}$	Capacitance reactance $\frac{1}{wC}$

In a sonar transducer the velocity $u = Ue^{jwt}$ is the velocity of the radiating surface. The complex mechanical impedance $\bar{z}_m$ is the ratio of the complex driving force to the complex velocity. The magnitude of this impedance

$$Z_m = |z_m| = \sqrt{R_m^2 + X_m^2}$$

is the ratio of the rms force to the rms velocity.

From this discussion it is clear that the inertia of the moving parts of an electroacoustic transducer will appear as an inductance in the equivalent circuit; friction in the moving parts, or any other process (such as hysteresis in magnetic materials) which dissipates mechanical energy, will appear as a resistance; and elastic properties of materials will appear inversely

as a capacitance. As a matter of fact, when electrical impedance measurements are made on a transducer, these effects are actually observed, the equivalent resistance, inductance, and capacitance being proportional to the corresponding mechanical quantities. We shall investigate these relationships for a piezoelectric transducer in a later section.

3. Radiation Impedance

In the preceding section we discussed the vibration of a mechanical system without any reference to the radiation of acoustic waves. That discussion may be considered as applying to a transducer operating in a vacuum. When the transducer is operating in its normal environment, the reaction of the water against the radiating surface imposes a load on the vibrating element, thereby altering the mechanical impedance of the system. Even without any mathematical analysis it is intuitively apparent that the radiation of acoustic power should give rise to a load which includes a resistance, since no power is dissipated in a pure reactance.

To obtain a little better insight into this problem, let us temporarily assume that the radiating surface is flat and has an area S. Let us assume further that the dimensions of the surface are large compared with the wavelength of the acoustic waves in the water. Under these assumptions the waves in the immediate vicinity of the radiating surface are essentially plane waves whose specific acoustic impedance is ρc, where ρ and c refer to the density and sound speed of the water. As we have seen earlier, the specific acoustic impedance is the ratio of the acoustic pressure to the particle velocity (of plane waves).

Now, since the water is assumed to be in contact with the radiating surface, and since by our assumption that the acoustic waves are essentially planar, it follows that the particle velocity of the water in the immediate vicinity of the radiating surface is the same as the velocity of the surface itself. Furthermore, the acoustic pressure of the water must equal the pressure exerted by the radiating surface, and hence the total force is equal

to the pressure times the area. If F represents the amplitude of the force exerted by the transducer on the water, and $\hat{U}$ denotes the amplitude of the velocity, then we see that

$$F = \rho c S \hat{U}$$

and the radiation impedance is

$$Z_r = \frac{F}{\hat{U}} = \rho c S \tag{610}$$

The force F is the force required to move the water and is in addition to the internal force required to drive the transducer mechanism. The overall mechanical impedance of the system consists of the combination of the internal impedance and the radiation impedance. It should be noted that the above definition of radiation impedance (i.e., as the ratio of the force to the particle velocity) is identical with the definition of mechanical impedance presented earlier, in the discussion of three-dimensional acoustic waves.

Under the preceding ideal assumptions - large flat surface - the radiation impedance is found to be a pure resistance. The analysis of more practical examples, where the dimensions of the radiating surface are of the same order of magnitude as a wavelength, is in general very complicated and will not be included in these notes. It is found in general that edge effects contribute an out-of-phase component represented by an inductive reactance. The radiation impedance is of the form

$$\bar{z}_r = \rho c S (r_1 + j x_1) \tag{611}$$

where r_1 and x_1 are dimensionless parameters representing the relative magnitudes of the resistive and reactive components. For example, for the case of a circular piston mounted in an infinite baffle, * the parameters r_1 and x_1 are functions of the product ka, where a is the radius of the circular

- - - - - - - - -

*That is, flush-mounted in an infinite rigid plane wall, with the acoustic radiation confined to the space on one side of the wall.

piston face and k is the wave number, $2\pi/\lambda$. These functions are plotted on the accompanying graph. For small values of ka (less than about 0.7), x_1 varies linearly and is the dominant term. As ka increases, x_1 reaches a maximum value of about 0.7 in the vicinity of ka = 1.5, and then gradually decreases in a series of damped oscillations toward zero. The resistive parameter r_1 increases at first as the square of ka. It reaches a maximum value of about 1.1 and then gradually approaches the value 1 in a series of damped oscillations. The behavior of r_1 and x_1 at large values of ka agrees with the results of our intuitive discussion which led to a pure resistance with $r_1 = 1$ and $x_1 = 0$.

Another interesting example is the case of a pulsating sphere which generates omnidirectional spherical waves. In our earlier analysis of spherical waves, we found that the acoustic pressure of a spherical wave is out of phase with the particle velocity, leading to a complex unit-area acoustic impedance.

$$z_a = \frac{\rho c}{1 + \frac{1}{k^2 r^2}} + j \frac{\frac{\rho c}{kr}}{1 + \frac{1}{k^2 r^2}} \tag{612}$$

If these waves are generated by a pulsating sphere whose mean radius is a, then the mechanical impedance per unit area of the spherical radiating surface is given by (611) with r = a, and the overall mechanical impedance is given by (611) with

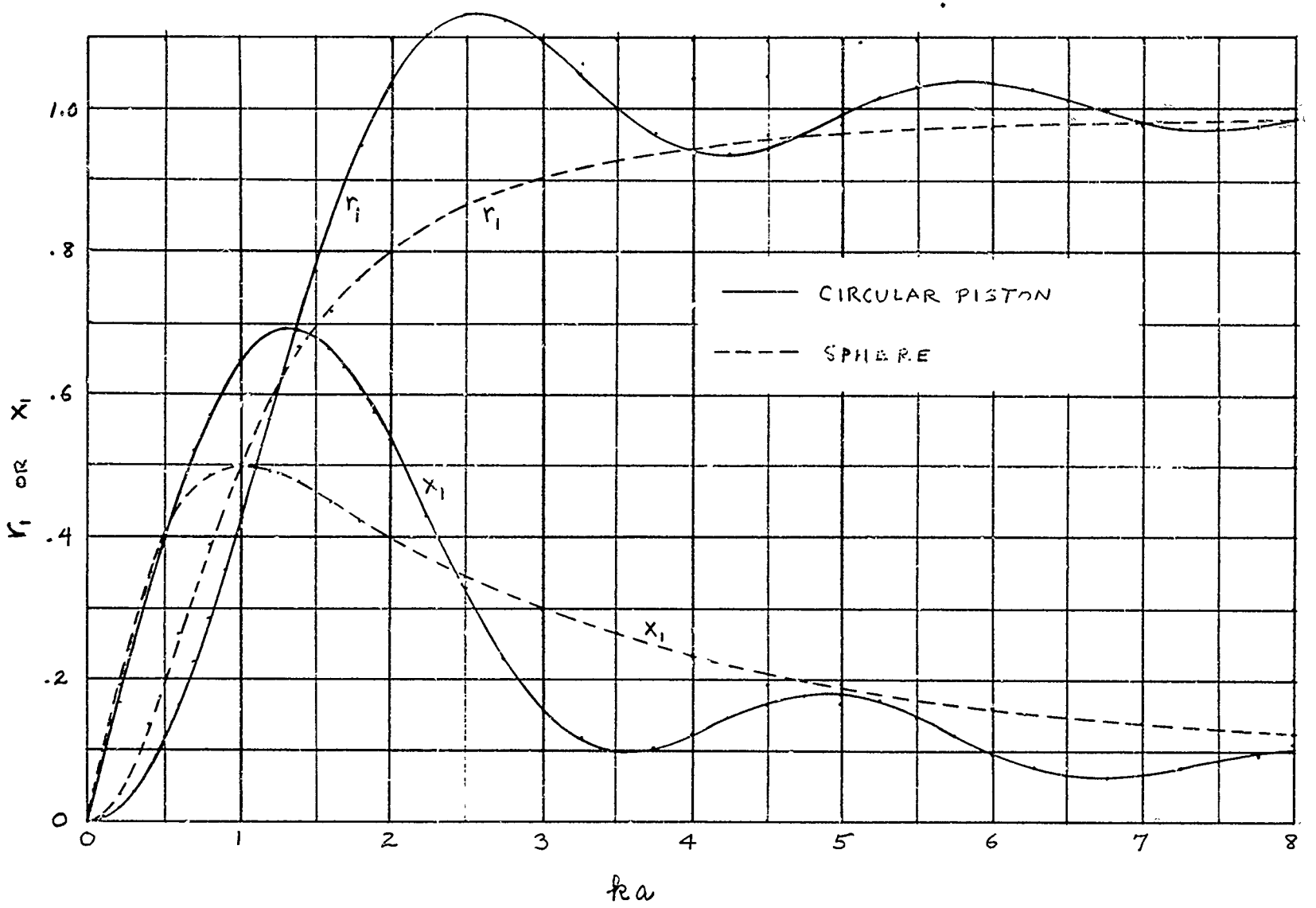
$$S = 4\pi a^2$$

$$r_1 = \frac{1}{1 + \frac{1}{k^2 a^2}} = \frac{k^2 a^2}{1 + k^2 a^2} \tag{613a}$$

$$x_1 = \frac{\frac{1}{ka}}{1 + \frac{1}{k^2 a^2}} = \frac{ka}{1 + k^2 a^2} \tag{613b}$$

These functions are plotted as dashed curves on the accompanying graph. A comparison of the two sets of curves shows a quite similar behavior for both

RADIATION IMPEDANCE FUNCTIONS



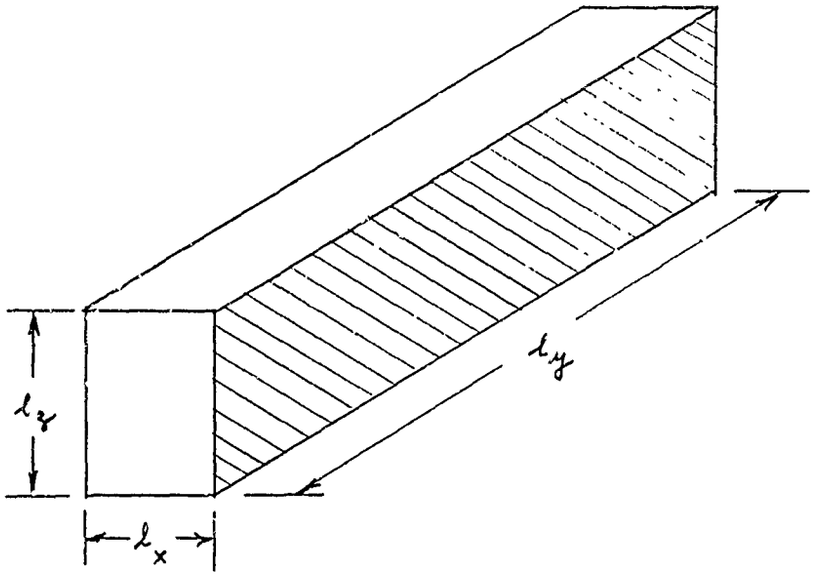
the circular piston and the sphere. Although only relatively few types of radiating surface have been analyzed, owing to mathematical difficulties, it is generally true that when the dimensions of the surface are large compared with the wavelength, the impedance is predominantly resistive with $r_1 \approx 1$, whereas when the dimensions are small compared with a wavelength, the impedance is largely reactive. The reactance is inductive, indicating that an inertia effect is present, an out-of-phase component in which mass is being moved without a corresponding generation of radiation.

4. Piezoelectric Transducers

The following discussion will present a somewhat over-simplified theory of one type of quartz crystal transducer, the purpose being merely to present some of the general concepts of piezoelectric transducers. In general, when a piezoelectric crystal is subjected to an electric field it experiences both compressional and shear strains whose values depend upon the way in which the crystal is cut and the direction in which the field is applied. We shall consider only the compressional strains.

a. The Piezoelectric Effect in a Quartz Crystal

We assume the crystal to be cut in rectangular form with sides of lengths ℓ_x, ℓ_y, and ℓ_z, as shown in the sketch



The two faces normal to the x axis are plated with conducting material (indicated by the shading). When a voltage is applied across these faces, the crystal suffers a change in length in the y-direction. Conversely, if a force is applied in the y-direction, electric charges of opposite polarity are produced on the plated faces in the x-direction. The behavior of the crystal in the presence of both an electric potential in the x-direction and a force in the y-direction is expressed approximately by the following two equations:

$$\sigma = \frac{\epsilon \epsilon_0 E}{\ell_x} - \frac{d_{12} F}{S} \tag{614}$$

$$\frac{\partial \eta}{\partial y} = -\frac{F}{Y_y S} + \frac{d_{12} E}{\ell_x} \tag{615}$$

where

σ = charge density (charge per unit area) on the plated faces

ϵ = "free" dielectric constant of the crystal in the x-direction, i.e., the dielectric constant measured when no external force is applied

ϵ_0 = 8.85×10^{-12} farad/meter = permittivity of free space

E = potential difference in volts, applied in the x-direction

F = applied compressional force in the y-direction

S = $\ell_x \ell_z$ = area of face normal to y-direction

d_{12} = one of the piezoelectric strain coefficients of the quartz crystal, measured in meters/volt or cm/volt, depending on the units used.

η = change in length of the y-dimension ℓ_y, due to the electrical and mechanical stresses applied

$\frac{\partial \eta}{\partial y}$ = longitudinal strain, i.e., fractional change in length

Y_y = Young's modulus of elasticity, measured in the y-direction.

Before we proceed with the analysis, a few words of explanation of equations (614) and (615) may be helpful. Assume for the moment that no external force is applied, i.e., $F = 0$. In this case equation (614) is the classical relation between the applied potential and the charge on the face of the capacitor. The charge is equal to the product of the charge density and the area of the face.

$$Q = \sigma S = \sigma \ell_y \ell_z$$

and the capacitance is the ratio of the charge to the potential difference, i.e.,

$$C = \frac{Q}{E} = \frac{\epsilon \epsilon_o S}{\ell_x}$$

This is the standard formula for the capacitance of a parallel plate capacitor.

If now a force F in the y-direction is applied instead of the voltage, a charge appears on the plates, whose magnitude depends upon the value of the piezoelectric strain coefficient d_{12}. This is called the direct piezoelectric effect. If both the voltage and the force are applied, the resultant charge density is given by (614).

Considering now (615), we see that the first term is simply the classical expression for Hooke's Law in the form

$$\text{Strain} = \frac{\text{Stress}}{\text{Modulus}}$$

The second term states that when an electric voltage is applied in the x-direction, it generates a strain in the y-direction whose value depends on the strain coefficient d_{12}. This is called the inverse piezoelectric effect.

In summary, the ordinary elastic properties of the quartz crystal are represented by the first term of (615). When two opposite faces are plated, the crystal acts as an ordinary parallel plate capacitor, as indicated by the first term of (614). In addition to the above, the piezoelectric

properties of the crystal are expressed by the second terms of both (614) and (615).

It will be convenient to express σ in terms of the voltage and strain by eliminating F between (614) and (615). The result is

$$\sigma = \frac{\epsilon' \epsilon_o E}{\ell_x} + Y_y d_{12} \frac{\partial \eta}{\partial y} \tag{616}$$

where

$$\epsilon' = \epsilon - \frac{Y_y d_{12}^2}{\epsilon_o} \tag{617}$$

The quantity ϵ' is called the clamped dielectric constant. It is the value which would be measured if the crystal could be clamped in such a way as to keep ℓ_y constant, i.e., $\partial\eta/\partial y = 0$.

b. Quartz Crystal as a Quarter-Wave Oscillator

To illustrate the behavior of a piezoelectric transducer, let us assume that one end of the crystal - one of the faces normal to the y-axis - is rigidly clamped to a heavy mass, so that it cannot move, and that the other end is in contact with the water (either directly or through some intermediate fluid having the same specific acoustic impedance). We shall also assume that the dimensions of the radiating surface are sufficiently large relative to the wavelength that the radiation impedance is purely resistive and has the value ρcS, where ρ and c are the density and sound speed of the water, and S is the area of the surface.

If an alternating voltage Ee^{jwt} is applied across the plated surfaces (normal to the x-axis), it will set up a mechanical oscillation in the y-direction, which can be expressed as a standing wave consisting of one wave traveling in the positive y-direction and a second wave traveling in the negative y-direction. The displacement η of the crystal material can then be expressed mathematically as a function of the y coordinate and of time by the following equation

$$\eta = (Ae^{-jk_y y} + Be^{jk_y y})\, e^{jwt} \tag{618}$$

where A and B are constant coefficients to be evaluated, and k_y is the wave number of the wave in the material

$$k_y = \frac{2\pi}{\lambda_y} = \frac{w}{c_y} \tag{619}$$

The sound speed c_y along the y-direction in the crystal is related to the density ρ_m of the material and the modulus of elasticity Y_y in the same manner as the sound speed in water (equation (5)), namely

$$c_y = \sqrt{\frac{Y_y}{\rho_m}} \tag{620}$$

In accordance with our assumptions, the boundary conditions for the evaluation of A and B are:

At $y = 0$ (clamped end): $\eta = 0$

At $y = \ell_y$ (radiating end): Mechanical Impedance = ρcS

Substitution of the first condition into (618) yields

$$A + B = 0$$

so that

$$\eta = -A \left(e^{jk_y y} - e^{-jk_y y}\right) e^{jwt}$$

or

$$\eta = -2jA \sin k_y y \; e^{jwt} \tag{621}$$

If we denote by Fe^{jwt} the force exerted by the water on the radiating face (which, of course, is equal and opposite to the force exerted by the crystal on the water), the second condition yields

$$Fe^{jwt} = \rho cS \left.\frac{\partial \eta}{\partial t}\right|_{y = \ell_y}$$

$$= \rho cS \left(-2j A \sin k_y \ell_y\right) j w e^{jwt}$$

so that A is found to be

$$A = \frac{F}{2 w \rho cS \sin k_y \ell_y}$$

and the displacement is

$$\eta = -\frac{j F \sin k_y y}{w\rho c S \sin k_y \ell_y} e^{jwt} \qquad (623)$$

The force F is related to the applied voltage E by the piezoelectric equation (615), evaluated at $y = \ell_y$. The strain is

$$\left.\frac{\partial \eta}{\partial y}\right|_{y=\ell_y} = -\frac{j k_y F \cos k_y \ell_y}{w\rho c S \sin k_y \ell_y} e^{jwt}$$

The force and applied voltage are, as previously stated, Fe^{jwt} and Ee^{jwt} Substitution into (615) and cancellation of the factor e^{jwt} yields

$$-\frac{j k_y F \cos k_y \ell_y}{w\rho c S \sin k_y \ell_y} = -\frac{F}{Y_y S} + \frac{d_{12} E}{\ell_x}$$

or

$$\frac{F}{Y_y S}\left(1 - j \frac{k_y Y_y \cos k_y \ell_y}{w\rho c \sin k_y \ell_y}\right) = \frac{d_{12} E}{\ell_x}$$

But from (619),

$$\frac{k_y}{w} = \frac{1}{c_y}$$

and from (620),

$$Y_y = \rho_m c_y^2$$

Therefore F solves to

$$F = \frac{\frac{d_{12} Y_y S}{\ell_x} E}{1 - j \frac{\rho_m c_y}{\rho c} \cot(k_y \ell_y)}$$

The factor $\frac{d_{12} Y_y S}{\ell_x}$, which we shall denote by the symbol ϕ, is called the transformation factor. It is measured in dynes/volt in the cgs system and in newtons/volt in the MKS system. Thus,

$$\phi = \frac{d_{12} Y_y S}{\ell_x} = d_{12} Y_y \ell_z \qquad (624)$$

and

$$F = \frac{\phi E}{1 - j \frac{\rho_m c_y}{\rho c} \cot k_y \ell_y} \qquad (625)$$

Substitution of (625) into (623) yields for the displacement η,

$$\eta = - \frac{j \phi E \sin k_y y \; e^{jwt}}{w \sin k_y \ell_y (\rho cS - j \rho_m c_y S \cot k_y \ell_y)} \qquad (626)$$

The velocity of the radiating surface (which is equal to the particle velocity of the water adjacent to the surface) is

$$u = \left.\frac{\partial \eta}{\partial t}\right|_{y=\ell_y} = \frac{\phi E \; e^{jwt}}{\rho cS - j \rho_m c_y S \cot k_y \ell_y}$$

If U denotes the complex amplitude of u, so that

$$u = U e^{jwt} \qquad (627)$$

we obtain the relation

$$U = \frac{\phi E}{\rho cS - j \rho_m c_y S \cot k_y \ell_y} \qquad (628)$$

Here we see that the numerator ϕE has the dimensions of a force, and that this applied force generates a velocity U of the transducer surface. The denominator of (628) thus represents the effective mechanical impedance of the system.

$$\bar{z}_m = \rho cS - j \rho_m c_y S \cot k_y \ell_y \qquad (629)$$

The form of equation (628) is the same as that of a series resonant electrical circuit. The "mechanical circuit" consists of a mechanical resistance ρcS in series with a mechanical reactance $- \rho_m c_y S \cot k_y \ell_y$. The absolute value of the current amplitude is

$$|U| = \frac{\phi E}{[(\rho cS)^2 + (\rho_m c_y S \cot k_y \ell_y)^2]^{1/2}} \qquad (630)$$

Recalling that the wave number k_y is a function of the frequency, that is,

$$k_y = \frac{2\pi}{\lambda_y} = \frac{w}{c_y} = \frac{2\pi f}{c_y}$$

we see that the reactance term containing $\cot k_y \ell_y$ will vary as the frequency is varied. Clearly the crystal will resonate at those frequencies for which $\cot k_y \ell_y = 0$, that is, for which

$$k_y \ell_y = \frac{2n-1}{2} \pi \;, \quad n = 1, 2, 3, \ldots \qquad (631)$$

The fundamental mode occurs for n = 1. In this case

$$k_y \ell_y = \frac{2\pi \ell_y}{\lambda_y} = \frac{\pi}{2} \tag{631}$$

or

$$\ell_y = \frac{\lambda_y}{4} \tag{632}$$

Thus, a crystal which is rigidly clamped at one end will resonate when its length is one quarter of the wavelength of the longitudinal waves in the crystal material.

Let us now investigate the behavior of the impedance as a function of the frequency. Consider first the case of very low frequencies, far below resonance, where $k_y \ell_y$ is very small. In this case

$$\cos k_y \ell_y \approx 1, \quad \sin k_y \ell_y \approx k_y \ell_y, \quad \text{and} \quad \cot k_y \ell_y \approx \frac{1}{k_y \ell_y}$$

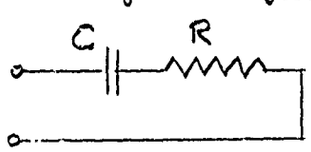
and the mechanical impedance (629) becomes approximately

$$\bar{z}_m \approx \rho c S - j \frac{\rho_m c_y S}{k_y \ell_y}$$

which may be transformed with the aid of (619) and (620) to the following

$$z_m \approx \rho c S + \frac{Y_y S}{j w \ell_y} \tag{633}$$

This is an impedance consisting of a series combination of resistance and capacitance, that is, of the form $\bar{z} = R + \frac{1}{j\omega C}$



The resistance term is the radiation resistance $\rho c S$. The factor $Y_y S/\ell_y$ in the reactance term is the mechanical analog of the reciprocal of the capacitance and, as we have seen previously, should correspond to the stiffness coefficient of a spring. As a matter of fact, it does represent the stiffness coefficient of the crystal in the y-direction. If a static compressional force F is applied in this direction across the crystal, it will cause the length ℓ_y to decrease by an amount η such that, according to Hooke's Law,

$$\frac{F}{S} = -\, Y_y \frac{\eta}{\ell_y}$$

or

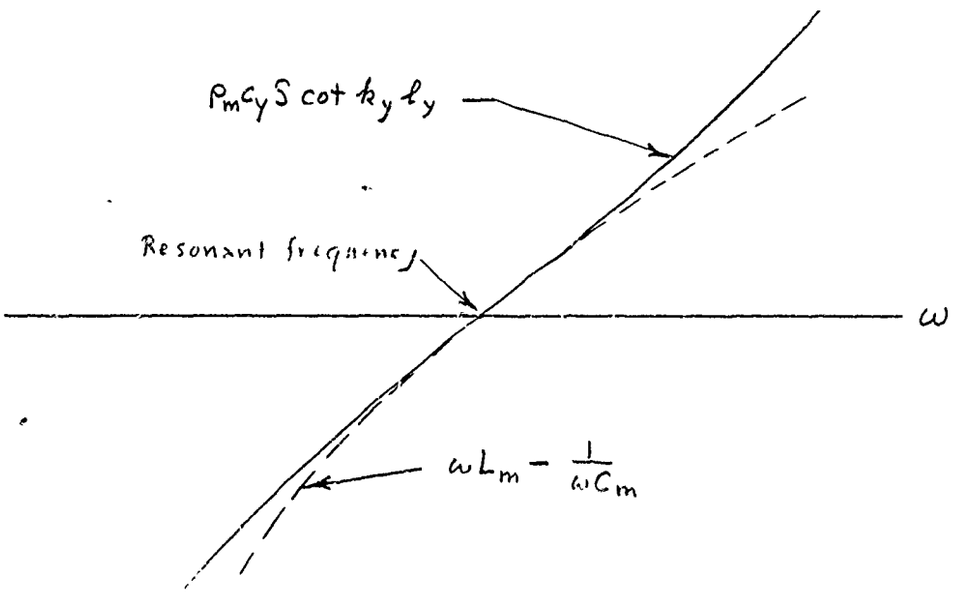
$$F = -\frac{Y_y S}{\ell_y}\,\eta \qquad (634)$$

Thus at low frequencies the crystal behaves essentially as a damped massless spring, the damping being caused by the acoustic radiation load.

Consider now the behavior of the crystal in the vicinity of the fundamental resonant frequency, that is, where $k_y \ell_y$ is approximately $\pi/2$ radians. According to our basic assumption the resistance term of the mechanical impedance is independent of frequency and retains its value of $\rho c S$. The reactance term, being proportional to cot $k_y \ell_y$, is zero at $k_y \ell_y = \frac{\pi}{2}$. Because of the minus sign in front, it is negative at lower frequencies and positive at higher. This is similar to the behavior of the reactance of a series combination of inductance and capacitance,

$$X_m = wL_m - \frac{1}{wC_m} \qquad (635)$$

Plots of the cotangent and the series LC reactance are shown in the diagram below. The best fit in the vicinity of the resonance frequency is obtained by



matching the slopes of the two curves at that point. The derivative of the reactive component of (629) with respect to w is

$$\frac{dXm}{dw} = - \rho_m c_y S \frac{d}{dw}\left(\cot k_y \ell_y\right)$$

Upon substitution of (619), the derivative is seen to be

$$\frac{dXm}{dw} = \frac{\rho_m \ell_y S}{\sin^2 k_y \ell_y}$$

At the resonant frequency, w_o, where $k_y \ell_y = \frac{\pi}{2}$, this becomes

$$\left.\frac{dXm}{dw}\right|_{w_o} = \rho_m \ell_y S \qquad (636)$$

The derivative of (635) is

$$\frac{dXm}{dw} = L_m + \frac{1}{w^2 C_m} \qquad (637)$$

Resonance occurs at the frequency at which the inductive reactance and capacitative reactance are equal. The resonant frequency is

$$w_o = \frac{1}{\sqrt{L_m C_m}} \qquad (638)$$

Substitution of (638) into (637) to eliminate C_m yields

$$\frac{dXm}{dw} = L_m\left(1 + \frac{w_o^2}{w^2}\right) \qquad (637a)$$

which, at the resonant frequency ($w = w_o$), is

$$\left.\frac{dXm}{dw}\right|_{w_o} = 2 L_m \qquad (639)$$

If the slopes of the two curves are to be equal, (636) and (639) must be equal. Hence

$$L_m = \tfrac{1}{2} \rho_m \ell_y S = m \qquad (640)$$

From our previous discussion of mechanical inductance, we should expect L_m to be a mass, as indeed it is. The product $\ell_y S$ is the volume of the crystal, and $\rho_m \ell_y S$ is its mass. Thus L_m is the <u>effective</u> mass, m, of the crystal, which is equal to one-half the actual mass.

The reciprocal of the mechanical capacitance, which represents the effective stiffness of the crystal, is obtained from (638)

$$\frac{1}{C_m} = w_o^2 L_m = \tfrac{1}{2} w_o^2 \rho_m \ell_y S$$

This expression may be transformed by application of (619) and (620). Since $k_y \ell_y = \frac{\pi}{2}$ at the resonant frequency, we see that

$$w_o^2 = k_y^2 c_y^2 = \frac{\pi^2 c_y^2}{4 \ell_y^2} = \frac{\pi^2 Y_y}{4 \rho_m \ell_y^2}$$

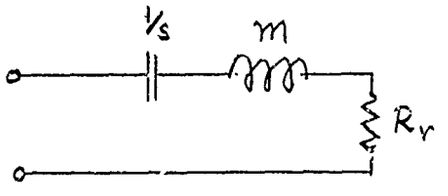
Hence

$$\frac{1}{C_m} = \frac{\pi^2 Y_y S}{8 \ell_y} = s \qquad (641)$$

where s denotes the effective stiffness. It will be noted that the effective stiffness at resonance is slightly larger, by a factor of $\frac{\pi^2}{8}$, than the static stiffness.

We see that in the vicinity of the fundamental resonance frequency the mechanical impedance has the characteristics analogous to a series resonant electrical circuit, the equation being

$$z_m = R_r + j\left(wm - \frac{s}{w}\right) \qquad (642)$$



where

$R_r = \rho c S$ = radiation resistance (643)

$m = \tfrac{1}{2} \rho_m \ell_y S$ = effective mass (640)

$s = \frac{\pi^2 Y_y S}{8 \ell_y}$ = effective stiffness (641)

At frequencies below resonance the stiffness term in the reactance is dominant and the velocity of the radiating surface leads the applied voltage in phase. At frequencies above resonance the mass term is dominant and the velocity lags

the voltage. At resonance the two terms cancel and the velocity is in phase with the voltage, the relation (628) being in this case

$$U = \frac{\phi E}{\rho c S} \quad \text{(at resonance)} \qquad (628a)$$

A further comment is in order regarding the significance of the transformation factor ϕ which appears in equation (628). The oscillations of the crystal are actually driven by the applied voltage Ee^{jwt} which acts through the piezoelectric properties of the crystal to produce the velocity Ue^{jwt}. Equation (628) states that the effect of the electric driving potential is equivalent to a mechanical driving force ϕEe^{jwt}. The piezoelectric transducer may thus be visualized as a sort of "transformer" between the electrical and mechanical systems, the "turns ratio" (secondary primary) being the reciprocal of the transformation factor.

In summary, our analysis has shown that when a quartz crystal is rigidly clamped at one end and is driven by an alternating voltage. it acts as a quarter-wave resonator, the mechanical system being analogous to a series-resonant A.C. electrical circuit. The analysis has been somewhat oversimplified, neglecting all secondary effects and making the assumption of an ideal infinite mass backing. The purpose has been to emphasize the basic concepts involved, rather than to discuss the details of hardware design. Also, this is only one of several types of transducer configurations. For example, instead of backing the crystal with a heavy mass, it is possible to design an efficient radiator by allowing the back end to vibrate freely in (ideally) a vacuum. A bar which vibrates at both ends will have a node in the center (or in the vicinity of the center if one end is loaded), and this node is equivalent to backing the front portion with an infinite mass. It is impossible, of course, to terminate the free end in a vacuum, but air is practically equivalent to a vacuum because of its low specific acoustic impedance relative to that of water. It is beyond the scope of these notes, however, to analyze other designs, since they differ more in engineering details than in basic principles.

c. Equivalent Circuit of Quarter-Wave Quartz Oscillator

To obtain the equivalent circuit of the piezoelectric oscillator we must compute the current

$$i = Ie^{jwt}$$

which flows when the voltage Ee^{jwt} is applied across the plated surfaces normal to the x-axis. The impedance of the equivalent circuit is then

$$\overline{z} = \frac{E}{I} \tag{644}$$

or, expressed in another way, the admittance of the equivalent circuit is

$$\overline{y} = \frac{1}{\overline{z}} = \frac{I}{E} \tag{644a}$$

By examining the mathematical form of $\overline{z}$ (or $\overline{y}$), we shall be able to recognize the combination of circuit elements which comprise the equivalent circuit.

In this analysis we shall consider only the crystal itself, as though the power source were connected directly across the plates. A complete transducer will of course have other electrical elements coupling the power source to the oscillator, and the complete equivalent circuit will include all of these components in addition to the equivalent circuit of the oscillator itself. A convenient starting point for the analysis is equation (616), which relates the charge density appearing on the plates to the applied voltage and to the strain $\frac{\partial \eta}{\partial y}$. The charge density is the charge per unit area, so that to obtain the total charge we must integrate σ over the face. Thus

$$q = \int_0^{\ell_y} \int_0^{\ell_z} \sigma \, dy \, dz \tag{645}$$

The current is the time derivative of the charge

$$i = \frac{dq}{dt} \tag{646}$$

The strain is found by taking the derivative of (626) with respect to y. When this is done, the charge density is found to be

$$\sigma = \left[\frac{\epsilon'\epsilon_o}{l_x} - j\,\frac{Y_y d_{12} k_y \phi \cos k_y y}{\omega \sin k_y l_y (pcS - j\rho_m c_y S \cot k_y l_y)}\right] E e^{j\omega t}$$

The first term is a constant and the second is a function only of y. The charge is therefore

$$q = \left[\frac{\epsilon'\epsilon_o l_y l_z}{l_x} - j\,\frac{Y_y d_{12} l_z k_y \phi}{\omega \sin k_y l_y (pcS - j\rho_m c_y S \cot k_y l_y)} \int_0^{l_y} \cos k_y y\, dy\right] E e^{j\omega t}$$

Integration plus substitution of (624) yields

$$q = \left[\frac{\epsilon'\epsilon_o l_y l_z}{l_x} - j\,\frac{\phi^2}{\omega(pcS - j\rho_m c_y^S \cot k_y l_y)}\right] E e^{j\omega t} \qquad (647)$$

and the current is

$$i = \underline{I}\, e^{j\omega t} = \left[j\omega\frac{\epsilon'\epsilon_o l_y l_z}{l_x} + \frac{1}{\frac{pcS}{\phi^2} - j\,\frac{\rho_m c_y^S \cot k_y l_y}{\phi^2}}\right] E e^{j\omega t} \qquad (648)$$

Here we see that the admittance of the equivalent circuit is of the form

$$\bar{y} = j\omega C_o + \frac{1}{R_R + j X_M} \qquad (649)$$

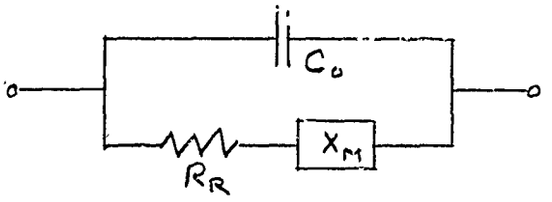
where

$$C_o = \frac{\epsilon' \epsilon_o l_y l_z}{l_x} \qquad (650)$$

$$R_R = \frac{pcS}{\phi^2} \qquad (651)$$

$$X_M = -\frac{\rho_m c_y^S S \cot k_y l_y}{\phi^2} \qquad (652)$$

This is the admittance of a parallel circuit having a capacitance C_o in one branch and a series combination of a resistance R and an inductance X in the other. The formula for C_o is similar to what we observed earlier for the



capacitance of the quartz crystal, the only difference being that it contains the clamped dielectric constant ϵ' instead of the free dielectric constant ϵ. C_o is therefore the capacitance which would be measured if the crystal were rigidly clamped so that its dimensions could not be changed. It will be noted that this is a real electrical property of the crystal, as contrasted with R_R and X_M, which represent the effect on the electrical circuit of mechanical properties.

Comparison of the resistance R_R (651) and the reactance X_M (652) with the corresponding components of the mechanical impedance (629), that is,

$$\bar{z}_m = R_r + j X_m \tag{629a}$$

where

$$R_r = \rho c S \text{ (radiation resistance)} \tag{643}$$

$$X_m = \rho m \, cy \, S \cot k_y \ell_y \tag{653}$$

shows that

$$R_R = \frac{R_r}{\phi^2} \tag{654}$$

$$X_M = \frac{X_m}{\phi^2} \tag{655}$$

(The small letter m denotes the mechanical component; the large letter M denotes the electrical equivalent; and the subscripts r and R refer to the radiation resistance.) Furthermore, we have seen that at frequencies in the vicinity of the fundamental resonance point the mechanical reactance can be expressed approximately as the sum of reactive components

$$X_m = wm - \frac{s}{w} \tag{656}$$

where $m = \frac{1}{2}\rho_m \ell_y S = \underset{\wedge}{\text{effective}}$ mass of crystal (640)

and $s = \frac{\pi^2 Y_y S}{8 \ell_y}$ = dynamic stiffness coefficient (641)

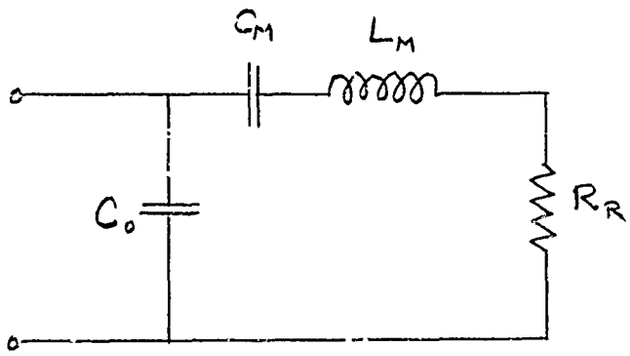
The electrical reactance of the equivalent circuit likewise consists of an inductance L_M and a capacitance C_M in series, such that

$$L_M = \frac{m}{\phi^2} \tag{657}$$

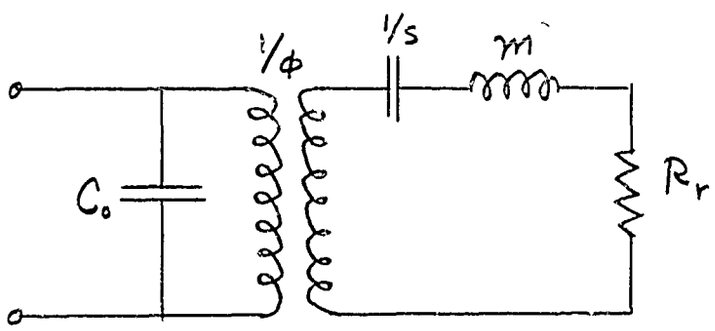
and

$$\frac{1}{C_M} = \frac{s}{\phi^2} \tag{658}$$

The electrical equivalent circuit for the vibrator operating near resonance is as shown in the following diagram.

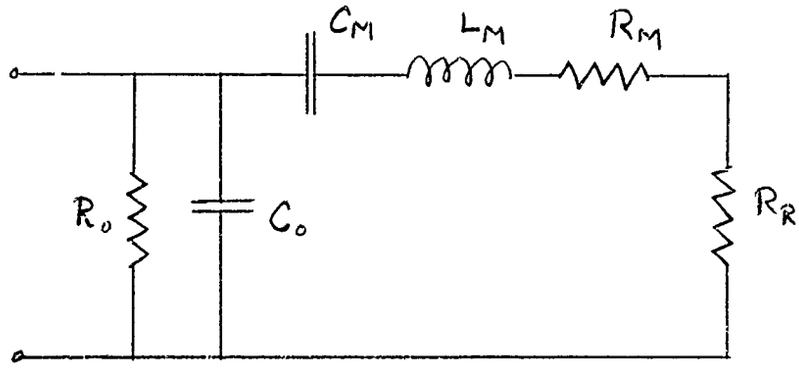


The same information may be presented in another form, using the previously mentioned concept of a hypothetical transformer with a "turns ratio" $1/\phi$.



It must be stressed that the concept of the transformer is symbolic in that th turns ratio is not a pure ratio but rather a dimensional factor relating mechanical to electrical quantities.

Two factors, present in a real crystal but neglected in the preceding discussion, are friction losses in the crystal vibration and dielectric losses in the crystal as a capacitor. If these are included, the equivalent circuit looks as follows.



The leakage resistance R_o is shunted across the crystal capacitance C_o, and the friction loss resistance R_M is in series with the radiation load resistance R_R.

Another practical problem encountered in the real world is the fact that the dimensions of a single crystal are seldom large enough relative to a wavelength to provide a pure resistive load. As we have seen, the radiation load of a small radiating surface has a large reactive component. To overcome this difficulty, a number of vibrators are mounted side-by-side, forming an extended array, and are driven in phase. (The size of the array is also important from a consideration of transducer directivity, as we shall see later.)

d. Coefficient of Mechanical Coupling

The ratio of the capacitance C_M in the motional branch of the equivalent circuit to the ordinary capacitance C_o of the crystal is a measure of the amount of coupling between the electrical and mechanical systems of the transducer. Obviously, any material for which this ratio is large will have a strong piezoelectric effect. From (650), (641), and (658) this ratio is found to be

$$\frac{C_M}{C_o} = \frac{8\,\phi^2\,\ell_y}{\pi^2\,Y_y\,S} \cdot \frac{\ell_x}{\epsilon'\epsilon_o\,\ell_y\,\ell_z}$$

or, since

$$S = \ell_x\,\ell_z$$

$$\phi = d_{12}\,Y_y\,\ell_z \qquad (624)$$

and

$$\epsilon' = \epsilon\,\left(1 - \frac{Y_y\,d_{12}^2}{\epsilon\,\epsilon_o}\right) \qquad (617)$$

the ratio is

$$\frac{C_M}{C_o} = \frac{8}{\pi^2}\,\frac{\frac{d_{12}^2\,Y_y}{\epsilon\,\epsilon_o}}{1 - \frac{d_{12}^2\,Y_y}{\epsilon\,\epsilon_o}} \qquad (659)$$

The quantity $\frac{d_{12}^2\,Y_y}{\epsilon\,\epsilon_o}$ is independent of the dimensions of the crystal; it is characteristic of the nature of the crystal material. It is a dimensionless quantity, called the coefficient of electromechanical coupling. We shall denote it by the symbol k^2.

$$k = d_{12}\sqrt{\frac{Y_y}{\epsilon\,\epsilon_o}} \qquad (660)$$

The capacitance ratio is thus

$$\frac{C_M}{C_o} = \frac{8k^2}{\pi^2\,(1 - k^2)} \qquad (659a)$$

The value of k for quartz is 0.1 and for barium titanate is 0.18.

e. The Quality Factor Q

The quality factor Q is a dimensionless ratio which is commonly used as a measure of the sharpness of resonance of an oscillating system. For a series resonant electrical system at its resonant frequency, Q is defined as

$$Q = 2\pi \frac{\text{Energy Stored in Inductance L}}{\text{Energy Dissipated in Resistance R per cycle}} \qquad (661)$$

In a resonant A.C. circuit the energy oscillates between the inductor and the capacitor. When the current is a maximum, the energy is stored in the magnetic field of the inductor. When the current drops to zero, the maximum charge appears across the plates of the capacitor, and the energy is stored in the electric field of the capacitor. Under steady state conditions these two amounts of energy are equal. Furthermore, the smaller the loss of energy in the resistance, the sharper will be the resonance. It is readily shown (see equation (680)) that for a series RLC circuit, (661) may be expressed in either of the following two forms

$$Q = \frac{w_o L}{R} = \frac{1}{w_o CR} \qquad (662)$$

In a mechanical system the kinetic energy of the mass corresponds to the magnetic energy of the inductor, and the potential energy of the elastic material corresponds to the electrostatic energy of the capacitor.

In describing sonar transducers it is customary to define two Q's, the mechanical Q_M and the electrical Q_E. Q_M is the Q of the mechanical system, and its value for the piezoelectric oscillator above is

$$Q_M = \frac{w_o L_M}{R_R + R_M} = \frac{w_o m}{R_r + R_m} \qquad (663)$$

where R_r and R_R are the radiation resistance (mechanical and electrical equivalent) and R_m and R_M represent the loss in the crystal. Although a system with a large Q_M is an efficient oscillator at its resonant frequency, it has a narrow bandwidth. In fact, it is readily shown that for a series

resonant circuit the bandwidth Δw between the 3 db-down points on either side of the resonant frequency is

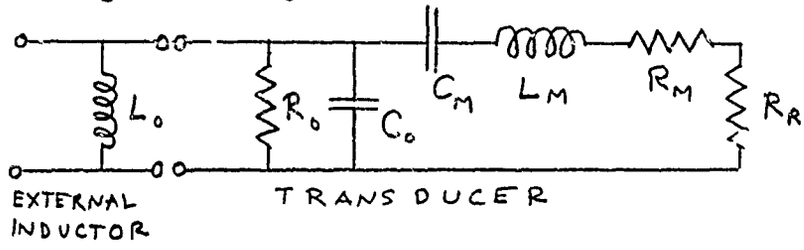
$$\Delta w = \frac{w_o}{Q}$$

In operational sonars, even for single-frequency pulses, there is a requirement for a significantly wide bandwidth to accommodate doppler shifts of the echo. Also, because of signal processing considerations, many modern sonars transmit pulses having an appreciable bandwidth. Reasonable values for the Q_M of a sonar projector are in the range from 5 to perhaps 10.

In the equivalent circuit of a piezoelectric transducer operating in water, the shunting capacitance C_o has a far greater admittance than the motional (mechanical) branch. As the frequency is varied from below to above resonance, the mechanical branch produces only a relatively small wiggle in the overall admittance curve. For this reason the electrical Q_E cannot be interpreted in the ordinary context of sharpness of resonance. At the resonant frequency the equivalent circuit consists of the capacitance C_o in parallel with the radiation and frictional resistance $R_R + R_M$. The Q_E of this lossy capacitor is to be interpreted in the context of energy storage, in the manner of (661). For this parallel circuit it is the reciprocal of what it would be for a series circuit, since the energy loss in a parallel circuit is inversely proportional to the shunting resistance. Therefore, at the resonant frequency,

$$Q_E = w_o C_o (R_R + R_M) \qquad (664)$$

It will be noted that since the equivalent circuit is highly reactive due to the large shunting capacitor, it therefore has a low power factor. The power factor can be improved by tuning the circuit by means of an external shunting inductor L_o:



f. Impedance Measurements

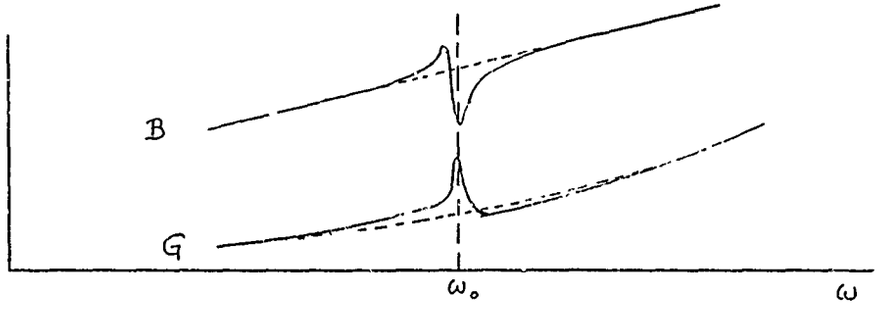
Examination of the equivalent circuit of a piezoelectric transducer shows that if we were able to open up the mechanical branch, the admittance of the circuit would consist only of the electrical portion. How does one "open up" the mechanical branch of the equivalent circuit? The analog of the electric current is the velocity of the radiating surface. Therefore to achieve zero current one must prevent the transducer from moving. Looked at from another point of view, if the medium into which the acoustic waves radiate had an infinite specific acoustic impedance, then R_R would be infinite. These remarks are consistent with our earlier finding that C is the clamped (or blocked) capacitance of the crystal, since it involves the clamped dielectric constant ϵ'.

In order to evaluate the components of the mechanical branch it would be logical to measure the admittance of the transducer under normal operating conditions as a function of the frequency, then measure the blocked admittance at the same frequencies. Since the blocked admittance is the admittance of the electrical branch, it follows that by subtracting the blocked admittance from the normally loaded admittance, we shall obtain the motional admittance, i.e., the admittance of the mechanical branch alone.

Note that in measuring admittance we must measure two quantities, the real (in-phase) component--the conductance G--and the imaginary (out-of-phase) component--the susceptance B,--since

$$\bar{y} = G + j B \tag{665}$$

The question now is, how do we measure the blocked admittance? Clearly it is not a simple matter to clamp the crystal and keep it from moving. One way to evaluate the blocked conductance and susceptance is to estimate them from plots of the normally loaded values vs. frequency. In the vicinity of resonance the normally loaded conductance and susceptance curves show wiggles, somewhat as shown below. However, as the frequency departs from resonsnce on either side, the



transducer velocity tends toward zero, and the difference between the blocked and normally loaded values of the admittance components likewise tent toward zero. A reasonable estimate of the blocked admittance components can be obtained by fairing smooth curves through the graphs of the normally loaded values, as suggested by the dotted lines in the preceding diagram.

When the resulting values of the motional susceptance B_M are plotted against the corresponding values of the motional conductance G_M, a very interesting graph is obtained. To investigate this, let us begin with the equation for the motional admittance, which, for simplicity, we shall write as follows:

$$\bar{y}_M = \frac{1}{R_M' + jX_M} = \frac{R_M'}{Z_M^2} - j\frac{X_M}{Z_M^2} \tag{666}$$

where $$R_M' = R_R + R_M \tag{667}$$

$$Z_M^2 = R_M'^2 + X_M^2 \tag{668}$$

$$X_M = wL_M - \frac{1}{wC_M} \tag{669}$$

Hence $$G_M = \frac{R_M'}{Z_M^2} \tag{670}$$

and $$B_M = -\frac{X_M}{Z_M^2} \tag{671}$$

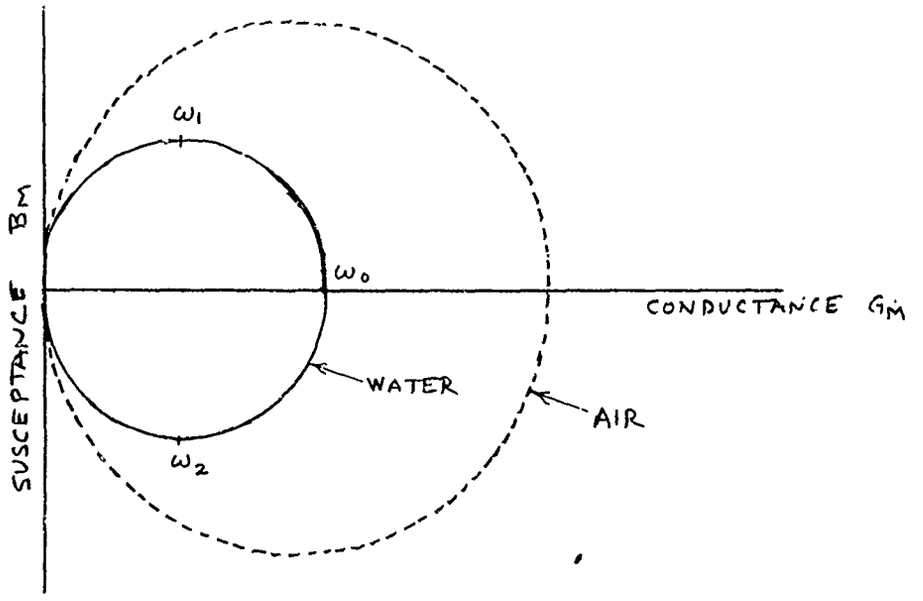
Squaring and adding (670) and (671) yields

$$G_M^2 + B_M^2 = \frac{R_M'^2 + X_M^2}{Z_M^4} = \frac{1}{Z_M^2} = \frac{G_M}{R_M'} \tag{672}$$

Equation (672) may be rewritten in the form

$$\left(G_M - \frac{1}{2R_M'} \right)^2 + B_M^2 = \left(\frac{1}{2R_M'}\right)^2 \tag{672a}$$

which is the equation of a circle of radius $1/2R_M'$ with its center on the G_M axis at $G_M = \frac{1}{2R_M'}$, as shown in the following sketch. This circle is called a <u>motional admittance circle</u>. It is the locus of points $[G_M(w), B_M(w)]$ as the (angular) frequency is varied from 0 to ∞. To see this, let us begin with a very low frequency, so small that $\frac{1}{wC_M}$ is the dominant term in both G_M and B_M.



$$\frac{1}{wC_M} \gg \begin{cases} R_M' \\ wL_M \end{cases}$$

At these frequencies, $Z_M^2 \approx \frac{1}{w^2 C_M^2}$ and

$$G_M \approx w^2 C_M^2 R_M' > 0$$

$$B_M \approx wC_M > 0$$

We thus see that the limit $w = 0$ corresponds to the origin, $G_M = B_M = 0$, and that as the frequency is increased, we move around the circle in a clockwise direction.

The resonant frequency w_0, where $X_M = 0$, corresponds to the point ($G_M = \frac{1}{R_M'}$, $B_M = 0$), which is on the G_M axis diametrically opposite the origin. At frequencies above resonance the reactance X_M is positive, and hence B_M is negative. This corresponds to the lower half of the circle. Finally, when $w \to \infty$, we return to the origin.

The purpose of obtaining values of G_M and B_M is to evaluate the components R_M, R_R, L_M, and C_M. We have already seen that R_M', which is the sum of R_R and R_M, is equal to the reciprocal of the circle diameter. In order to separate the radiation resistance R_R from the internal loss resistance R_M, we must remove the radiation load. Ideally this requires operation of the transducer in a vacuum, but air will suffice for practical purposes, since its specific acoustic impedance is less than 0.0003 times that of water. Thus, if G_0 represents the conductance measured at the resonant frequency, then

$$R_R + R_M = \frac{1}{(G_o)_{water}} \tag{673}$$

$$R_M = \frac{1}{(G_o)_{air}} \tag{674}$$

Typical admittance circles for operation in water and air are shown in the preceding sketch. The circle for air, of course, is considerably larger than the circle for water, since the conductance is larger.

The motional inductance and capacitance can be obtained from measurement of the two quadrantal frequencies, w_1 and w_2. These are the frequencies at the top and bottom of the admittance circle. At the top B_M is equal to G_M, or

$$\frac{1}{w_1 C_M} - w_1 L_M = R_M' \tag{675}$$

and at the bottom B_M is equal to the negative of G_M, or

$$\frac{1}{w_2 C_M} - w_2 L_M = -R_M' \tag{676}$$

Equations (675) and (676) may be solved simultaneously, yielding

$$L_M = \frac{R_M'}{w_2 - w_1} \tag{677}$$

$$C_M = \frac{w_2 - w_1}{w_1 w_2 R_M'} \tag{678}$$

Several additional interesting observations may be made from these relations. First, the resonant frequency w_o, which is

$$w_o = \frac{1}{\sqrt{L_M C_M}} \tag{679}$$

is equal to the geometric mean of the two quadrantal frequencies, as may be seen by substituting (677) and (678). Thus,

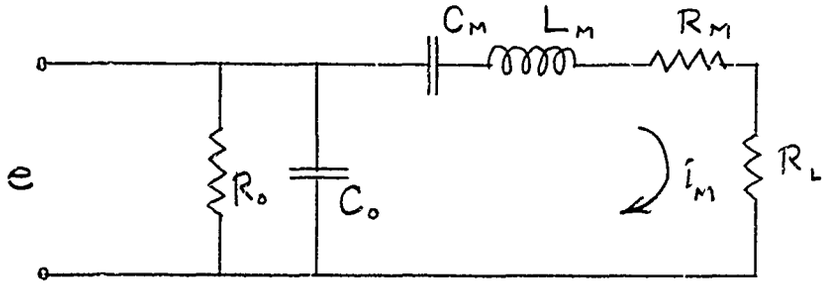
$$w_o = \sqrt{w_1 w_2} \tag{679a}$$

Secondly, the quadrantal frequencies occur at the 3 db-down points, since when $X_M = \pm R_M'$, the amplitude of the oscillation is $1/\sqrt{2}$ times its value at resonance. Third, substitution of (677) into (662) shows that the mechanical Q may be expressed in terms of the three frequencies as follows:

$$Q_M = \frac{w_o L_M}{R_M'} = \frac{w_o}{w_2 - w_1} \tag{680}$$

g. Efficiency.

The efficiency of the transducer is the ratio of the power expended in the load resistance R_R to the total power expended in the transducer. For this purpose we shall include both the friction loss resistance R_R and the dielectric loss resistance R_o, the latter being shunted across the capacitance C_o.



The efficiency is computed in two steps. The electromechanical efficiency η_{EM} is the ratio of the power in the motional branch to the total power in both branches. If an rms voltage e is applied across the terminals, the power lost in R_o is e^2/R_o. The current in the motional branch is

$$i_M = \frac{e}{Z_M}$$

where Z_M is the magnitude of the motional impedance

$$Z_M = (R_R + R_M)^2 + (wL_M - \frac{1}{wC_M})^2$$

and the power is

$$i_M^2(R_R + R_M) = \frac{e^2(R_R + R_M)}{Z_M^2}$$

Therefore

$$\eta_{EM} = \frac{\frac{e^2(R_R + R_M)}{Z_M^2}}{\frac{e^2}{R_o} + \frac{e^2(R_R + R_M)}{Z_M^2}}$$

$$= \frac{R_o(R_R + R_M)}{Z_M^2 + R_o(R_R + R_M)} \tag{681}$$

The second step is the mechanoacoustic efficiency η_{MA}, which is the fraction of the power in the motional branch which is transmitted as acoustic radiation. This is

$$\eta_{MA} = \frac{R_R}{R_R + R_M} \tag{682}$$

The overall efficiency, the electroacoustic efficiency η_{EA} is the product of the two,

$$\eta_{EA} = \frac{R_O R_R}{Z_M^2 + R_O(R_R + R_M)} \tag{683}$$

At the resonant frequency the reactance is zero and $Z_M = R_R + R_M$. In this case (683) simplifies to

$$\eta_{EA} = \frac{R_O R_R}{(R_R + R_M)(R_O + R_R + R_M)} \tag{684}$$

h. Piezoelectric Crystal as a Receiving Hydrophone.

When the crystal is used as a hydrophone the power input is in the form of a mechanical force generated by the acoustic waves. The output is in the form of a voltage at the electrical terminals. In analyzing the hydrophone circuit we shall make the customary simplifying assumption that the output terminals are open-circuited. The hydrophone may be analyzed either by direct application of the piezoelectric equations (615) and (616) to the mechanical system, or from the point of view of the equivalent circuit. We shall outline the first approach, omitting many of the detailed mathematical steps, and shall then show that the second leads to the same result.

Before we begin, it should be noted that if the output circuit is open, so that no current flows, no power will be drawn from the acoustic waves in the water, except for internal losses in the crystal itself, which we shall temporarily neglect. If no power is drawn, the force exerted on the crystal face by the water must be 90 degrees out of phase with the velocity. Since the force is equal to the product of the acoustic pressure and the area of the crystal face, and since the velocity of the surface is equal to the particle velocity of the water, this means that the pressure and particle velocity must be 90 degrees out of phase. In order for such a state of affairs to exist, the incoming wave must be reflected at the surface, and the relation between the incident and reflected waves must be such that the resultant pressure at the surface has the required character-

istics as determined by the transducer system.

The goal of the analysis is to obtain the relationship between the input acoustic pressure

$$p = P e^{jwt} = \frac{F e^{jwt}}{S} \tag{685}$$

and the output voltage $E e^{jwt}$. This relation can be obtained from (615), but to do so we must express the strain $\partial\eta/\partial y$ in terms of E or F. Assuming the same mechanical configuration as before, namely, that the crystal is clamped at one end, we see that equation (621) relating η to y and t, is applicable to the present mode of operation, the only difference being in the boundary conditions by which the coefficient A is determined. When operating as a hydrophone, the velocity of the crystal face must be equal to the particle velocity of the water, that is,

$$\dot{\eta} = u = Ue^{jwt} \qquad \text{when } y = l_y$$

The resulting value of A is

$$A = \frac{U}{2w \sin k_y l_y}$$

so that the strain is

$$\frac{\partial\eta}{\partial y} = -j \frac{U k_y \cos k_y v}{w \sin k_y l_y} e^{jwt} \tag{686}$$

We may now express U in terms of the output voltage E by application of (616). It will be noted that if there is no output current, the net charge on the plated faces of the crystal is zero, that is,

$$\int_0^{l_y}\int_0^{l_z} \sigma \, dy\, dz = 0$$

Since the electrical potential is constant over the face, the first term on the right integrates to

$$\int_0^{l_y}\int_0^{l_z} \frac{\epsilon\epsilon_0 E e^{jwt}}{l_x} dy\, dz = C_0 E e^{j\omega t}$$

where C_0 is the clamped capacitance of the crystal (650). Inserting the value of $\frac{\partial\eta}{\partial y}$ given by (686), and carrying out the integration, we obtain

$$wC_0E = j\phi U \tag{687}$$

where ϕ is the transformation factor defined by (624). Substitution of (687)

yields

$$\frac{\partial \eta}{\partial y} = -\frac{k_y C_0 E \cos k_y Y}{\phi \sin k_y l_y} e^{jwt} \tag{686a}$$

If now we insert this value of the strain into (615), we find at the boundary $y = l_y$,

$$E = \frac{F}{\left(\frac{d_{12}}{l_x} + \frac{k_y C_0}{\phi} \cot k_y l_y\right) Y_y S} \tag{688}$$

The above result appears rather complicated. However, in special cases it reduces to simpler forms. One case of great practical interest is the low-frequency case. If a hydrophone is to receive broadband signals, it should have a flat response over a wide range of frequencies. For this application it must operate well below resonance, that is, in the frequency range where $k_y l_y \ll 1$, and $\cot k_y l_y \approx \frac{1}{k_y l_y}$. In this case (688) becomes

$$E = \frac{F}{\left(\frac{d_{12}}{l_x} + \frac{C_0}{\phi l_y}\right) Y_y S} \tag{689}$$

which, upon substitution of (624), (650), and (617), reduces to

$$E = \frac{k^2 F}{\phi} = \frac{k^2 SP}{\phi} \tag{689a}$$

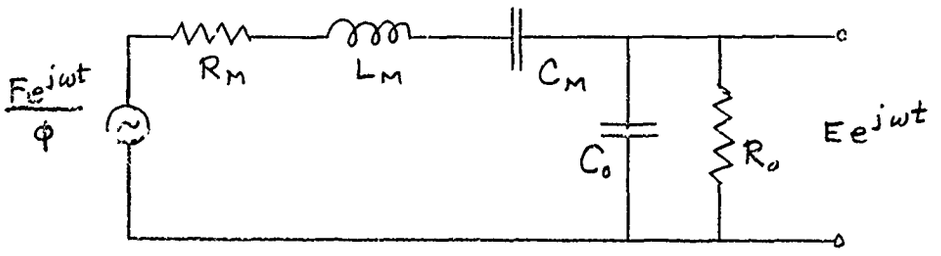
where k is the electromechanical coupling factor (660). In this expression F/ϕ is the electrical equivalent input voltage.

Another case of interest is operation at the resonant frequency. At resonance, $k_y l_y = \pi/2$ and $\cot k_y l_y = 0$, so that the output voltage is

$$E = \frac{F l_x}{d_{12} Y_y S} = \frac{F}{\phi} \tag{690}$$

Here the output voltage is equal to the electrical equivalent of F.

Turning now to the equivalent circuit, we see that the radiation load resistance R_R is removed from the circuit and an input voltage $\frac{Fe^{jwt}}{\phi}$ is inserted in its place. Since the transducer is open-circuited, there is no load at the other end. Including the internal friction loss resistance R_M and the dielectric loss resistance R_0, the circuit looks as follows:



This is a series circuit whose complex impedance $\bar{z}$ is the sum of the motional impedance $\bar{z}_M$ and the electrical impedance $\bar{z}_E$

$$\bar{z} = \bar{z}_M + \bar{z}_E \tag{691}$$

where

$$\bar{z}_M = R_M + j\left(wL_M - \frac{1}{wC_M}\right) \tag{692}$$

and

$$\bar{z}_E = \frac{1}{\frac{1}{R_o} + jwC_o} \tag{693}$$

The output voltage is the potential drop across $\bar{z}_E$. Hence

$$E = \frac{\bar{z}_E}{\bar{z}_M + \bar{z}_E}\frac{F}{\phi}$$

$$E = \frac{\dfrac{1}{\frac{1}{R_o} + jwC_o}}{R_M + j\left(wL_M - \frac{1}{wC_M}\right) + \dfrac{1}{\frac{1}{R_o} + jwC_o}}\frac{F}{\phi} \tag{694}$$

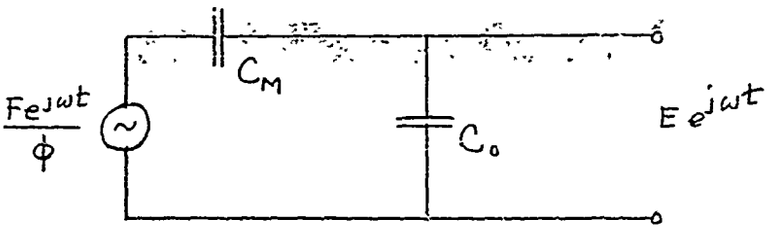
If the resistances R_M and R_o are neglected, as was done in the preceding analysis, the output voltage becomes

$$E = \frac{\frac{C_M}{C_o}}{1 + \frac{C_M}{C_o} - w^2L_MC_M}\frac{F}{\phi} \tag{695}$$

In the first of the two special cases considered above, i.e., operation at frequencies well below resonance, we found earlier that the motional reactance is a pure capacitance whose value (from (658) and (633)) is

$$C_M = \frac{\phi^2}{s} = \frac{\phi^2 l_y}{Y_y S} \tag{696}$$

In this case the equivalent circuit consists merely of two capacitances in series



and the output voltage simplifies to

$$E = \frac{C_M}{C_o + C_M}\frac{F}{\phi} \tag{697}$$

$$= \frac{F}{Y_y S\left(\frac{d_{12}}{l_x} + \frac{C_o}{\phi l_y}\right)} = \frac{k^2 F}{\phi}$$

which agrees with our previous result (689a).

In the other special case, where the hydrophone operates at the resonant frequency w_o, it is seen that the output voltage reduces to (690), which confirms our previous result. It should also be noted that whenever the resistances in the circuit have negligible effect, the circuit is reactive at all frequencies and the output voltage is in phase with the input pressure and 90 degrees out of phase with the particle velocity.

i. Numerical Data

The following table, adapted from Kinsler and Frey, Fundamentals of Acoustics, Second Edition, Wiley, 1962, lists the electromechanical constants of three common piezoelectric transducer materials.

Quantity	Quartz X-cut	ADP(20°C) 45° Z-cut	Barium Titanate (25°C)	Units MKS	Units CGS
Density, ρ	2.65	1.80	5.50	10^3 kg/m^3	1 gm/cm^3
Young's modulus, Y	7.9	2.9	1.11	10^{10}newt./m^2	10^{11}dyne/cm^2
Piezoelec. strain coef., d	2.3	24	56	10^{-12} m/volt	10^{-10}cm/volt
Free diel. const.,	4.5	15.3	12.0	-	-
Coupling coefficient, k	0.1	0.29	0.18	-	-
Longitudinal velocity, c	5.45	3.28	4.50	10^3 m/sec	10^5 cm/sec
Longit. impedance, ρc	14.5	5.9	6.2	10^6kg/m^{2}sec	10^5gm/cm^{2}sec

5. Magnetostrictive Transducers

We have gone into considerable detail in discussing piezoelectric transducers primarily for the purpose of illustrating some of the basic concepts of mechanical impedance and the interrelation between the electrical and mechanical systems. Since, except for details of application, these basic concepts apply also to magnetostrictive transducers, this section will be limited to a brief discussion of the magnetostrictive effect and its application to the vibrating element of a transducer. A complete discussion of transducer theory and design is beyond the scope of these notes.

When a longitudinal magnetic field is applied to a rod or tube of ferromagnetic material (parallel to the direction of the axis), the presence of the field will cause a change in physical dimensions, the most significant of which is a change in length. If the material was originally unmagnetized, the direction of the change, i.e. expansion or contraction, is independent of the direction of the field. That is, if the nature of the material is such that it expands, it will expand whether the field is directed toward one end or toward the other. In most such materials it is found that the strain (fractional change in length) produced is approximately proportional to the square of the magnetic flux density, B, although at very high flux densities there is a tendency toward saturation. Some materials such as permalloy expand, while others, such as nickel, contract. The most commonly used materials in magnetostrictive transducers are alloys of nickel.

The magnetic field is generated by a current in a coil wound around the tube. The magnetic field strength H produced in the material is proportional to the strength of the current i and to the number of turns per unit length of the coil (or to the total number of turns n divided by the length l of the coil).

$$H = \text{const} \cdot \frac{n\,i}{l}$$

The constant of proportionality depends upon the system of units employed. In the old CGS system H is measured in oersteds and l in centimeters, and if i is measured in amperes, the constant of proportionality is 0.4π. In the MKS system, which is replacing the CGS system in engineering applications, H is measured in ampere-turns/meter, i in amperes, and l in meters, and the constant of proportionality is unity. The relation between ampere-turns/meter and oersteds is

$$1 \text{ amp-turn/m} = \frac{4\pi}{10^3} \approx 0.0126 \text{ oersted}$$

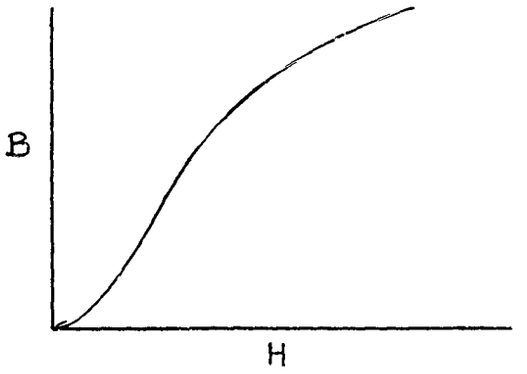
The relation between the magnetic flux density and the field strength depends upon the material and is usually expressed in the form

$$B = \mu' \mu_0 H = \mu H$$

where μ is the permeability of the material, μ' the relative permeability (relative to a vacuum), and μ_0 is a constant whose value depends upon the system of units used. It is called the permeability of free space. B is measured in gauss in the CGS system and in webers/meter^2 in the MKS system.

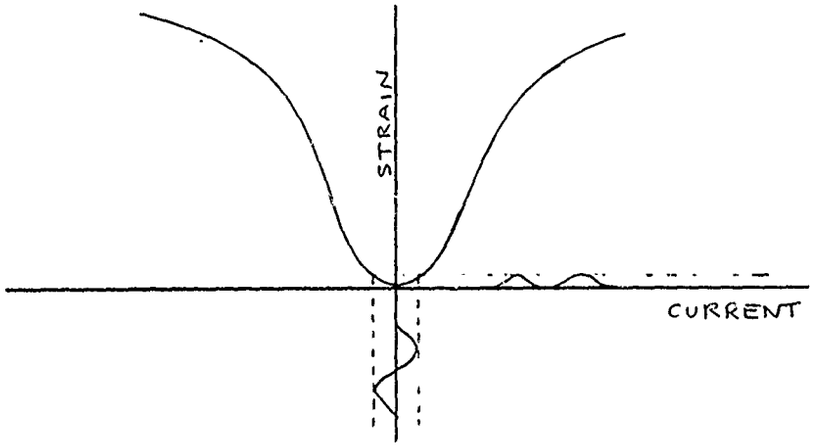
$$1 \text{ weber/m}^2 = 10^4 \text{ gauss}$$

The constant μ_0 has a value of 1 in the CGS system and $4\pi \cdot 10^{-7}$ in the MKS system. A typical magnetization curve of B vs. H is shown below. It is seen that the curve



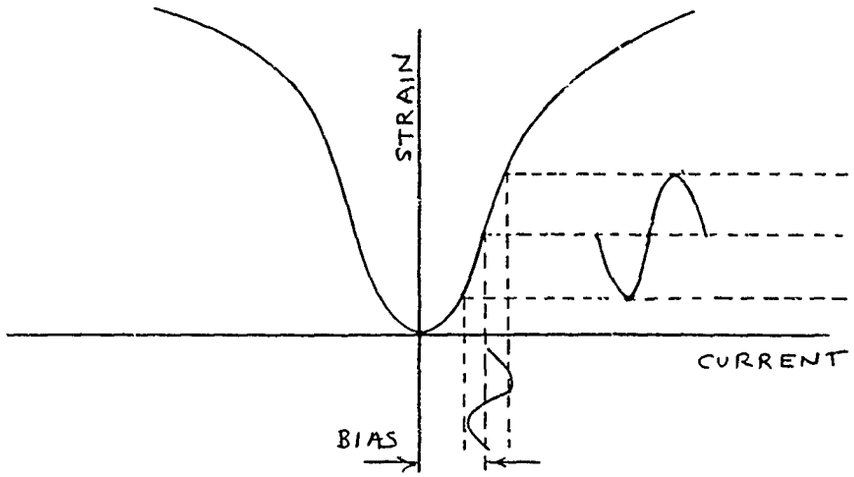
is non-linear, so that μ is not a constant but varies with the field strength. At large field strengths the material saturates.

If we plot the strain $\frac{\partial \xi}{\partial x}$ produced in a nickel tube as a function of the driving current in the coil, we obtain a curve similar to that shown below. This

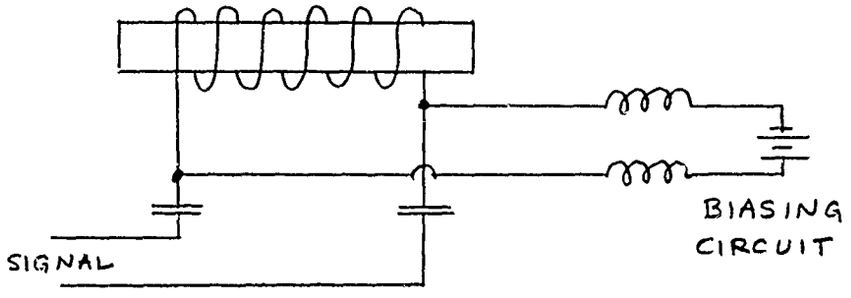


curve exhibits the combined characteristics of the magnetization curve and the parabolic relationship between the strain and the flux density, and is therefore symmetric with respect to the current. It is seen that if an alternating current is applied to the coil, the system will behave as a rectifier, as indicated in the above diagram. Furthermore, because the curve has a small slope near the origin, the output for small currents will be very small. (A large current would introduce gross distortions in waveform.)

In practical applications it is desirable that the system be as linear as practicable. Examination of the curve shows that the most nearly linear portion--and, fortunately, the largest slope--occurs in the neighborhood of the point of inflection. Optimum operation can therefore be achieved by biasing the material with a constant field strength sufficient to rais the flux density to this level when no current flows. In this way approximately linear operation with a relatively high sensitivity can be achieved, as illustrated in the diagram below. The bias

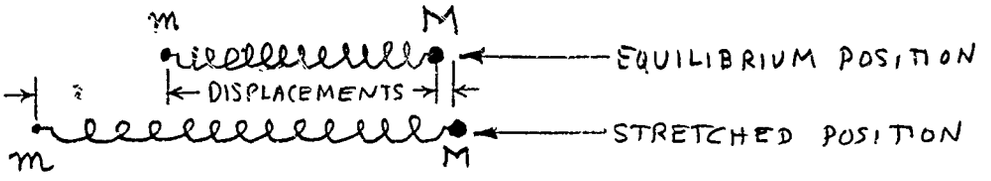


may be produced either by means of a permanent magnet or by means of another coil which is energized by direct current. An alternate method of current biasing is to use the same coil for both currents and to separate the currents by means of inductors and capacitors as shown below.



A commonly used form of magnetostrictive element is a thin-walled tube. This form has the advantage of low eddy-current losses, but its mass is insufficient for efficient coupling to the water. Solid rods are undesirable because of large eddy currents. The coupling problem is solved by attaching a heavy mass to one end of the tube, the face of this mass being coupled to the water. In many practical transducer designs a large number of tubes are attached to a single large plate. Each tube is provided with a separate coil and all coils are energized in phase. If the mass of the plate is large compared with the mass of the tubes, each tube will behave approximately as a quarter-wavelength resonator. The free ends of the tubes are placed in air or some pressure-release material. Since the heavy mass (the plate) is near the node of the tube vibration, it will be driven with great force and small amplitude. In a properly designed transducer the mechanical impedance of the system will match the impedance of the water.

If we consider the dynamics of a single tube operating at resonance with a mass M attached to one end, it can be shown that the system may be represented by two lumped masses M and m connected by a spring, as shown below. The lumped mass M is located at the center of gravity of the actual attached mass. The lumped mass m represents the vibrating tube. This mass is located at the far



end of the tube and has a value equal to one half the actual mass of the tube. The stiffness s of the spring which connects the masses is found to be

$$s = \frac{\pi^2 YS}{8\ell}$$

where Y = Young's modulus

S = cross-sectional area of the tube

$\ell \approx \lambda_m/4$ = length of the tube

λ_m = wavelength of the wave in the tube.

It will be recalled that both the effective mass and the spring stiffness are the same as we derived previously for a quarter-wave crystal oscillator at resonance. The systems differ, of course, in that the crystal was clamped at one end and radiated into the water at the other end, whereas in the magnetostrictive

oscillator the free end is on the back side away from the water and the "clamped" end is not really clamped but is attached to the heavy mass which is located just beyond the node and vibrates with a small amplitude. A simple analysis shows that the displacements, velocities, and accelerations of the lumped masses on either end of the spring are inversely proportional to the respective masses. The vibration of the free end with a small mass and large amplitude thus balances the vibration of the large mass on the other end. The node of the actual vibration is located at the center of gravity of the two lumped masses. The net effect is the same as would be obtained if the free end were removed and the bar were rigidly clamped (or attached to an infinite mass) at the node.

It will be noted that the electric input to the piezoelectric projector was expressed in terms of the applied voltage, whereas the input to the magnetostrictive projector has been expressed in terms of the current. This is but one aspect of a general relationship between the equivalent circuits of the two types. The circuits are related by the principle of <u>duality</u>, by which the current in one circuit corresponds to the voltage in the other, impedance in one to admittance in the other, series connections in one to parallel connections in the other, etc.

TRANSDUCERS (continued)

B. Directional Characteristics

1. Introduction

We have seen that a small sound source, such as a pulsating sphere whose radius is small compared with a wavelength, generates a spherically symmetric (omnidirectional) sound field in which the pressure is a function of only the radial coordinate, as indicated by equation (136)

$$p = \frac{A}{r} e^{j(wt-kr)} \qquad (136)$$

The situation is different in the case of a transducer having extended dimensions. The simplest case to consider is the case of two omnidirectional sources, each of the same strength (that is, same value of A in (136)) and operating at the same frequency and with the same phase. If the two sources are separated from each other by a finite distance, it is clear that to any given point in the sound field the distances traveled by the waves from the two sources will in general not be the same, and hence the two waves will interfere with one another. There will be some directions in space where the two individual pressures will be in phase and resultant pressure will be large. At other locations the waves will differ in phase by various amounts and the resultant pressure will be smaller. When the resultant pressure is measured or computed as a function of the direction relative to a fixed set of axes, there results a pattern called the directivity pattern of the transducer. In the case of an array of two omnidirectional sources, the pattern is a function of the ratio of the separation distance to the wavelength of the sound waves.

In the case of a transducer whose radiating surface covers a finite area S, the directivity pattern may be computed by dividing the area

into infinitesimal elements dS and considering each little element as an omnidirectional source. The resultant pressure at any point in the sound field is then computed by adding up all the incremental pressures due to the elements dS, that is, by an integration process.

In the case of a transducer having only a linear dimension, such as a straight line or a ring, the linear dimension may be divided into infinitesimal increments, each increment being considered an infinitesimal omnidirectional sound source, and the resultant pressure at any point in the sound field may be computed by a similar integration process. Real transducers, of course, are never one-dimensional lines, but a cylinder, for example, whose length is large compared with its diameter may be approximated by a hypothetical ideal linear transducer.

In describing the directivity pattern of a transducer we are concerned only with relative (not absolute) values of the pressure. In general, for any transducer, there is a single direction, or a locus of directions, along which the pressure is a maximum. In the case of a circular piston, for example, the pressure is a maximum along a line normal to the circular face. This is the direction of the axis of the sonar beam. The actual value of the pressure at a specified reference distance from the transducer along this axis, when considered in relation to the input electrical current, is a measure of the sensitivity of the transducer, and will be discussed in a separate section. In considering the directivity pattern we are concerned only with the relative pressure response, which will be defined as the ratio of the pressure in any direction (θ, ϕ) to the pressure along the axis, both pressures measured at the same distance from the transducer. If $p(\theta, \phi)$ represents the pressure at a given distance along the direction (θ, ϕ), and p_o represents the pressure along the beam axis at the same distance, then

$$\text{Relative pressure response} = \frac{p(\theta, \phi)}{p_o} \qquad (701)$$

Since the intensity is proportional to the square of the pressure, the relative intensity response, or, since we are talking about a projector.

the relative transmitting (intensity) response, $\eta(\theta, \phi)$, is

$$\eta(\theta, \phi) = \left[\frac{p(\theta, \phi)}{p_o}\right]^2 \tag{702}$$

In describing directivity patterns we must distinguish between the near field and the far field. The near field may be loosely considered to be that region in space which lies within a few wavelengths of the transducer (or transducer array). If we divide the radiating surface into infinitesimal elements and consider each element as a simple source, then the distances from these elements to a point in the near field will exhibit large percentage variations, so that the amplitudes of the waves from the closest elements will be appreciably larger than the amplitudes of the waves from the more distant elements. Thus, in the near field, the contributions from the closest elements will have a proportionately larger effect than the contributions from the more distant elements. The structure of the near field is therefore in general quite complicated, and exact mathematical computations are all but impossible. Fortunately, in most sonar applications the near field is of little concern to any one except the transducer designer. It should be pointed out, however, that interest in the near field has recently been generated by practical problems associated with the calibration of large transducer arrays.

The far field is that region of space where the distance from the transducer is large compared with the transducer dimensions. In this case the transducer may be considered as effectively equivalent to a point source having the appropriate directional characteristics. The distance must be great enough to permit the two following assumptions. First, the spreading loss to any point in the far field is the same for all elements of the transducer. Differences in path length produce differences in phase but not in amplitude. Second, ray paths from all elements of the transducer to any point in the far field are parallel lines. This second assumption permits

us to talk about directions in space, as we did in defining the relative pressure response, without worrying about the differences in direction associated with finite transducer dimensions. (In the near field the paths from different elements of the transducer have significantly different directions.) In these notes we shall discuss only far-field directivity patterns.

The discussion thus far has been concerned with acoustic projectors. Similar considerations apply to hydrophones. In this case we consider plane waves impinging on the hydrophone (or hydrophone array). When a wave of a given intensity arrives along the maximum response axis, which we shall call the "beam" axis, we shall designate the output voltage as e_o. When a wave of the same intensity arrives from any other direction (θ, ϕ), we shall designate the output voltage as $e(\theta, \phi)$. The relative voltage response is then

$$\text{Relative voltage response} = \frac{e(\theta, \phi)}{e_o} \tag{703}$$

The corresponding power response, or relative receiving response, $\eta'(\theta, \phi)$, is the square of this,

$$\eta'(\theta, \phi) = \left[\frac{e(\theta, \phi)}{e_o}\right]^2 \tag{704}$$

A transducer which has the same relative transmitting and receiving response, that is, for which

$$\eta'(\theta, \phi) = \eta(\theta, \phi) \tag{705}$$

is called a reversible transducer.

When the relative response is expressed in decibels, it is called the deviation loss, and will be designated by a large N. Thus

$$N(\theta, \phi) = 10 \log \frac{1}{\eta(\theta, \phi)} = -10 \log \eta(\theta, \phi) \tag{706}$$

$$N'(\theta, \phi) = 10 \log \frac{1}{\eta'(\theta, \phi)} = -10 \log \eta'(\theta, \phi) \tag{706a}$$

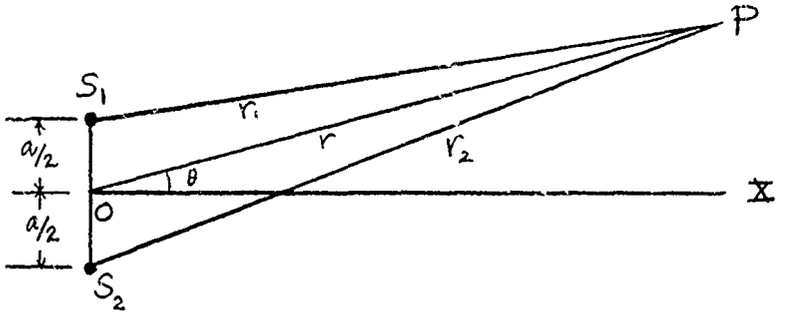
In the derivation of directivity patterns for specific transducer types we shall talk in terms of projectors. However, the same patterns will apply to the identical (ideal) transducers if used as hydrophones.

An introductory note is in order relative to the specification of directions in space in terms of the angles (θ, ϕ). For the transducer types to be discussed we shall employ spherical polar coordinates. Depending upon the transducer type, two different forms will be used. In both types the angle ϕ will correspond to longitude on the earth. In one type θ will correspond to latitude on the earth, that is, θ will be measured north and south from the equator. In the other type θ will correspond to the co-latitude, measured away from the pole.

2. Directivity Patterns for Specific Transducer Types

a. Two-spot Array

We shall consider first the pattern of an array consisting of two identical omnidirectional sources separated by a distance a and operating in phase with each other. Let us set up a coordinate system having its origin at the midpoint of the line joining the two sources S_1 and S_2. We



desire to compute the resultant pressure at the point P located at a distance r from the origin, along a line making an angle θ with the line OX, which is perpendicular to $S_1 S_2$. Assuming the individual pressure due to each source separately to be given by an expression of the form (136), the resultant pressure is

$$p(\theta) = \frac{A}{r_1} e^{j(wt-kr_1)} + \frac{A}{r_2} e^{j(wt-kr_2)} \qquad (707)$$

where r_1 and r_2 are the respective distances S_1P and S_2P from the sources to the point P, and k is the wave number,

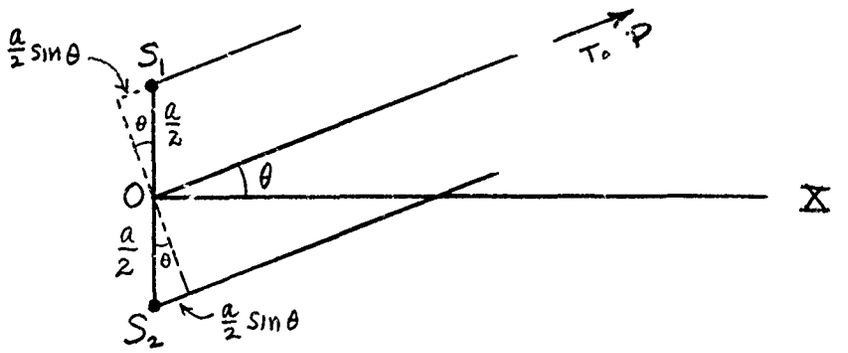
$$k = \frac{2\pi}{\lambda}$$

If the point P is in the near field, as suggested by the preceding sketch, the expressions for r_1 and r_2 in terms of r, θ, and a involve radicals and are very complicated. We shall therefore restrict our analysis to the far field, where $r >> a$. In this case the three lines S_1P, OP, and S_2P are essentially parallel, and

$$r_1 = r - \frac{a}{2} \sin\theta \tag{708}$$

and

$$r_2 = r + \frac{a}{2} \sin\theta \tag{708a}$$



Furthermore, since a is very small compared to r, we may with negligible error replace both r_1 and r_2 with r in the denominators of (707), obtaining

$$p(\theta) = \frac{A}{r} e^{j(wt-kr+\frac{ka}{2}\sin\theta)} + e^{j(wt-kr-\frac{ka}{2}\sin\theta)}$$

$$= \frac{Ae^{j(wt-kr)}}{r} (e^{\frac{1}{2}jka\sin\theta} + e^{-\frac{1}{2}jka\sin\theta}) \tag{709}$$

To simplify the notation, let

$$\psi = \frac{ka}{2}\sin\theta = \frac{\pi a}{\lambda}\sin\theta \tag{710}$$

Then

$$p(\theta) = \frac{2Ae^{j(wt-kr)}}{r} \cos\psi \tag{709a}$$

It is clear that the resultant pressure is a maximum in the direction of the axis OX, since in this direction the path lengths r_1 and r_2 are equal and the pressures add in phase. In this direction

$$\theta = \psi = 0$$

and therefore the maximum pressure at a distance r is

$$p_o = \frac{2A\, e^{j(\omega t - kr)}}{r} \tag{711}$$

Hence the relative pressure response (701) is

$$\frac{p(\theta)}{p_o} = \cos \psi \tag{712}$$

or, in an alternate form

$$\frac{p(\theta)}{p_o} = \frac{\sin 2\psi}{2 \sin \psi} \tag{712a}$$

The above procedure illustrates the basic far-field assumptions which will be employed in computing directivity patterns of other transducer types. First, the fractional difference in path length between r_1 and r_2 is so small that they may both be replaced by the average distance r in the expressions for the amplitude of the waves. The path difference affects only the relative phases of the waves at the field point P where the resultant pressure is computed. Second, the paths r_1 and r_2 are so nearly parallel that their difference in length may be computed on the assumption that they are parallel lines.

The pattern of this simple array is obtained by plotting the pressure ratio (712) as a function of θ. It is seen first of all that the pressure is a maximum at $\theta = 0$ and that the pattern is symmetric (the same for positive and negative values of θ). From the definition of ψ (710) it is seen that the shape of the pattern depends upon the ratio $\frac{a}{\lambda}$, that is, upon the spacing of the two sources in relation to the wavelength of the sound waves. Suppose, for example, that

$$a = 2\lambda$$

Then, as θ increases from 0 to $\pi/2$, ψ will increase from 0 to 2π, and cos ψ will vary from 1 to 0 to -1 to 0 and back to +1. The points at which the response is zero ($\psi = \pm \pi/2$ and $\pm 3\pi/2$ in this example) are called nulls. The central portion between the first null on either side, and containing the direction of maximum response, is called the beam or the major or primary lobe. The portions between adjacent nulls on either side of the major lobe are called secondary lobes or side lobes. In the present example there are two side lobes on either side. It is seen that in the first side lobe ($\pi/2 < \psi < 3\pi/2$) the pressure ratio is negative, indicating a phase reversal. In the second side lobe ($3\pi/2 < \psi < 2\pi$) the ratio is positive again. This is a general characteristic of most transducers - the phase alternates between successive side lobes. The relative intensity response, being the square of the pressure response, is of course positive in all lobes. It is seen that for a two-spot array the maximum intensity in each of the side lobes is the same as in the primary lobe. This, of course, is not a good design for a practical sonar transducer, and, as we shall see, appreciable reduction in the heights of the side lobes can be achieved in more sophisticated designs.

Another factor of practical importance is the response in the "end-fire" direction, that is, in the direction of the line joining the sources ($\theta = 90°$). In many applications it is desirable to have a low response in the direction at right angles to the beam in order to reject interfering noise. In the example we have chosen the response at 90° is a maximum. It can be seen that there will be a null here if the separation distance of the two sources is equal to an odd number of half wavelengths.

An important measure of the directional characteristics of a transducer is the beamwidth. Unfortunately, there is no universally accepted standard definition of beamwidth. From a theoretical standpoint it would be convenient to define the beamwidth as the angular distance between the first null on either side of the beam. Such a definition is not particularly practical, since the region in the vicinity of the null is of no practical value.

Three different definitions are in common use, all of them being expressed as the angular spread between points of various relative intensities, that is, between the

3 db-down points

6 db-down points

10 db-down points

The corresponding intensity ratios are 1/2, 1/4, and 1/10, and the corresponding pressure ratios are

Db Down	Rel. Press. Resp.
3	0.707
6	0.500
10	0.316

In the above example, where $a = 2\lambda$, the beamwidth is 2θ, where

$$\cos(2\pi \sin\theta) = \frac{p(\theta)}{p_o}$$

The computation of beamwidths for the two-spot array with a 2λ spacing is shown in the following table

Db	$\cos(2\pi \sin\theta)$	$2\pi \sin\theta$	$\sin\theta$	2θ
3	.707	$.250\,\pi$	.125	14.4°
6	.500	$.333\,\pi$	.167	19.2°
10	.316	$.398\,\pi$	.199	22.9°

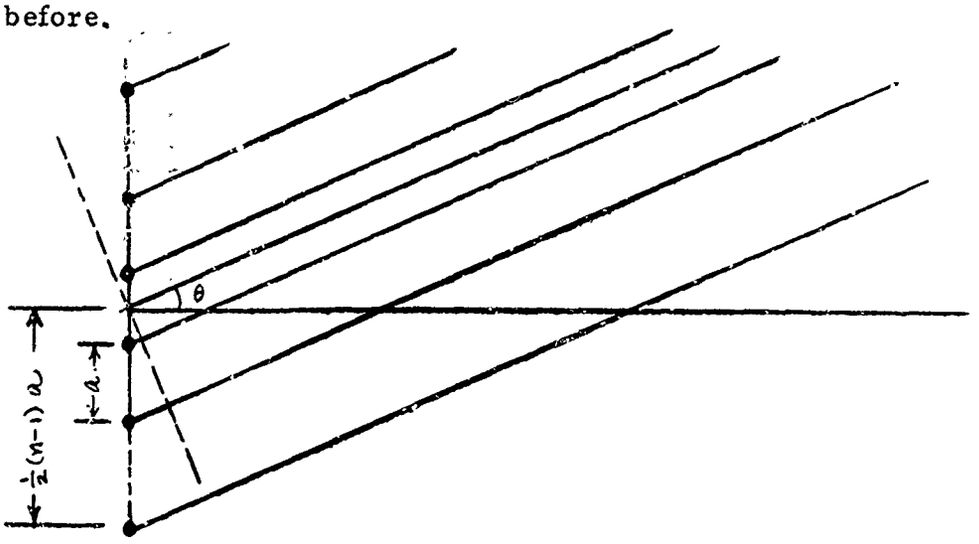
It can be seen that the three definitions lead to significantly different numerical values. For this reason it is advisable when specifying beamwidths to state the definition to which the values apply.

It will be noted that in (709) and (712) the pressure is written as a function only of the angle θ. The angle ϕ is measured in a plane perpendicular to the line $S_1 S_2$. From the geometry of the situation it is clear that the pattern is symmetric about the axis $S_1 S_2$. The actual pattern, of course, is three-dimensional, and the two-dimensional pattern discussed above is only a trace of the three-dimensional pattern in the plane of the paper. The actual pattern is therefore obtained by rotating the two-dimensional pattern about the axis $S_1 S_2$. We thus see that the "beam," or major lobe, is not a beam at all in the sense of a searchlight beam, but rather a doughnut-shaped affair; and the direction of maximum response is not a single direction but rather a locus of all such directions in the equatorial plane. The side lobes are conically-shaped regions. Comparing our polar coordinate system with that of the earth we see that the angle ϕ corresponds to longitude, as stated earlier, and θ corresponds to latitude. The primary lobe covers the equatorial region, and the pole is the line joining the two sources. This coordinate system is useful in describing all axially symmetric beam patterns whose maximum response is in the equatorial plane.

b. Multi-Spot Array

Let us now consider a linear array consisting of n equally spaced omnidirectional sources, the spacing between adjacent ones being a. All sources are assumed to be identical and to be operating in phase. The mathematical formulation is slightly different, depending on whether the number of elements is even or odd, but the end result is the same. For convenience we shall assume n to be even. We choose the origin of the

coordinate system at the midpoint of the array and define the angle θ as before.



To obtain the far-field pattern, we assume equal amplitudes for the pressures due to all individual sources, and we treat all the ray paths as being parallel. The resultant pressure at a distance r in the direction θ is merely an extension of (709)

$$p(\theta) = \frac{Ae^{j(wt-kr)}}{r}\left[e^{\frac{1}{2}j(n-1)ka\sin\theta} + e^{\frac{1}{2}j(n-3)ka\sin\theta} + \ldots + e^{\frac{1}{2}jka\sin\theta}\right.$$

$$\left. + e^{-\frac{1}{2}jka\sin\theta} + \ldots + e^{-\frac{1}{2}j(n-3)ka\sin\theta} + e^{-\frac{1}{2}j(n-1)ka\sin\theta}\right] \qquad (713)$$

The maximum response occurs at $\theta = 0$ and is the sum of the individual pressures, all with the same phase

$$p_o = \frac{nAe^{j(wt-kr)}}{r} \qquad (714)$$

Upon substitution of ψ from (710), the relative pressure response is found to be

$$\frac{p(\theta)}{p_o} = \frac{1}{n}\left[e^{(n-1)j\psi} + e^{(n-3)j\psi} + \ldots + e^{j\psi} + e^{-j\psi} + \ldots + e^{-(n-1)j\psi}\right] \qquad (715)$$

This result may be expressed in either of two alternate forms. First, from the cosine formula

$$\cos x = \tfrac{1}{2}(e^{jx} + e^{-jx})$$

we obtain

$$\frac{p(\theta)}{p_o} = \frac{2}{n}\left[\cos\psi + \cos 3\psi + \ldots + \cos(n-1)\psi\right] \tag{715a}$$

To obtain the alternate expression, factor out $e^{(n-1)j\psi}$ from the series in brackets, leaving

$$1 + e^{-2j\psi} + e^{-4j\psi} + \ldots + e^{-2(n-1)j\psi}$$

This series is of the form

$$1 + x + x^2 + \ldots + x^{n-1} = \frac{1-x^n}{1-x}$$

Hence $$\frac{p(\theta)}{p_o} = \frac{e^{(n-1)j\psi}}{n} \cdot \frac{1-e^{-2nj\psi}}{1-e^{-2j\psi}}$$

$$= \frac{1}{n}\,\frac{e^{nj\psi} - e^{-nj\psi}}{ej\psi - e^{-j\psi}}$$

And, since $$\sin x = \frac{1}{2j}\left(e^{jx} - e^{-jx}\right)$$

we obtain

$$\frac{p(\theta)}{p_o} = \frac{\sin n\psi}{n \sin\psi} \tag{715b}$$

where $$\psi = \frac{ka}{2}\sin\theta = \frac{\pi a}{\lambda}\sin\theta \tag{710}$$

The nulls in the multispot pattern occur for values of $\psi = \psi_\nu$ such that

$$n\psi_\nu = \nu\pi, \qquad \nu = 1, 2, \ldots \tag{716}$$

The corresponding beam pattern angles θ_ν are obtained by inserting (710)

$$\sin\theta_\nu = \frac{\nu\lambda}{na} \tag{717}$$

Since $\sin\theta_\nu \leq 1$, the maximum number of nulls on either side of the main lobe is

$$\nu \leq \frac{na}{\lambda} \tag{718}$$

The points of maximum response (the peaks) of the lobes can be obtained by setting the derivative of (715) equal to zero. The resulting equation is

$$\tan n\psi = n \tan\psi \tag{719}$$

This equation is most conveniently solved for given value of n by plotting both tan $n\psi$ and $n \tan \psi$ vs. ψ and reading off the values of ψ at the intersections of the curves. If the array contains more than two elements, the values of $n\psi$ which satisfy (719) are reasonably close to odd integral multiples of $\pi/2$, and hence the corresponding values of $\sin n\psi$ are fairly close to unity. Therefore, if we substitute $\sin n\psi = 1$ in (715), the resulting envelope equation

$$\left(\frac{p(\theta)}{p_o}\right)_{env.} = \frac{1}{n \sin \psi} \tag{720}$$

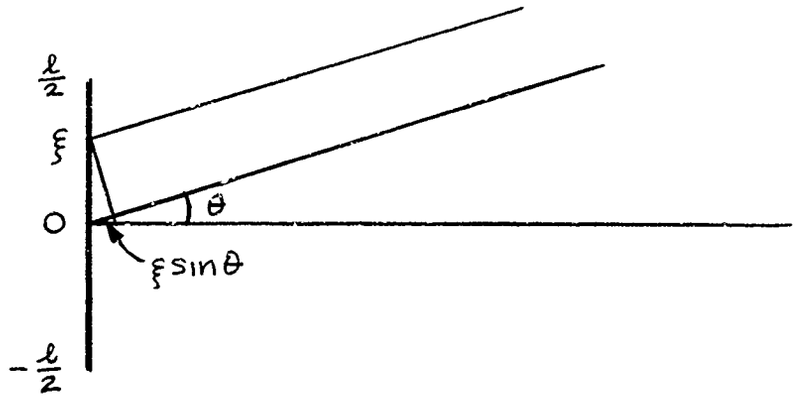
gives a good measure of the maximum response in the side lobes at the appropriate values of ψ. Since it is usually desirable to reduce the response in the side lobes as much as possible, we should like $\sin \psi$ to continue to increase as the angle θ increases from 0 to 90°. Reference to (710) reveals that the spacing, a, between adjacent elements should not exceed one-half wavelength. If $a = \frac{1}{2}\lambda$, then

$$\psi = \frac{\pi}{2} \sin \theta$$

so that when $\theta = \frac{\pi}{2}$, $\psi = \frac{\pi}{2}$. With this spacing it is seen that $\sin \psi$ increases steadily from 0 to 1 as θ goes from 0 to 90°. If the spacing is less than $\frac{1}{2}\lambda$, $\sin \psi$ increases with θ, but the maximum value is less than 1. On the other hand, if the spacing is greater than $\frac{1}{2}\lambda$, $\sin \psi$ reaches its maximum value (and the envelope of the beam pattern reaches its minimum value) at a value of θ less than 90°. Beyond this point the side lobes begin to build up again.

c. Continuous Linear Transducer

In analyzing any continuous line type transducer, whether a straight line or any other shape, we divide the line into infinitesimal elements of length $d\xi$ and treat each of these as an omnidirectional source of strength $Ad\xi$, where the constant A is a measure of the pressure produced per unit length of the line. In the present discussion we shall analyze a straight-line transducer of length ℓ. We shall set the origin of our coordinate system at the midpoint of the line, as indicated in the sketch. Consider an element $d\xi$



at a distance ξ from the origin. The length of the path from this element to a point P in the far field, located at a distance r at an angle θ from the origin, is $r - \xi \sin\theta$. Neglecting ξ in comparison with r in the amplitude factor, the incremental pressure at P due to $d\xi$ is seen to be

$$dp(\theta) = \frac{A d\xi}{r} e^{j(wt-kr+k\xi \sin\theta)}$$

and the resultant pressure is

$$p(\theta) = \frac{A}{r} e^{j(wt-kr)} \int_{-\ell/2}^{\ell/2} e^{jk\xi \sin\theta} d\xi$$

which readily integrates to

$$p(\theta) = \frac{A\ell}{r} e^{j(wt-kr)} \frac{\sin\left(\frac{k\ell}{2} \sin\theta\right)}{\frac{k\ell}{2} \sin\theta}$$

As in the case of the multispot array, the pressure is a maximum on the axis $\theta = 0$, where all elements of the line add in phase, giving

$$p_o = \frac{A\ell}{r} e^{j(wt-kr)}$$

The relative pressure response is then

$$\frac{p(\theta)}{p_o} = \frac{\sin\psi}{\psi} \tag{721}$$

where

$$\psi = \frac{k\ell}{2} \sin\theta = \frac{\pi\ell}{\lambda} \sin\theta \tag{722}$$

It should be noted that the same result can be obtained from the multispot array by the limiting process of letting the number of elements go to infinity while maintaining a fixed length of the array. To do this, we let

$$n \rightarrow \infty$$

$$a \rightarrow 0$$

$$na = \ell$$

in equation (715). When this is done, it is seen that $\frac{\pi a}{\lambda} \sin\theta$ becomes an increasingly small angle, so that

$$n \sin\left(\frac{\pi a}{\lambda} \sin\theta\right) \rightarrow \frac{\pi na}{\lambda} \sin\theta \rightarrow \frac{\pi \ell}{\lambda} \sin\theta$$

while $$\sin\left(\frac{\pi na}{\lambda} \sin\theta\right) \rightarrow \sin\left(\frac{\pi \ell}{\lambda} \sin\theta\right)$$

Hence

$$\frac{p(\theta)}{p_o} \rightarrow \frac{\sin\left(\frac{\pi \ell}{\lambda} \sin\theta\right)}{\frac{\pi \ell}{\lambda} \sin\theta}$$

which agrees with (721).

As may be expected, the pattern of a continuous line transducer is similar to that of a multispot array containing a large number of elements. The nulls occur where

$$\psi = n\pi, \qquad n = 1, 2, 3, \ldots \tag{723}$$

or

$$\sin\theta = \frac{n\lambda}{\ell} \tag{724}$$

The locations of the maxima of the side lobes are found by setting the derivative of (721) equal to zero. The corresponding values of ψ are obtained as solutions of the transcendental equation

$$\tan\psi = \psi \tag{725}$$

The solutions are listed in many books of mathematical tables. The first five values are tabulated below.

Side Lobe No.	ψ at Max. (radians)	Height of lobe (db down)
1	4.493	13.26
2	7.725	17.83
3	10.904	20.79
4	14.066	22.99
5	17.221	24.73

For any given value of $\frac{\ell}{\lambda}$ the maximum value of ψ occurs at $\theta = 90°$ and is

$$\psi_{max} = \frac{\pi \ell}{\lambda} \tag{726}$$

The maximum number of maxima on either side of the beam can be found by comparing ψ_{max} with the values listed in the above table. The value of the response at each of the maxima is found by substituting the values from the table into (721). The decibel equivalents of these maxima, that is,

$$-20 \log \frac{p(\vartheta)}{p_o}$$

are also listed in the table.

The beamwidth of a linear transducer is obtained by setting

$$-20 \log \frac{p(\theta)}{p_o} = \begin{Bmatrix} 3 \\ 6 \\ 10 \end{Bmatrix} \text{ db}$$

or

$$\frac{\sin \psi}{\psi} = \begin{Bmatrix} 0.707 \text{ (3 db)} \\ 0.500 \text{ (6 db)} \\ 0.316 \text{ (10 db)} \end{Bmatrix}$$

which leads to

$$\sin \theta = \begin{Bmatrix} 0.443 \ \lambda/\ell \text{ (3 db)} \\ 0.602 \ \lambda/\ell \text{ (6 db)} \\ 0.738 \ \lambda/\ell \text{ (10 db)} \end{Bmatrix}$$

The beamwidth, of course, is 2θ.

As in the case of multispot arrays, the beam pattern in three dimensions is obtained by rotating the two-dimensional pattern about the transducer axis.

d. Ring Transducer

Another example of a continuous line transducer is the circular ring. The procedure for deriving the beam pattern for a ring is basically the same as for a line. There is, however, a difference in the nature of the pattern, which leads to the selection of a different coordinate system. The beam of a circular ring transducer is like that of a searchlight. The beam axis is a single line through the center of the ring normal to its plane. It is logical therefore to measure the angle θ from the axis, that is, the co-latitude measured from the pole, rather than the latitude measured from the equator. The angle ϕ, of course, is still the longitude angle. From symmetry considerations the pattern of a circular ring is independent of ϕ.

We shall merely state the resulting equation of the relative pressure response without deriving it.

$$\frac{p(\theta)}{p_o} = J_o(\psi) \tag{726}$$

where

$$\psi = \frac{\pi d}{\lambda} \sin\theta$$

d = diameter of ring

J_o = Bessel function of zero order

e. Circular Piston

The circular piston is one of the simplest and most commonl used forms of transducer involving a continuous area, and we shall derive its beam pattern to illustrate the method of analysis employed for area type transducers.

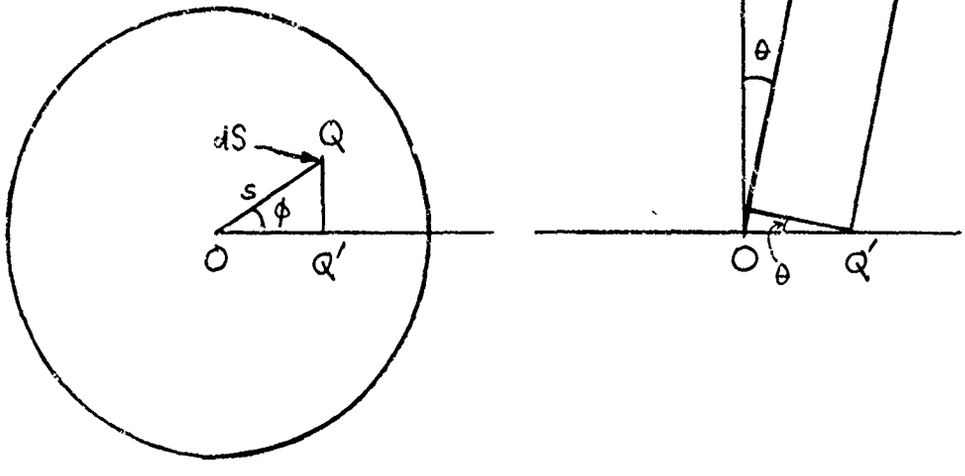
We choose the center of the circular face as the origin of our coordinate system. The axis of symmetry is the line through the origin perpendicular to the plane of the face. This is the axis of the beam. We shall define θ as the polar angle measured away from this axis. The plane of the radiating face thus corresponds to the equatorial plane of the earth and

θ corresponds to the co-latitude. The longitude angle ϕ is measured in the plane of the face relative to an arbitrarily chosen reference direction (Greenwich meridian). From the geometry of the situation it is clear that the pressure in the sound field generated by this transducer is symmetric about the axis and is thus independent of ϕ. Therefore, without loss of generality we may select the point P in the far field at which we compute the pressure to lie in the $\phi = 0$ plane, the other coordinates being r and θ.

To compute the pressure $p(\theta)$ at the point P we divide the radiating face into infinitesimal elements dS and consider each element to be an omnidirectional source of strength AdS, where A in this case is a measure of the pressure produced per unit area. In accordance with the standard far-field assumptions, the amplitude of the pressure at P due to each element is $\frac{AdS}{r}$ (the same value of r being used for all elements).

Let the radius of the piston face be a, and let the element dS be located at the point Q at a distance s from the center, its location being described by the polar coordinates (s, ϕ). The element of area is

$$dS = s\, d\, s\, d\, \phi$$



We must now compute the distance QP from the radiating element to the far-field point. By hypothesis the point P lies in a plane defined by $\phi = 0$ and the distance OP is equal to r. If we drop a perpendicular from Q to Q' on the $\phi = 0$ axis, it is seen that all points on QQ' are equidistant from P. From the sketch it is seen that

$$Q'P = OP - OQ' \sin\theta$$

or

$$Q'P = r - s \sin\theta \cos\phi$$

The incremental pressure at P is

$$dp(\theta) = \frac{A}{r} e^{j(wt-kr)} e^{jks \sin\theta \cos\phi} s \, d s \, d \phi$$

The resultant pressure is obtained by integrating $dp(\theta)$ over the area of the face. It is

$$p(\theta) = \frac{A}{r} e^{j(wt-kr)} \int_0^a \int_0^{2\pi} e^{jks \sin\theta \cos\phi} s \, d s \, d \phi \qquad (727)$$

This integral cannot be evaluated in closed form. However, a series solution can be obtained by expanding the exponential in powers of $jks \sin\theta \cos\phi$.

If we integrate over ϕ first, we note first of all that from symmetry considerations the integral from 0 to π is equal to the integral from π to 2π. Second, the integrals of odd powers of cos ϕ from 0 to π vanish. Third, if n is even,

$$\int_0^{\pi} \cos^{2n} \phi \, d \phi = \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2 \cdot 4 \cdot 6 \cdots 2n}$$

Fourth, on the axis where $\theta = 0$, the reference pressure po may be evaluated directly, yielding

$$p_o = \frac{\pi a^2 A}{r} e^{j(wt-kr)}$$

The relative pressure response therefore is

$$\frac{p(\theta)}{p_o} = \frac{2}{a^2}\int_o^a \left[1 - \frac{1}{2!}\frac{1}{2}(ks\sin\theta)^2 + \frac{1}{4!}\frac{1.3}{2.4}(ks\sin\theta)^4 - \cdots\right] sds$$

$$= 2\left[\frac{1}{2} - \frac{1}{2!}\frac{1}{4}\frac{1}{2}\psi^2 + \frac{1}{4!}\frac{1}{6}\frac{1.3}{2.4}\psi^4 - \frac{1}{6!}\frac{1}{8}\frac{1.3.5}{2.4.6}\psi^6 + \cdots\right] \quad (728)$$

where

$$\psi = ka\sin\theta = \frac{2\pi a}{\lambda}\sin\theta \quad (729)$$

Equation (728) may be transformed as follows

$$\frac{p(\theta)}{p_o} = \frac{2}{\psi}\left[\frac{\psi}{2} - \frac{\psi^3}{2^3\,1!\,2!} + \frac{\psi^5}{2^5\,2!\,3!} - \frac{\psi^7}{2^7\,3!\,4!} + \cdots\right] \quad (728a)$$

The series in brackets is the Bessel function of first order, $J_1(\psi)$. The pressure ratio may therefore be written as

$$\frac{p(\theta)}{p_o} = \frac{2J_1(\psi)}{\psi} \quad (728b)$$

The first order Bessel function is an oscillatory function resembling a damped sine wave, and the pattern as a function of θ is similar to that of a linear transducer. The first five nulls occur at the following values of ψ :

Null No.	ψ (radians)
1	3.832
2	7.016
3	10.174
4	13.324
5	16.471

The maxima of the first five side lobes are as follows:

Lobe No.	ψ	Db down
1	5.136	17.57
2	8.417	23.81
3	11.620	27.96
4	14.796	31.08
5	17.960	33.60

It should be noted that the three-dimensional beam pattern of a circular piston is fundamentally different from that of a line. The primary lobe is directed along the axis of the circular face rather than being spread out over the equatorial plane.

f. Rectangular Plate

The method of analysis for a rectangular plate transducer is similar to that of the circular piston, except that the integration is carried out over a rectangular instead of a circular area. There is a significant difference, however, in the resulting beam pattern. The pattern is no longer symmetric about the axis; it is a function of ϕ as well as θ. We shall omit the derivation. If the length and width of the rectangle are ℓ and w, respectively, and the arbitrary reference direction from which ϕ is measured is taken parallel to the side ℓ, the relative pressure response is found to be

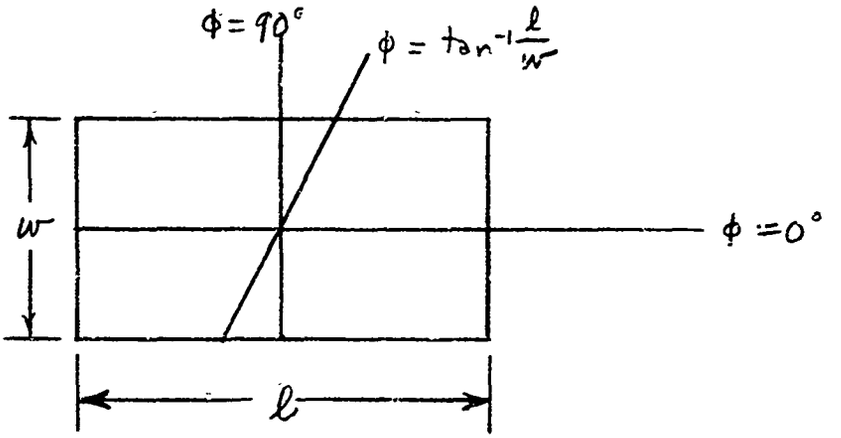
$$\frac{p(\theta,\phi)}{p_o} = \frac{\sin\psi_\ell}{\psi_\ell} \frac{\sin\psi_w}{\psi_w} \tag{730}$$

where

$$\psi_\ell = \tfrac{1}{2} k \ell \sin\theta\cos\phi = \frac{\pi\ell}{\lambda}\sin\theta\cos\phi \tag{731}$$

$$\psi_w = \tfrac{1}{2} k w \sin\theta\sin\phi = \frac{\pi w}{\lambda}\sin\theta\sin\phi \tag{731a}$$

This pattern has several interesting characteristics. Consider first the case where $\phi = 0$. This gives the pattern in the plane



perpendicular to the side w. It is seen that

$$\frac{\sin \psi_w}{\psi_w} = 1$$

since $\psi_w = 0$. Hence

$$\frac{p(\theta, 0)}{p_o} = \frac{\sin \psi_\ell}{\psi_\ell} \tag{732}$$

The pattern in this plane is the same as that of a linear transducer of length ℓ. Similarly in the plane $\phi = 90°$, $\psi_\ell = 0$ and

$$\frac{p(\theta, \frac{\pi}{2})}{p_o} = \frac{\sin \psi_w}{\psi_w} \tag{733}$$

This is the same as the pattern of a line of length w. Next, consider a value ϕ such that

$$w \sin \phi = \ell \cos \phi$$

or

$$\phi = \tan^{-1} \frac{\ell}{w} \tag{734}$$

This leads to

$$w \sin \phi = \ell \cos \phi = \frac{\ell w}{\sqrt{\ell^2 + w^2}}$$

The relative voltage response is

$$\frac{p(\theta,\ \tan^{-1}\frac{\ell}{w})}{p_o} = \left[\frac{\sin\frac{\pi\,\ell\,w\,\sin\theta}{\lambda\sqrt{\ell^2+w^2}}}{\frac{\pi\,\ell\,w\,\sin\theta}{\lambda\sqrt{\ell^2+w^2}}}\right]^2 \qquad (735)$$

In the special case where $w = \ell$, the value of ϕ is 45° and the relative pressure response is

$$\frac{p\ (\theta,\ \frac{\pi}{4})}{p_o} = \left[\frac{\sin\frac{\pi\,\ell\,\sin\theta}{\sqrt{2}\lambda}}{\frac{\pi\,\ell\,\sin\theta}{\sqrt{2}\lambda}}\right]^2 \qquad (735a)$$

It is seen that in the plane defined by (734) the relative pressure response is the square of a $\frac{\sin x}{x}$ function. It therefore does not exhibit the usual phase reversals which normally occur between successive side lobes, and furthermore the side lobes are much weaker than those of a linear transducer of equivalent length; they are twice as many db down. In the special case of a square radiating surface this condition occurs in the plane of the diagonal.

g. Pressure Gradient Hydrophone

Consider a hydrophone array consisting of two omnidirectional elements separated by a distance a, the two outputs being subtracted rather than added. Suppose that a plane wave is traveling along the direction of the line joining the two elements. If the instantaneous value of the pressure at the midpoint of this line is $P\,e^{jwt}$, then the pressures at the two elements are

$$p_1 = P\,e^{j(wt+\frac{ka}{2})}$$

and

$$p_2 = P\,e^{j(wt-\frac{ka}{2})}$$

and the output voltage of the array is

$$e_o = BP\left[e^{j(wt+\frac{ka}{2})} - e^{j(wt-\frac{ka}{2})}\right]$$

$$= 2jBP\sin\frac{ka}{2}\ e^{jwt} \qquad (736)$$

where B is a constant representing the sensitivity of the elements. If a single omnidirectional element were placed at the location of the midpoint of the

two-element array its output would be

$$BP e^{jwt}$$

It is seen that, relative to the single hydrophone at the center, the output of the two-hydrophone differential array is shifted in phase by 90 degrees and is multiplied by $2 \sin \frac{ka}{2}$.

If the system is to be used as a practical instrument, certain restrictions must be placed upon the spacing, a. For example, if a were equal to a whole wavelength, $a = \lambda$, then the output would be zero, since the two waves would be in phase, and

$$\sin \frac{ka}{2} = \sin \frac{\pi a}{\lambda} = \sin \pi = 0$$

On the other hand, if a were extremely small compared with a wavelength, then

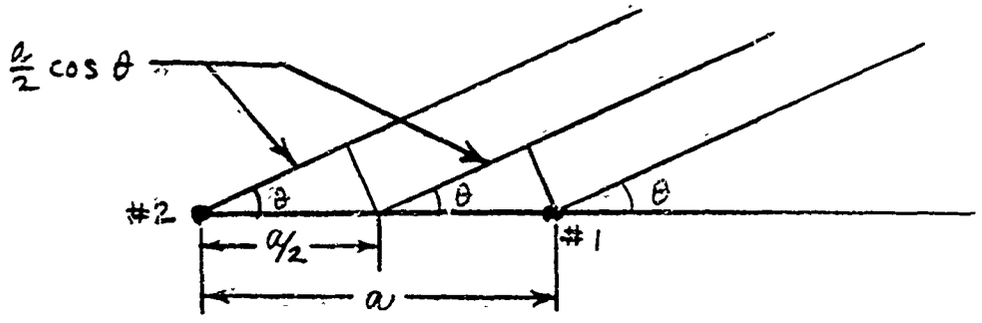
$$\sin \frac{ka}{2} << 1$$

and the output would be extremely small. Although the maximum output occurs when $\sin \frac{ka}{2} = 1$, or $a = \frac{1}{2}\lambda$, it is desirable because of beam pattern considerations (to be discussed below) that the spacing be short enough that the small angle approximation $\sin \frac{ka}{2} \approx \frac{ka}{2}$ may be applied, the upper limit being about 1/6 of a wavelength. If the small angle approximation is applied to (736), the result is

$$e_o \approx jka\, BP\, e^{jwt} \qquad (736a)$$

Comparison with equation (23) shows that the output (736a) is proportional to the pressure gradient at the midpoint of the line joining the two hydrophones. The transducer is therefore called a pressure-gradient hydrophone.

To derive the beam pattern let us suppose that a plane wave is received, whose direction of propagation makes an angle θ with the axis of the hydrophone array. If the pressure at the midpoint is $P e^{jwt}$, then



$$p_1 = P\, e^{j\left(wt + \frac{ka}{2}\cos\theta\right)}$$

$$p_2 = P\, e^{j\left(wt - \frac{ka}{2}\cos\theta\right)}$$

and the output voltage is

$$e(\theta) = 2j\,BP \sin\left(\frac{ka}{2}\cos\theta\right) e^{jwt} \qquad (738)$$

The relative voltage response is then

$$\frac{e(\theta)}{e_o} = \frac{\sin\left(\frac{ka}{2}\cos\theta\right)}{\sin\frac{ka}{2}} \qquad (739)$$

If the spacing of the elements is close enough that $\frac{ka}{2}$ may be considered a small angle, the relative voltage response is approximately

$$\frac{e(\theta)}{e_o} = \cos\theta \qquad (739a)$$

The beam pattern is thus a cosine pattern. A simple calculation will show that larger values of $\frac{ka}{2}$ (such as $\frac{ka}{2} = \frac{\pi}{2}$) tend to broaden the beam.

It is easily shown that the same type of response is obtained from a rectangular block of length a, which is mounted in such a way as to permit it to oscillate in a single dimension.

One of the difficulties in using pressure-gradient hydrophones is the ambiguity in sensing the direction from which a wave is arriving. The only difference in output between a wave having a direction θ and a wave having a direction $\pi - \theta$ is a phase reversal. The rms outputs are identical. To overcome this difficulty, an auxiliary omnidirectional hydrophone is sometimes located in the vicinity of the midpoint and its output is added to that of the pressure-gradient hydrophone. If the output of the auxiliary hydrophone is

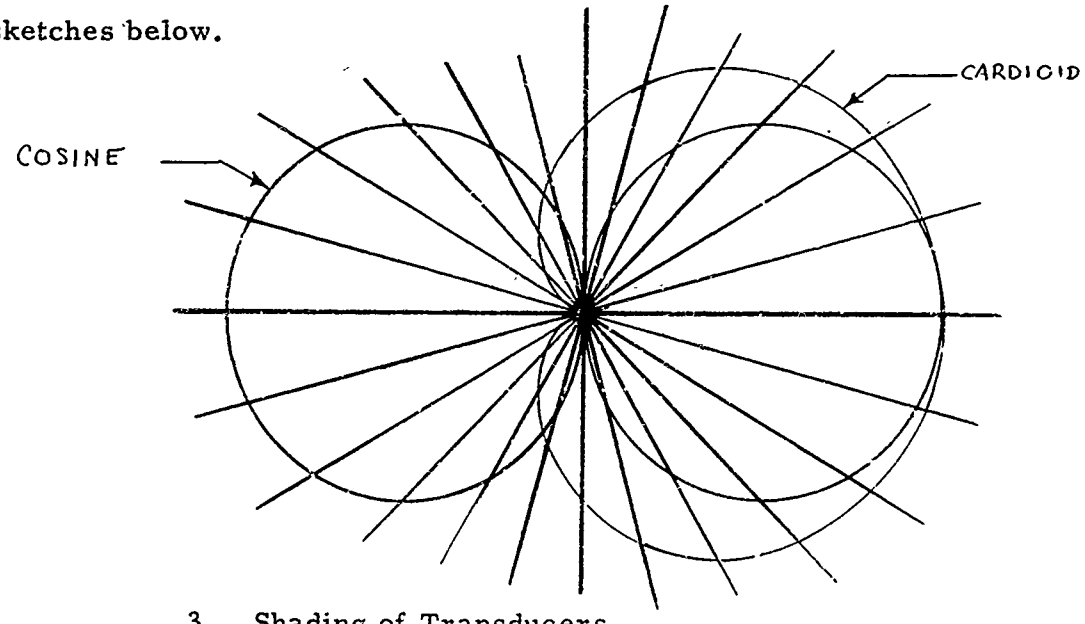
adjusted to be equal to that of the pressure-gradient hydrophone in the direction of maximum response, that is, is equal to ka BP e^{jwt}, and if the phases of the two outputs are shifted relative to each other by 90 degrees, so that both have the same phase, then the combined output is (assuming $a < \frac{\lambda}{6}$)

$$e_{sum}(\theta) = ka\ BP\ (1 + \cos\theta)\ e^{jwt}$$

and the relative pressure response is

$$\left[\frac{e(\theta)}{e_o}\right]_{sum} = \tfrac{1}{2}(1 + \cos\theta) \qquad (740)$$

The auxiliary hydrophone serves to provide phase discrimination. If the incoming wave is arriving from the direction $\theta = 0$, the two outputs are in phase and the resultant pressure is twice that of each separately. If the wave is coming from the direction $\theta = 180°$, the two outputs are 180° out of phase and the resultant is zero. The pattern described by (740) is called a cardioid pattern. The cosine and cardioid patterns are illustrated in the sketches below.



3. Shading of Transducers

The presence of side lobes in a transducer beam pattern is usually undesirable. In a projector the energy radiated into the side lobes is wasted and may under some circumstances produce harmful reverberation. In a hydrophone the noise introduced through the side lobes introduces

additional interference. Also, a signal detected through a side lobe is capable of causing great confusion.

It has been found that the side lobe response of a transducer can be reduced by a technique known as shading. The essence of the technique is to reduce the sensitivity of outer portions of the transducer surface. In the case of a linear multispot array, shading may be accomplished either by reducing the sensitivity of the outer elements relative to those in the center, or by spacing the outer elements farther apart. In a single area-type transducer an equivalent effect can be obtained by breaking up the surface into annular zones and progressively reducing the sensitivity of the outer zones.

The following derivation will illustrate one method of shading a linear multispot array. It is based on an article by C.L. Dolph in the June 1946 issue of the Proceedings of the IRE. Rather than deriving the general formulas, we shall illustrate the method by applying it to the specific case of a 6-element array with a fixed spacing, a, between adjacent elements. The relative pressure response for an unshaded array is given by (715a) with n = 6. We shall assume the shading to be symmetric about the center, so that symmetrically located elements have the same sensitivity. Let A_1 represent the sensitivity of the two center elements, 3 and 4, A_3 the sensitivity of the elements 2 and 5, and A_5 the sensitivity of the outer elements 1 and 6. The relative pressure response of the shaded array is then

$$\frac{p(\theta)}{p_o} = \frac{A_1 \cos \psi + A_3 \cos 3\psi + A_5 \cos 5\psi}{A_1 + A_3 + A_5} \tag{741}$$

If we designate

$$\cos \psi = \mu \tag{742}$$

then the cosines of the multiple angles can be expressed as polynomials in μ.

$$\cos 3\psi = 4\mu^3 - 3\mu \tag{742a}$$

$$\cos 5\psi = 16\mu^5 - 20\mu^3 + 5\mu \tag{742b}$$

These polynomials are known as Chebyshev polynomials and are designated as $T_n(\mu)$ where n represents the highest power of μ appearing in the

polynomial. As ψ varies from 0 to 180 degrees, μ varies from 1 to -1. Since

$$T_n(\mu) = \cos n\psi$$

it is seen that $T_n(\mu)$ has n-1 maxima and minima in the interval $-1 < \mu < +1$, and that its value is either + 1 or -1 at each of these points.

Substitution of these polynomials into (741) yields

$$\frac{p(\theta)}{p_o} = \frac{16A_5\mu^5 - (20A_5 - 4A_3)\mu^3 + (5A_5 - 3A_3 + A_1)\mu}{A_5 + A_3 + A_1} \tag{741a}$$

Since the numerator of this expression is a 5th degree polynomial in μ, let us choose values of the A's to make this a Chebyshev polynomial of the 5th degree. Let z be a parameter (whose value is to be determined) such that the numerator of (741a) is equal to

$$T_5(z\mu) = 16z^5\mu^5 - 20z^3\mu^3 + 5z\mu \tag{743}$$

The A's may be evaluated in terms of z by equating coefficients of like powers of μ. Thus,

$$A_5 = z^5 \tag{744a}$$

$$A_3 = 5z^5 - 5z^3 \tag{744b}$$

$$A_1 = 10z^5 - 15z^3 + 5z \tag{744c}$$

and

$$A_1 + A_3 + A_5 = 16z^5 - 20z^3 + 5z = T_5(z) \tag{744d}$$

The relative pressure response is

$$\frac{p(\theta)}{p_o} = \frac{T_5(z\mu)}{T_5(z)} \tag{741b}$$

We note that the direction of maximum response, $\theta = 0$, corresponds to $\mu = 1$, and to a relative pressure response of unity. The peaks of the side lobes will occur at the maxima and minima of $T_5(z\mu)$, whose value at these points is ± 1. Thus the maximum (absolute) value of the relative pressure response is the same for all the side lobes and is equal to $\frac{1}{T_5(z)}$. Since the side lobe response must be less than the response on the beam axis, it follows that

$$T_5(z) > 1$$

and hence

$$z > 1$$

We are thus concerned with a Chebyshev polynomial outside its normal range.

The problem is now to find the value of z which will produce the desired amount of side lobe reduction. To this end, let r denote the ratio of the response on the axis to the response at the peaks of the side lobes, so that the side lobes are reduced by

$$\text{Side lobe reduction} = 20 \log r \tag{745}$$

For a pre-selected value of r we must evaluate z from the equation

$$r = 16 z^5 - 20 z^3 + 5 z \tag{746}$$

Although this equation can be solved for z by numerical methods, the process becomes increasingly tedious as the number of elements in the array is increased. For side lobe reductions of about 20 db or more a very good approximation is obtained from the formula

$$z \approx \tfrac{1}{2}\left[(2r)^{\frac{1}{n-1}} + (2r)^{-\frac{1}{n-1}}\right] \tag{747}$$

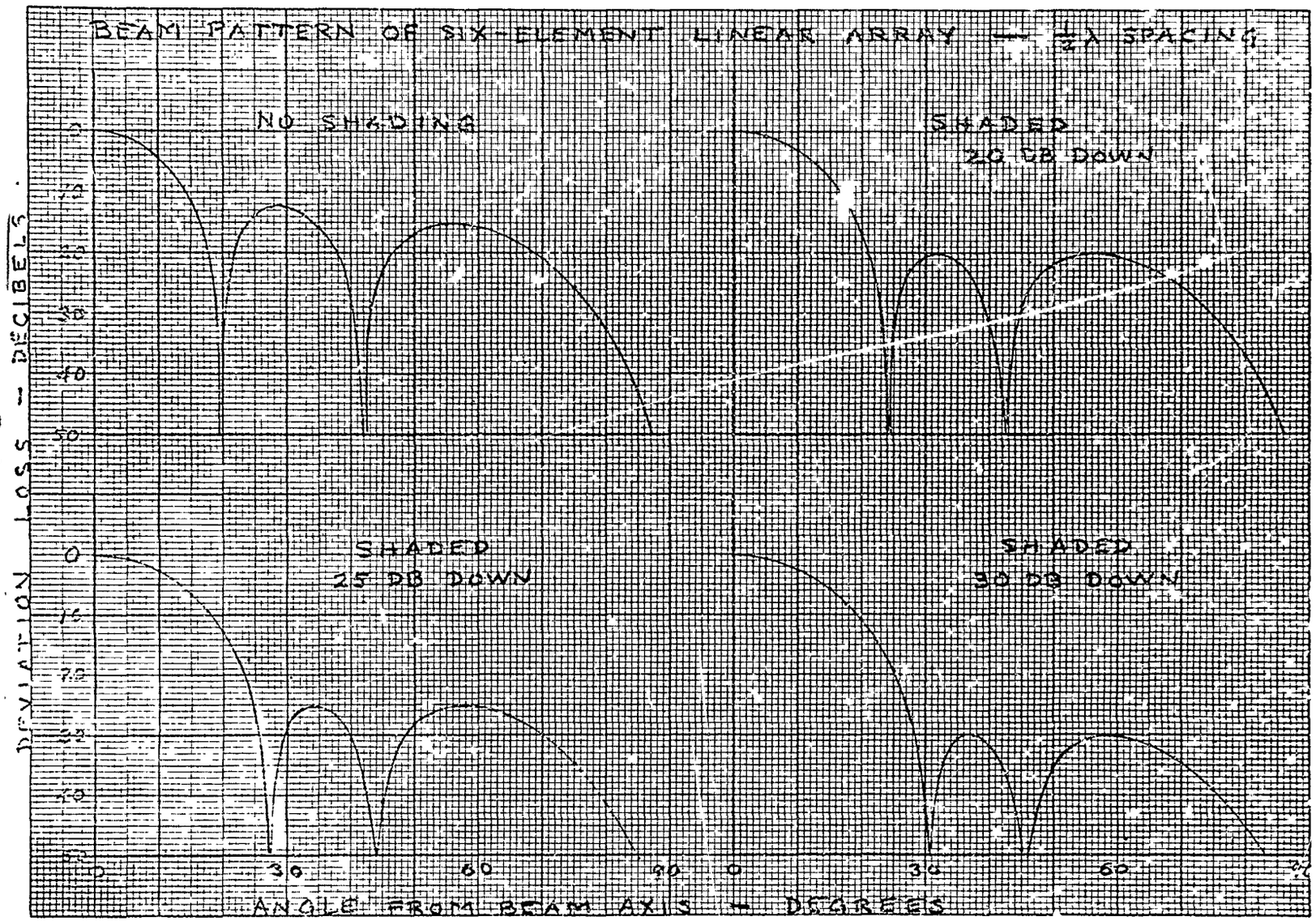
Using this value of z, the coefficients A_1, A_3, and A_5 may be evaluated from (744a), (744b), and (744c), and the beam pattern may then be computed from (741).

To illustrate the theory with a numerical example, the following table lists the numerical values of r, z, and the coefficients A_1, A_3, and A_5 for 6-element linear shaded arrays whose side lobes are 15, 20, 25, and 30 db down relative to the main lobe.

CONSTANTS OF A 6-ELEMENT SHADED MULTISPOT ARRAY

Db Down	r	z	A_1	A_3	A_5
15	5.623	1.1186	2.113	1.759	1.752
20	10.000	1.1846	4.315	3.352	2.333
25	17.783	1.2660	8.415	6.115	3.252
30	31.623	1.3641	15.976	10.923	4.273

It will be noted that as the side lobe reduction is increased, relatively more weight is given to the central elements A_1 and less to the outer elements A_5.


BEAM PATTERN OF SIX-ELEMENT LINEAR ARRAY — ½λ SPACING
NO SHADING
SHADED
20 DB DOWN
SHADED
25 DB DOWN
SHADED
30 DB DOWN
DEVIATION LOSS - DECIBELS
0
10
20
30
40
50
60
30
60
30
0
30
60
ANGLE FROM BEAM AXIS - DEGREES

The values of z listed in the above table were calculated from the exact equation (746). The corresponding approximate values from (747) are 1.1192, 1.1849, 1.2661, and 1.3641.

Beam patterns for shaded and unshaded arrays with an element spacing $a = \frac{1}{2}\lambda$ are shown in the accompanying graphs.

It is to be expected that a price must be paid for the side lobe reduction. As may be seen from the graphs, the price is a broadening of the beam. The approximate beamwidths of the four patterns are:

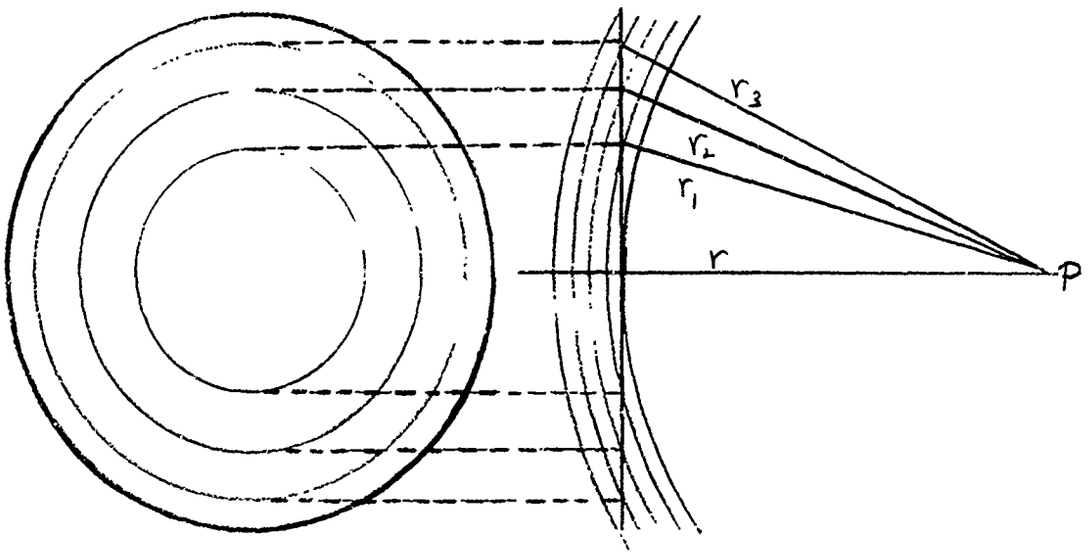
Array	Beamwidth (10 db) (Degrees)
Unshaded	28
Shaded, 20 db	33
Shaded, 25 db	36
Shaded, 30 db	38.5

4. The Near Field

Although a general analysis of the near field of transducers is beyond the scope of these notes, it will be useful to include a brief discussion of the behavior of the pressure along the beam axis in the near field. We have already seen that pressure in the far field varies inversely as the first power of the distance from the transducer. As we move inward along the axis toward the transducer, we find that this law gradually begins to break down. The rate of increase of pressure with diminishing distance begins to slacken off, and in general a number of oscillations are observed in the pressure amplitude. These oscillations are due to interference of portions of the wave from different portions of the transducer surface and are related to Fresnel diffraction phenomena observed in optics.

The usual example discussed in textbooks is the circular piston. This is a rather convenient example because it lends itself readily to a qualitative physical explanation. We shall discuss this example in a rather rough manner and shall then consider a second type of transducer - the continuous line.

We wish to consider the resultant pressure at the point P on the axis at a distance r from the center of the transducer. In particular,



we desire to investigate qualitatively how the pressure varies as the distance r is varied. Let us begin by drawing an arc of a tangent circle of radius r with center at P. We next draw a concentric arc whose radius r_1 is one-half wavelength larger, i.e.,

$$r_1 = r + \tfrac{1}{2}\lambda$$

and continue drawing arcs whose radii increase in steps of $\frac{1}{2}\lambda$

$$r_2 = r_1 + \tfrac{1}{2}\lambda, \text{ etc.}$$

until we reach the edge of the disc. These arcs are traces (in the plane of the paper) of spherical surfaces which intersect the disc in circles as shown above, dividing the disc into annular zones. By virtue of this construction, we see that the phase at P of all the radiation originating within any one zone will have a maximum spread of 180 degrees. Furthermore, the areas of all the zones are approximately equal. This means that the resultant pressure due to two adjacent zones will be approximately zero, since on the average the separate pressures due to the two zones are

approximately 180 degrees out of phase. We may therefore cancel the effects of the zones in pairs, and the net result will be roughly equal to the contribution from the fraction of a zone left over at the edge of the disc.

If, therefore, the distance r has such a value that the disc contains an even integral number of zones, the resultant pressure will be approximately zero. If the distance r is such that the disc contains an odd integral number of zones, the pressure will be a maximum. We see therefore that as r is varied from 0 to infinity, the pressure will exhibit a number of oscillations equal to the number of wavelengths in the radius of the disc. At short distances (small r) the pressure fluctuates rapidly with r, and at longer distances it fluctuates more slowly. If a denotes the radius of the disc, the last zero occurs when r reaches a value such that

$$\sqrt{r^2 + a^2} - r = \lambda$$

or, if $r >> a$,

$$r \approx \frac{a^2}{2\lambda} \tag{748}$$

Beyond this point a gradual transition to the far field occurs.

Consider now a linear transducer of length ℓ. The pressure at a point P at a distance r on the perpendicular bisector of the transducer line is

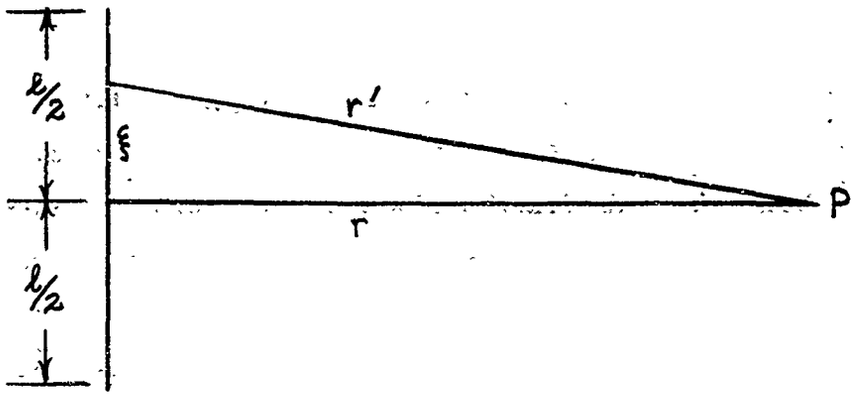
$$p = A \int_{-\frac{\ell}{2}}^{\frac{\ell}{2}} \frac{e^{j(wt-kr')}}{r'} \, d\xi \tag{749}$$

where

A = strength of source per unit length

$r' = \sqrt{r^2 + \xi^2}$ = slant distance from element $d\xi$ to P

ξ = distance from center of line to element $d\xi$



Equation (749) cannot be integrated in closed form, but a reasonable approximation can be obtained for values of r greater than about $\frac{\ell}{2}$ by expanding the radical and neglecting powers of ξ above the second. This approximation yields

$$p = \frac{2Ae^{j(wt-kr)}}{r} \int_0^{\frac{\ell}{2}} \left(1 - \frac{\xi^2}{2r^2}\right) e^{-\frac{jk\xi^2}{2r}} \, d\xi \qquad (749a)$$

where from symmetry considerations the integral over the whole line is replaced by twice the integral over half of the line.

The real part of the pressure is

$$p = \frac{2A}{r}\left[G_c \cos\left(t-kr\right) + G_s \sin\left(t-kr\right)\right] \qquad (749.1)$$

where

$$G_c = \int_0^{\ell/2} \left(1 - \frac{\xi}{2r^2}\right) \cos \frac{k\xi^2}{2r} \, d\xi$$

$$G_s = \int_0^{\ell/2} \left(1 - \frac{\xi^2}{2r^2}\right) \sin \frac{k\xi^2}{2r} \, d\xi$$

The amplitude of the real pressure is

$$p_m = \frac{2A}{r}\sqrt{G_c^2 + G_s^2} \qquad (749.2)$$

By substitution of

$$v = \xi\sqrt{\frac{k}{\pi r}} = \xi\sqrt{\frac{2}{\lambda r}} \tag{749.3}$$

and integration by parts, the integrals G_c and G_s may be expressed in terms of the standard Fresnel integrals

$$C(v) = \int_0^v \cos\frac{\pi u^2}{2}\,du \tag{750}$$

$$S(v) = \int_0^v \sin\frac{\pi u^2}{2}\,du \tag{750a}$$

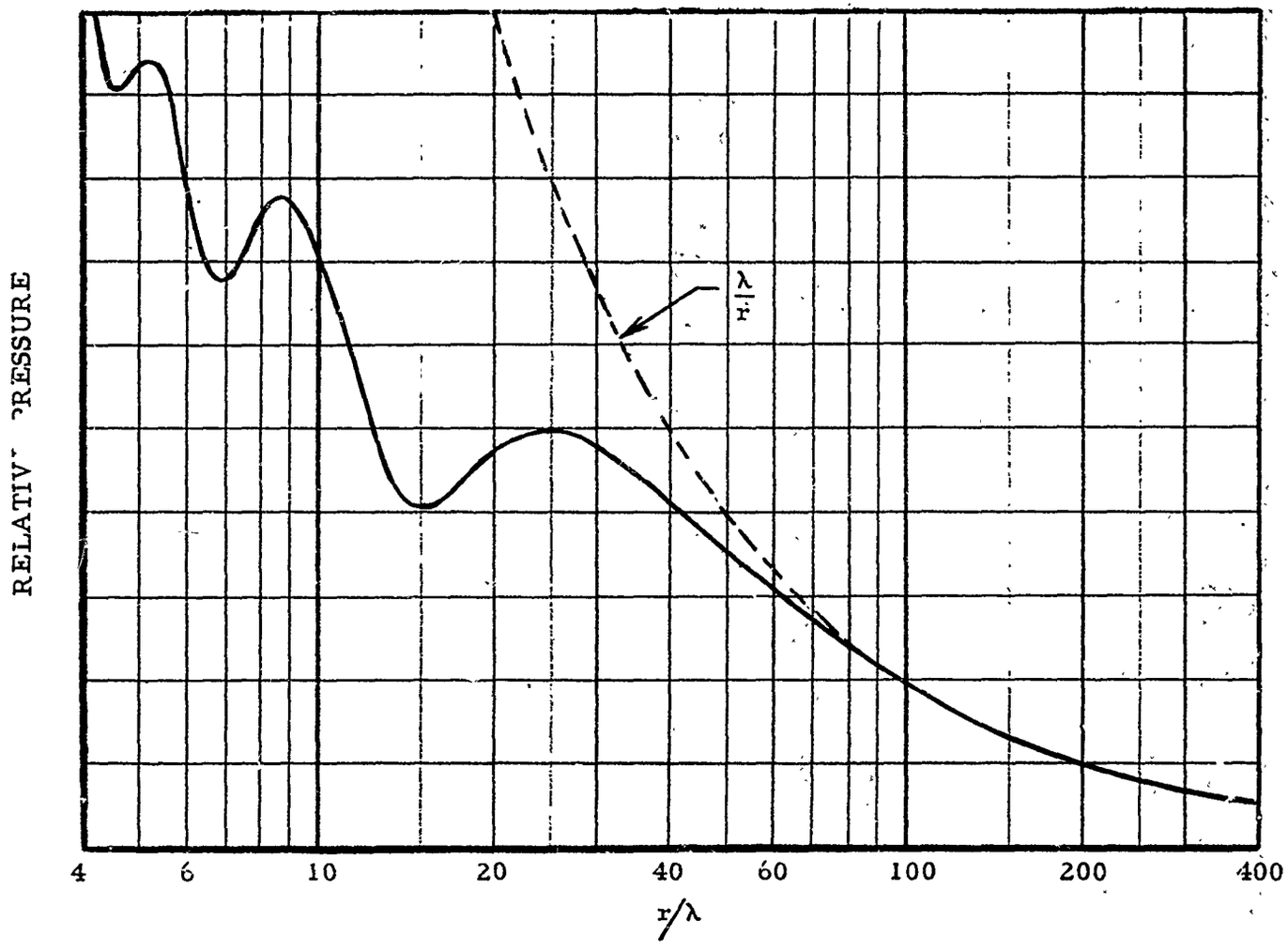
The results are

$$G_c = C\left(\frac{\ell}{\sqrt{2\lambda r}}\right) + \frac{\lambda}{4\pi r} S\left(\frac{\ell}{\sqrt{2\lambda r}}\right) - \frac{\lambda}{4\pi r}\frac{\ell}{\sqrt{2\lambda r}} \cos\frac{\pi \ell^2}{4\lambda r} \tag{750.1}$$

$$G_s = S\left(\frac{\ell}{\sqrt{2\lambda r}}\right) - \frac{\lambda}{4\pi r} C\left(\frac{\ell}{\sqrt{2\lambda r}}\right) + \frac{\lambda}{4\pi r}\frac{\ell}{\sqrt{2\lambda r}} \sin\frac{\pi \ell^2}{4\lambda r} \tag{750.2}$$

NEAR-FIELD PRESSURE ALONG THE AXIS OF A CONTINUOUS LINE TRANSDUCER

$\ell = 10\lambda$



270b

The accompanying graph is a plot of the pressure amplitude (in arbitrary units) vs. distance along the beam axis (expressed in wavelength units) for a transducer which is 10 wavelengths long, that is,

$$\ell = 10\lambda$$

The approximation we have made begins to break down at ranges less than 7 or 8 wavelengths and cannot be used below about 4 or 5 wavelengths. The fluctuations in the pressure field of the linear transducer are considerably smaller than those of a circular piston. The reason for this is that instead of annular zones we now have linear segments, and these segments are not of equal length. Hence only partial cancellation occurs at the minima.

The dashed curve shows the inverse first power law, which applies in the far field. This curve represents the pressure which would be produced by an omnidirectional source having the same far-field pressure distribution as the linear transducer (along its beam axis). The divergence between the two curves is a measure of the near-field effect. Because of the continuous nature of the transition from near to far field, any decision as to where the far field begins must be somewhat arbitrary. A reasonable boundary between the fields for the 10λ line would appear to be about 100 wavelengths. At this distance, the path length from the center of the line is 100λ and the path length from either end of the line is $(100\lambda)^2 + (5\lambda)^2$. The difference in path length is

$$\sqrt{(100\lambda)^2 + (5\lambda)^2} - 100\lambda \approx \frac{1}{8}\lambda$$

As a rough rule of thumb, the far field may be said to begin at a range such that maximum spread of path lengths from points on the transducer line to a point on the beam axis is about 1/8 of a wavelength. Applying this criterion to linear transducers in general, we obtain the following result for the distance r to the far field

$$\sqrt{r^2 + (\frac{\ell}{2})^2} - r = \frac{\lambda}{8}$$

$$r = \frac{\ell^2 - \frac{\lambda^2}{16}}{\lambda}$$

$$r \approx \frac{\ell^2}{\lambda} \qquad (751)$$

where ℓ is the length of the transducer. Equation (751) may, in fact, be applied to a transducer of any shape, provided ℓ is interpreted as the maximum dimension of the transducer.

These considerations are important in transducer calibration procedures, where measurements should be taken in the far field. We see that the required separation distance increases as the square of the maximum linear dimension of the transducer being calibrated. The requirements are thus far more severe for highly directive transducers.

5. Directivity Factor; Directivity Index

In the previous discussion of beam patterns we have considered in some detail the response of transducers in directions away from the beam axis. It should be noted, however, that the off-axis performance is of interest largely from a negative point of view. That is to say, the chief purpose of designing directivity into a projector is to concentrate the energy into a beam so that a larger fraction of the input power is usefully employed for detecting targets and less is lost in stray directions where it generates unwanted reverberation. Likewise, in a receiving array, directivity serves to reject noise and other interference arriving from directions other than the target direction, thereby enhancing the signal to noise ratio. Therefore, to evaluate the performance of a projector against targets located in the vicinity of the beam axis it is useful to be able to summarize the directive effect of the beam pattern in a single number which measures the gain of the system relative to an omnidirectional projector. Similarly, it is useful to summarize the gain of a directional receiving array relative to an omnidirectional hydrophone. The number used for this purpose is called the directivity factor and its decibel

equivalent is called the directivity index. For a projector these quantities are called the transmitting directivity factor and transmitting directivity index; for a hydrophone array they are called the receiving directivity factor and receiving directivity index. In the analysis which follows we shall derive formulas for the transmitting directivity factor and directivity index of a number of different transducer types and shall then discuss briefly their applicability to receiving arrays.

It is clear that a directional projector, by reducing the amount of radiation transmitted in off-axis directions, tends to concentrate its radiated power into a beam. Therefore, a directional transducer will generate a higher intensity at any given distance along its beam axis than an omnidirectional transducer operating at the same power level. Or, stated in another way, to produce the same acoustic intensity at any point on the beam axis, the total acoustic power radiated by a directional transducer is less than that which must be radiated by an omnidirectional transducer. The ratio of these amounts of power is the gain of the directional transducer. The directivity factor is sometimes defined as this gain, (a number greater than 1) and sometimes as the reciprocal of the gain (a number less than 1). We shall use the latter definition. Let us define an "equivalent omnidirectional transducer" as an omnidirectional transducer which generates the same far-field acoustic intensity as the directional transducer generates along its beam axis. Then the directivity factor d is defined as follows:

$$d = \frac{\text{Total acoustic power of directional transducer}}{\text{Total acoustic power of equivalent omnidirectional transducer}} \qquad (752)$$

and the directivity index D is

$$D = 10 \log \frac{1}{d} = -10 \log d \qquad (753)$$

The distance at which the intensity is measured is immaterial, so long as we use the same distance for both transducers. We may therefore use any distance r, it being assumed, of course, that the far-field directivity pattern applies. If the intensity at this distance is I_o, then for the omnidirectional transducer the intensity is constant over the entire surface of

a sphere of radius r surrounding the transducer, and the power is

$$P_o = 4\pi r^2 I_o \tag{754}$$

For the directional transducer the intensity varies over the sphere in accordance with the beam pattern, being proportional to the relative transmitting (intensity) response, which is equal to the square of the relative pressure response. The total power is therefore

$$P = \int\int \left[\frac{p(\theta, \phi)}{p_o}\right]^2 I_o \, dS \tag{755}$$

where dS is an element of area, and the integration is carried out over the entire surface of the sphere surrounding the transducer.

In two dimensions the angle θ in radians subtended by an arc of length s of a circle of radius r is

$$\theta = \frac{s}{r} \text{ (radians)}$$

Similarly in three dimensions the solid angle Ω subtended by an area S of a sphere of radius r is

$$\Omega = \frac{S}{r^2} \text{ (steradians)}$$

We may therefore replace the element dS by

$$dS = r^2 \, d\Omega$$

Inserting this into (755) we see that the directivity factor (752) is

$$d = \frac{1}{4\pi} \int\int \frac{p(\theta, \phi)}{p_o}^2 \, d\Omega \tag{756}$$

a. Two-Spot Array

In the coordinate system we have used for linear arrays the angle θ is measured from the equatorial plane and ranges in value from $-\pi/2$ to $+\pi/2$. The element of solid angle is

$$d\Omega = \cos\theta \, d\theta \, d\phi \tag{757}$$

Inserting the pressure response function (712) for a two-spot array, we obtain for the directivity factor

$$d = \frac{1}{4\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_{0}^{2\pi} \cos^2\psi \cos\theta \, d\theta \, d\phi \tag{758}$$

where $$\psi = \frac{\pi a}{\lambda} \sin\theta$$

or $$\cos\theta d\theta = \frac{\lambda}{\pi a} d\psi$$

Since the integrand is independent of ϕ and is symmetric in θ, we may integrate immediately over ϕ and replace the θ-integral with twice the integral from 0 to $\pi/2$. Changing variables from θ to ψ, we obtain

$$d = \frac{\lambda}{\pi a} \int_{0}^{\frac{\pi a}{\lambda}} \cos^2\psi \, d\psi$$

$$= \frac{\lambda}{2\pi a} \int_{0}^{\frac{\pi a}{\lambda}} (1 + \cos 2\psi) \, d\psi$$

or

$$d = \tfrac{1}{2}\left(1 + \frac{\sin\frac{2\pi a}{\lambda}}{\frac{2\pi a}{\lambda}}\right) \tag{759}$$

As is to be expected, the directivity factor is a function of the spacing, a, of the two elements. If the spacing is very small, that is, if

$$a << \lambda$$

the "sinc" function $(\frac{\sin x}{x})$ is approximately 1, so that $d \approx 1$ and $D \approx 0$ db, and the array behaves as an omnidirectional transducer.

As the spacing is increased, the sinc function decreases in value and the directivity index begins to build up. At a spacing of one half wavelength, where

$$\sin \frac{2\pi a}{\lambda} = \sin \pi = 0$$

we see that $d = \frac{1}{2}$ and D = 3 db. As the spacing is further increased, the directivity index reaches a maximum value of 4.1 db at

$$\frac{2\pi a}{\lambda} = 4.493 \quad \text{or} \quad a = 0.715\,\lambda$$

Beyond this point the directivity index a curve of D vs. a shows a series of damped oscillations about its average value of 3 db.

b. Multi-Spot Array

The most useful form of the relative pressure response of a multi-element linear array for the computation of the directivity factor is the exponential form (715). As noted previously, this function has a slightly different appearance, depending upon whether the number of elements is odd or even. Equation (715) assumes an even number. At the outset we note that the integrand is independent of ϕ and is symmetric in θ, and that it is expeditious to change the variable from θ to ψ. When these operations are carried out, we obtain

$$d = \frac{\lambda}{\pi a n^2} \int_0^{\frac{\pi a}{\lambda}} \left[e^{(n-1)j\psi} + e^{(n-3)j\psi} + \ldots + e^{j\psi} + e^{-j\psi} + \ldots + e^{-(n-1)j\psi} \right]^2 d\psi ,$$

$$d = \frac{\lambda}{\pi a n^2} \int_0^{\frac{\pi a}{\lambda}} \left[e^{2(n-1)j\psi} + 2e^{2(n-2)j\psi} + \ldots + (n-m)e^{2mj\psi} + \ldots + (n-1)e^{2j\psi} + n + (n-1)e^{-2j\psi} + \ldots + (n-m)e^{-2mj\psi} + \ldots + e^{-2(n-1)j\psi} \right] d\psi$$

$$d = \frac{\lambda}{\pi a n^2} \left[\frac{e^{2(n-1)j\psi} - e^{-2(n-1)j\psi}}{2(n-1)j} + 2\frac{e^{2(n-2)j\psi} - e^{-2(n-2)j\psi}}{2(n-2)j} + \ldots + (n-m)\frac{e^{2mj\psi} - e^{-2mj\psi}}{2mj} + \ldots + (n-1)\frac{e^{2j\psi} - e^{-2j\psi}}{2j} + n\psi \right]_0^{\frac{\pi a}{\lambda}}$$

This expression may be written in the following compact form

$$d = \frac{1}{n} + \frac{2}{n^2} \sum_{m=1}^{n-1} \left[(n-m) \frac{\sin \frac{2\pi ma}{\lambda}}{\frac{2\pi ma}{\lambda}} \right] \tag{760}$$

It is readily seen that this formula boils down to (759) when n = 2.

The same general comments apply to the directivity factor of the multispot array as were made for the two-element array. If a is very small compared with λ, we see that d is approximately

$$d = \frac{1}{n} + \frac{2}{n^2} \sum_{m=1}^{n-1} (n - m)$$

$$= \frac{1}{n} + \frac{2}{n^2} \frac{n(n-1)}{2} = 1$$

and the array acts as an omnidirectional transducer. Whenever the spacing between elements is an integral number of half wavelengths, all the sinc terms are zero and the directivity index is

$$D = 10 \log n \tag{761}$$

As the spacing is increased, the directivity index executes a series of damped oscillations about the value given by (761).

c. Continuous Line Transducer

The relative pressure response for a linear transducer is given by (721). Insertion of this function into (756) gives

$$d = \frac{1}{4\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_{0}^{2\pi} \left(\frac{\sin\psi}{\psi} \right)^2 \cos\theta \, d\theta \, d\phi$$

$$= \frac{\lambda}{\pi \ell} \int_{0}^{\frac{\pi \ell}{\lambda}} \left(\frac{\sin\psi}{\psi} \right)^2 d\psi$$

This integral may be reduced by integration by parts, yielding

$$d = \frac{\mathrm{Si}\left(\frac{2\pi\ell}{\lambda}\right)}{\frac{\pi\ell}{\lambda}} - \left(\frac{\sin\frac{\pi\ell}{\lambda}}{\frac{\pi\ell}{\lambda}}\right)^2 \tag{761}$$

where Si(x) is the sine integral

$$\mathrm{Si}(x) = \int_0^x \frac{\sin\xi}{\xi}\, d\xi \tag{762}$$

If the length ℓ of the array is greater than about one wavelength, the sine integral is approximately

$$\mathrm{Si}(x) \approx \frac{\pi}{2} - \frac{\cos x}{x} \tag{763}$$

so that the directivity factor is approximately

$$d = \frac{\lambda}{\pi\ell}\left(\frac{\pi}{2} - \frac{\lambda}{2\pi\ell}\right) \tag{764}$$

This formula is excellent for most linear transducers of practical interest.

For rough calculations a still simpler formula may be obtained by neglecting the second term of (764), yielding

$$d \approx \frac{\lambda}{2\ell} \tag{764a}$$

and

$$D \approx 10 \log \frac{2\ell}{\lambda} \tag{765}$$

For transducer lengths greater than a wavelength the error in decibels introduced into the directivity index by this last approximation is approximately $\frac{0.4\lambda}{\ell}$.

d. Circular Piston

The circular piston transducer differs from the linear types discussed above in that its beam axis is a single line normal to the piston face. As noted previously, it is desirable for this type of transducer to measure the angle θ from the pole instead of from the equator, so that the element of solid angle is

$$d\Omega = \sin\theta\, d\theta\, d\phi \tag{766}$$

In designing transducers of this type it is usually desirable to concentrate the radiation in the forward direction ($\theta = 0$) by providing a baffle in the rearward direction ($\theta = 180°$). In practical sonar transducers it is impossible to prevent the side lobes from spreading into the rear hemisphere, although the desired condition can be approximated in large arrays. In the present discussion we shall make the assumption of an ideal infinite baffle which confines the entire sound field to the forward hemisphere. Under this assumption the general formula for the directivity factor of an area-type transducer is

$$d = \frac{1}{4\pi} \int_0^{\frac{\pi}{2}} \int_0^{2\pi} \left[\frac{p(\theta, \phi)}{p_o}\right]^2 \sin\theta d\theta d\phi \tag{767}$$

Substituting (728b) for the circular piston, we obtain

$$d = \tfrac{1}{2} \int_0^{\frac{\pi}{2}} \left[\frac{2J_1(\psi)}{\psi}\right]^2 \sin\theta d\theta d\phi \tag{768}$$

where $$\psi = \frac{2\pi a}{\lambda} \sin\theta$$

This integral has been evaluated, and the result is

$$d = \left(\frac{\lambda}{2\pi a}\right)^2 \left[1 - \frac{J_1\left(\frac{2\pi a}{\lambda}\right)}{\frac{2\pi a}{\lambda}}\right] \tag{769}$$

If the radius a is larger than about half a wavelength, the second term in (769) is relatively small and the directivity factor is approximately

$$d \approx \left(\frac{\lambda}{2\pi a}\right)^2 \tag{770}$$

It will be noted that since the area of the piston face is

$$A = \pi a^2$$

the directivity factor may be expressed in the form

$$d = \frac{\lambda^2}{4\pi A} \tag{771}$$

e. Rectangular Plate

The directivity factor of a rectangular plate of dimensions ℓ and w is (assuming again an infinite baffle)

$$d = \frac{1}{4\pi} \int_0^{\pi/2} \int_0^{2\pi} \left(\frac{\sin \psi_\ell}{\psi_\ell} \frac{\sin \psi_w}{\psi_w} \right)^2 \sin \theta d\theta d\phi \qquad (772)$$

where

$$\psi_\ell = \frac{\pi \ell}{\lambda} \sin \theta \cos \phi$$

$$\psi_w = \frac{\pi w}{\ell} \sin \theta \sin \phi$$

This integral does not lend itself readily to evaluation. However, we can obtain an approximate solution for the special case where the dimensions ℓ and w are large compared with a wavelength. Let us change variables as follows

$$x = \sin \theta \cos \phi$$

$$y = \sin \theta \sin \phi$$

If ℓ/λ and w/λ are large, then both sinc functions in (772) will drop to small values before θ becomes very large. Since the integrand becomes extremely small at large values of θ, we may replace $\pi/2$ with ∞ in the upper limit of integration. The corresponding limits of integration for x and y are $-\infty$ to $+\infty$. The element of area is

$$dx\, dy = \left| \begin{matrix} \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} \\ \frac{\partial x}{\partial \phi} & \frac{\partial y}{\partial \phi} \end{matrix} \right| d\theta\, d\phi$$

$$= \sin \theta \cos \theta d\theta d\phi$$

Since we are interested only in small values of θ, we may replace cos θ by 1. Substitutions of these approximations into (772) yields

$$d = \frac{1}{4\pi} \int_{-\infty}^{\infty} \left(\frac{\sin \frac{\pi \ell x}{\lambda}}{\frac{\pi \ell x}{\lambda}} \right)^2 dx \int_{-\infty}^{\infty} \left(\frac{\sin \frac{\pi w y}{\lambda}}{\frac{\pi w y}{\lambda}} \right)^2 dy$$

which integrates to

$$d = \frac{\lambda^2}{4\pi \ell w} \tag{773}$$

Note that the product ℓw is the area A of the transducer face, so that once again we find

$$d = \frac{\lambda^2}{4\pi A} \tag{771}$$

It has been found that (771) is a good approximation to area-type transducers of <u>any shape,</u> provided none of the linear dimensions is less than about two wavelengths.

We noted earlier that the beam pattern of a rectangular transducer is related to the patterns of two linear transducers of lengths ℓ and w, at right angles to each other. It is of interest to compare the directivity index of the rectangular transducer with those of the lines. Let $D_{\ell w}$ refer to the rectangular plate and D_ℓ and D_w refer to two lines. Then, approximately,

$$D_{\ell w} = 10 \log \left(\frac{4\pi \ell w}{\lambda^2} \right)$$

$$D_\ell = 10 \log \left(\frac{2\ell}{\lambda} \right)$$

$$D_w = 10 \log \left(\frac{2w}{\lambda} \right)$$

Therefore

$$D_{\ell w} = D_\ell + D_w + 10 \log \pi \tag{774}$$

or

$$D_{\ell w} = D_\ell + D_w + 5 \text{ db} \tag{774a}$$

The directivity index of the plate is about 5 db larger than the sum of the directivity indices of the crossed lines.

f. Effect of Shading

The reduction in side lobes due to shading would be expected to increase the directivity index. This effect, however, is more than compensated by the broadening of the beam. To investigate this effect, let us go back to the earlier example of a 6-element linear array with Chebyshev polynomial shading. Going back to (741), let

$$a_i = \frac{A_i}{A_1 + A_3 + A_5}, \quad i = 1, 3, 5 \tag{775}$$

Integration of the complex exponential equivalent of (741) in lieu of the cosine form yields

$$d = \tfrac{1}{2}\left[a_1^2 + a_3^2 + a_5^2 + (a_1^2 + 2a_1 a_3 + 2a_3 a_5)\frac{\sin\frac{2\pi a}{\lambda}}{\frac{2\pi a}{\lambda}}\right.$$

$$+ (2a_1 a_3 + 2a_1 a_5)\frac{\sin\frac{4\pi a}{\lambda}}{\frac{4\pi a}{\lambda}} + (a_3^2 + 2a_1 a_5)\frac{\sin\frac{6\pi a}{\lambda}}{\frac{6\pi a}{\lambda}} + 2a_3 a_5 \frac{\sin\frac{8\pi a}{\lambda}}{\frac{8\pi a}{\lambda}}$$

$$\left. + a_5^2 \frac{\sin\frac{10\pi a}{\lambda}}{\frac{10\pi a}{\lambda}}\right] \tag{776}$$

In our previous numerical example we selected a spacing of one half wavelength

$$a = \frac{\lambda}{2}$$

For such a spacing, all the sinc terms of (776) vanish, leaving

$$d = \tfrac{1}{2}(a_1^2 + a_3^2 + a_5^2)$$

Numerical calculations for the cases discussed previously yield the following values of directivity index

Array	D db
Unshaded	7.78
Shaded, 15 db	7.75
Shaded, 20 db	7.53
Shaded, 25 db	7.26
Shaded, 30 db	7.02

6. Noise Rejection by Directional Hydrophones and Arrays

At the outset we must distinguish between two different types of noise - (1) noise originating from a single discrete source, and (2) general ambient noise present in the sea. Noise of the first type arrives at the hydrophone as a plane wave traveling in a specific direction, or at most a small angular spread of directions. The sensitivity of the system to noise of this type is determined by the deviation loss in the direction from which the sound is coming. For example, in the case of a helicopter dipped sonar the noise generated by the vehicle arrives at the transducer from a small angular region about the vertical. The effect of this noise in interfering with horizontally arriving signals is determined by the deviation loss of the transducer in the vertical direction.

Our primary concern in this section is with the general ambient noise present in the ocean. The bulk of this noise originates from a very large number of sources distributed throughout the surface of the ocean and the near-surface region. For this reason the noise arrives at the receiving transducer not from any one direction but from a continuous distribution of directions. In describing the noise intensity (or, more accurately, intensity per unit band), it is useful to introduce the concept of intensity per unit solid angle. Because of the continuous distribution, the amount of energy traveling in any discrete direction is zero; the amount traveling in cone of solid angle $d\Omega$ is proportional to $d\Omega$.

Let $J_\Omega(\theta, \phi)$ represent the intensity per unit band per unit solid angle arriving from the direction (θ, ϕ). Then the intensity per unit band

arriving from a small cone of solid angle $d\Omega$ is $J_\Omega(\theta, \phi)\, d\Omega$. The total noise intensity per unit band received by an omnidirectional hydrophone is

$$J_o = \iint J_\Omega(\theta, \phi)\, d\Omega \tag{777}$$

where the integration is carried over the entire sphere of unit radius surrounding the hydrophone. In the special case of an isotropic noise field, where $J_\Omega(\theta, \phi)$ is a constant independent of θ and ϕ, we see that

$$J_o = \iint J_\Omega\, d_\Omega = 4\pi\, J_\Omega$$

or

$$J_\Omega = \frac{1}{4\pi}\, J_o \tag{778}$$

Thus for an isotropic field the intensity per unit band per unit solid angle (that is, per steradian) is equal to $\frac{1}{4\pi}$ times the overall intensity per unit band measured by an omnidirectional hydrophone.

If now we consider a directional receiving array, we see that the effect of the noise arriving from the small cone of solid angle $d\Omega$ in the direction (θ, ϕ) is modified by the relative receiving response of the array in that direction. The effective noise per unit band sensed by the array is thus

$$J_{eff} = \iint \left[\frac{e(\theta, \phi)}{e_o}\right]^2 J_\Omega(\theta, \phi)\, d\Omega, \tag{779}$$

where, as before, the integration is carried out over the entire sphere. The ratio of the total noise as sensed by the directional array to the total noise as sensed by an omnidirectional hydrophone is

$$\frac{J_{eff}}{J_o} = \frac{\iint \left[\frac{e(\theta, \phi)}{e_o}\right]^2 J_\Omega(\theta, \phi)\, d\Omega}{\iint J_\Omega(\theta, \phi)\, d\Omega} \tag{780}$$

From (780) it is apparent that the gain in noise rejection by a directional transducer relative to an omnidirectional transducer is dependent upon the directional properties of the ambient noise field. In the special case of an isotropic noise field, where $J_\Omega(\theta, \phi)$ is constant, equation (780) reduces to

$$\frac{J_{eff}}{J_o} = \frac{1}{4\pi} \iint \left[\frac{e(\theta, \phi)}{e_o}\right]^2 d\Omega \tag{781}$$

which, when compared with (756) is seen to be the receiving directivity factor of the array. In other words, the directivity factor and directivity index apply to a receiving array only when the array is operating in an isotropic noise field. If the field is isotropic, the effective noise intensity per unit band sensed by the directional array is less by a factor d than the intensity per unit band sensed by an omnidirectional hydrophone

$$J_{eff} = d\, J_o \tag{782}$$

and the corresponding effective spectrum level of the noise is

$$N_{eff} = N_o - D \tag{783}$$

where N_o is the spectrum level corresponding to J_o.

If the noise field is not isotropic, equation (783) is not valid, and the effective noise sensed by the transducer should be computed directly from (779). Noise is normally expressed in decibels. Ideally a directional field should be described in terms of spectrum level per unit solid angle, which we shall designate as

$$N_\Omega(\theta, \phi) \text{ in db//}\mu b^2/\text{cps steradian}$$

To evaluate the effect of the noise on a directional transducer, we may express $J_\Omega(\theta, \phi)$ in (pressure)^2 units by the relation

$$J_\Omega(\theta, \phi) = 10^{0.1 N_\Omega(\theta, \phi)} \tag{784}$$

Values of $J_\Omega(\theta, \phi)$ are then substituted into (779). The integration must usually be carried out by numerical methods. The effective noise level is

$$N_{eff} = 10 \log J_{eff} \tag{785}$$

The experimental determination of the intensity per steradian in a directional noise field is not a simple matter. If an ideal hydrophone array could be built which had a very narrow beam covering a small solid angle $\Delta\Omega$, without side lobes, then the intensity per unit band per steradian could be obtained by dividing the observed intensity per unit band by $\Delta\Omega$

$$J_\Omega = \frac{J}{\Delta\Omega}$$

In practice, however, the beam does not have a sharply defined cut-off, and side lobes are always present. The analytical process of estimating the true field from measurements by practical transducer arrays is complicated and will not be discussed in these notes. Considerable work of this type has been done in the field of radio astronomy.

Measurements made at great depths in the ocean have revealed a considerable angular dependence of ambient sea noise in a vertical plane. The highest intensity is received from the vertical direction. The intensity received horizontally many be as much as 18 db down, depending upon the sea state and the frequency. The directional effect increases with increasing sea state. A transducer array having a horizontal beam will experience a lower ambient noise background when operating at great depth than when operating near the surface.

7. Reverberation Factor: Reverberation Index

Although the subject of reverberation will not be discussed until later, a brief account of the effect of the directional characteristics of transducers upon the received reverberation will be given in this section. We shall restrict our attention to the ideal case of volume reverberation in an infinite homogeneous ocean. It is assumed that when a ping is sent out by the transducer, every cubic yard of water scatters back to the transducer a constant fraction of the incident acoustic radiation. Considered as a function of the direction in space relative to the beam axis, the intensity of the incident radiation over a sphere surrounding the transducer is proportional to the relative transmitting response $\eta(\theta, \phi)$. Since the returning reverberation comes in at the same angle as the outgoing radiation which caused it, the effect on the receiving array is proportional to the relative receiving response $\eta'(\theta, \phi)$. Thus, the overall effect of the reverberation is proportional to the product of these two functions The reverberation factor is the integral

of this combined relative response over the sphere,

$$d_R = \frac{1}{4\pi} \iint \eta(\theta, \phi)\, \eta'(\theta, \phi)\, d\Omega \tag{786}$$

The reverberation index is the decibel equivalent of d_R,

$$D_R = 10 \log \frac{1}{d_R} \tag{787}$$

If we assume that the transducer is reversible, then

$$\eta(\theta, \phi)\, \eta'(\theta, \phi) = \eta^2(\theta, \phi) = \left[\frac{p(\theta, \phi)}{p_o}\right]^4$$

and

$$d_R = \frac{1}{4\pi} \iint \left[\frac{p(\theta, \phi)}{p_o}\right]^4 d\Omega \tag{788}$$

We shall apply this result to two representative cases, the continuous line and the rectangular plate. For the continuous line we obtain

$$d_R = \frac{1}{4\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_0^{2\pi} \left(\frac{\sin \psi}{\psi}\right)^4 \cos\theta\, d\theta\, d\phi$$

$$= \frac{\lambda}{\pi \ell} \int_0^{\frac{\pi \ell}{\lambda}} \left(\frac{\sin \psi}{\psi}\right)^4 d\psi \tag{789}$$

Successive integration by parts leads to the expression

$$d_R = \frac{\lambda}{3\pi \ell} \left[4\, \mathrm{Si}\left(\frac{4\pi \ell}{\lambda}\right) - 2\, \mathrm{Si}\left(\frac{2\pi \ell}{\lambda}\right) - \frac{\lambda}{\pi \ell}\left(6 \sin^2 \frac{\pi \ell}{\lambda} - 8 \sin^4 \frac{\pi \ell}{\lambda}\right) \right.$$

$$\left. - 2 \left(\frac{\lambda}{\pi \ell}\right)^2 \sin^3 \frac{\pi \ell}{\lambda} \cos \frac{\pi \ell}{\lambda} - \left(\frac{\lambda}{\pi \ell}\right)^3 \sin^4 \frac{\pi \ell}{\lambda} \right] \tag{789a}$$

When the transducer is more than one or two wavelengths long this formula may be approximated by

$$d_R \approx \frac{\lambda}{3\ell} \tag{789b}$$

The approximate reverberation index is

$$D_R \approx 10 \log \frac{3\ell}{\lambda}$$

$$= 10 \log \frac{2\ell}{\lambda} + 10 \log \frac{3}{2}$$

$$\approx D + 1.76 \text{ db} \qquad (790)$$

Thus the reverberation index is larger than the directivity index by about 1.8 db.

A similar approximation may be obtained for a rectangular plate, the result being

$$d_R \approx \frac{\lambda^2}{9\pi A} \qquad (791)$$

$$D_R \approx 10 \log \frac{\lambda^2}{4\pi A} + 10 \log \frac{9}{4}$$

$$= D + 3.52 \text{ db} \qquad (792)$$

As a rough rule of thumb the reverberation index is about 1.8 db larger than the directivity index for transducers whose beams are spread cut over the equatorial plane (for example, linear transducers) and about 3.5 db larger for transducers with a searchlight pattern (for example, circular piston and rectangular plate).

8. Capture Area

When a transducer is used as a hydrophone the input power is obtained from the acoustic radiation. The maximum power which can be obtained from the radiation is called the available power. It is obtained when the waves are coming in the direction of the maximum response axis. Since the acoustic intensity is a measure of the power per unit area, the available power must be equal to the product of the intensity by the effective area of the hydrophone, which is called the capture area, A_c.

For a hydrophone which is in the form of a plate whose dimensions are large compared to the wavelength of sound, the capture area is approximately equal to the actual area of the plate. This area is related to the directivity factor and the wavelength by equation (771). Thus

$$A_c = \frac{\lambda^2}{4\pi d} = \frac{c^2}{4\pi f^2 d} \tag{793}$$

It has been found from radio antenna theory that equation (793) yields approximately the correct value of the capture area for all types of antennas; it likewise yields approximately the correct value of the capture area for all types of acoustic hydrophones. The capture area is equal to the actual area only for hydrophones having a relatively large area. For other types this is not the case.

Let us consider two examples. First, a linear hydrophone. The directivity factor, according to (764a), is approximately

$$d = \frac{\lambda}{2\ell}$$

Therefore the capture area is

$$A_c = \frac{\lambda^2}{4\pi} \cdot \frac{2\ell}{\lambda} = \frac{\lambda\ell}{2\pi} = \frac{\ell c}{2\pi f} \tag{794}$$

The capture area of a linear hydrophone is proportional to the length and inversely proportional to the frequency.

Second, an omnidirectional hydrophone. The directivity index is by definition unity.

$$d = 1$$

Therefore the capture area is

$$A_c = \frac{\lambda^2}{4\pi} = \frac{c^2}{4\pi f^2} \tag{795}$$

The capture area is inversely proportional to the square of the frequency and is independent of the actual area of the hydrophone.

TRANSDUCERS (continued)

C. Thermal Noise.

1. Introduction.

In the detection of acoustic signals the sonar operator must always compete with background interference. Some sources of interfering noise can be controlled to varying degrees; others, such as ambient sea noise, depend on meteorological conditions and other accidents of nature. Even though under ideal conditions it is conceivable that all these sources of noise might be reduced to negligible proportions, there is one basic factor that sets a lower limit below which signals cannot be detected. This factor is thermal agitation.

According to the kinetic theory of matter, all the elementary particles of which matter is composed are in a continual state of agitation. This energy of agitation is proportional to the absolute temperature. In a gas each molecule possesses on the average an amount of energy equal to

$$\tfrac{1}{2} n k_o T_K$$

where

n = number of degrees of freedom per molecule

k_o = Boltzmann's constant = 1.371×10^{-16} erg/° Kelvin
= 1.371×10^{-23} joule/ °Kelvin

T_k = absolute temperature in degrees on the Kelvin scale (= 273.2 plus Centigrade temperature)

The simplest molecules are those of the inert gases, like neon, which consist of a single atom. These molecules possess three degrees of freedom by virtue of their translational motion in three dimensions. More complicated molecules may have rotational and vibrational degrees of freedom in addition to the three translational degrees of freedom. The theorem of equipartition of energy states that on the average the energy of a molecule is equally divided among all of its degrees of freedom.

As a result of the pioneering work of J. B. Johnson and H. Nyquist of the Bell Telephone Company, and the developments of statistical theory which have grown out of their work, the concepts of classical statistical mechanics have been extended to broadband random processes in general and thermal noise in particular. If, for example, we consider the noise voltage generated in a resistor by the thermal motions of the electrons of which it is composed, we find that this is a typical broadband random process. Suppose this noise is filtered through a bandpass filter having a bandwidth W. It has been shown that if this voltage, considered as a function of time, is sampled at a rate not exceeding $\frac{1}{2W}$ samples per second, these samples are statistically independent random variables. The maximum number of statistically independent samples which can be obtained in any time t is 2 Wt. Each independent sample may be considered as equivalent to a degree of freedom. By the equipartition theorem, the total thermal energy generated by the electrons in the resistor in a bandwidth W in time t is

$$\tfrac{1}{2}(2Wt)k_o\, T_K = W t k_o\, T_K$$

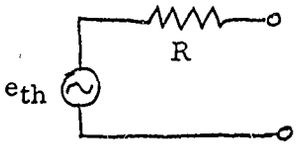
The average thermal power generated in the resistor in the frequency band of width W is the energy per unit time,

$$P_{th} = k_o\, T_K W \tag{801}$$

and the average power per unit band is

$$U_{th} = \frac{P_{th}}{W} = k_o\, T_K \tag{802}$$

To compute the rms noise voltage in the band W, we picture the resistor as being equivalent to an a.c. generator having

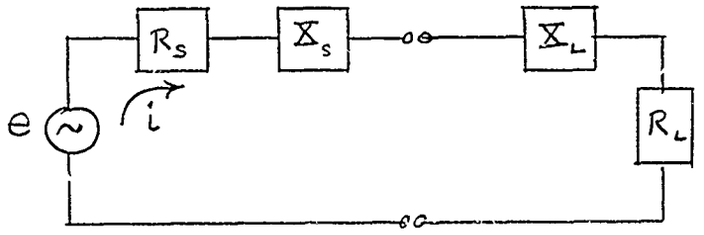


an emf whose rms value is the thermal noise voltage e_{th} and whose internal resistance is the resistance R of the resistor under consideration. The value of e_{th} is such that the available power from this equivalent generator is equal to the thermal power of (801). The available power is the maximum power

which can be drawn when the load impedance is matched to the source impedance. The actual power which is drawn from a power source depends upon both the load impedance and the internal impedance of the source. If the source and load impedances are

$$z_S = R_S + j\,X_S$$

$$z_L = R_L + j\,X_L$$



the rms current is

$$i = \frac{e}{\left[(R_S + R_L)^2 + (X_S + X_L)^2\right]^{1/2}}$$

and the power delivered to the load is

$$P_L = i^2 R_L = \frac{e^2 R_L}{(R_S + R_L)^2 + (X_S + X_L)^2} \qquad (803)$$

If the impedance of the source is fixed, the maximum output power is obtained by matching the load impedance to the source impedance. The optimum values are obtained by taking the derivatives of P_L with respect to R_L and X_L and setting these equal to zero. Thus,

$$\frac{\partial P_L}{\partial R_L} = \frac{e^2}{Z^2} - \frac{2(R_S + R_L)e^2 R_L}{Z^4} = 0$$

$$\frac{\partial P_L}{\partial X_L} = -\frac{2(X_S + X_L)e^2 R_L}{Z^4} = 0$$

where $Z^2 = (R_S + R_L)^2 + (X_S + X_L)^2$

These equations solve to

$$R_L = R_S \qquad (804a)$$

$$X_L = -X_S \qquad (804b)$$

The maximum power is drawn from a generator when the load impedance is the complex conjugate of the source impedance, that is, when the load resistance is equal to the source resistance and the load reactance is equal to

the negative of the source reactance (inductive vs. capacitance, or vice versa). In this case the overall circuit is purely resistive and half of the power generated by the emf is dissipated in the internal resistance and half in the load resistance. Substituting (804a) and (804b) into (803), we see that the maximum load power, or available power from the source, is

$$P_{Avl} = \frac{e^2}{4R_s} \tag{805}$$

In the case of thermal noise in the resistor R, (805) leads to the following

$$P_{th} = k_o T_K W = \frac{e^2_{th}}{4R}$$

or

$$e_{th} = \sqrt{4R k_o T_K W} \tag{806}$$

2. Thermal Equivalent Water Noise.

Thermal agitation affects a hydrophone in two ways: (a) the molecules of water, which are constantly in motion, strike the hydrophone in a random manner and generate a noise voltage; (b) the electrons in the electrical circuit elements of the hydrophone experience thermal agitation and generate noise in the circuit.

The thermal agitation of the water molecules is a basic property of the medium and its effect upon a hydrophone is most conveniently described in terms of a hypothetical equivalent acoustic wave which, when received in the conventional manner by an omnidirectional hydrophone, produces the same electrical output as the molecular motions. The total power per unit band received from the water is the same as that generated by the load resistor in the equivalent circuit of the hydrophone, its value being given by (802). The thermal equivalent intensity per unit band, J_{th}, is that value which, when multiplied by the capture area of the hydrophone, is equal to the thermally generated power per unit band

$$J_{th} A_c = Uth \tag{807}$$

The capture area of an omnidirectional hydrophone is given by (795)

$$A_c = \frac{c^2}{4\pi f^2} \qquad (795)$$

Hence

$$J_{th} = \frac{4\pi f^2 k_o T_K}{c^2} \qquad (808)$$

If the Boltzmann constant is expressed in joules/°K, c in cm/sec, and f in cps, then J_{th} is in watts/cm² cps and the thermal equivalent spectrum level, L_{th} is

$$L_{th} = 10 \log J_{th}$$
$$= 20 \log f - 10 \log \frac{c^2}{4\pi k_o T_K} \text{ db//watt/cm}^2 \text{ cps} \qquad (809)$$

As a numerical example, if the water temperature is 60°F, so that $T_K = 288.7$°K, the thermal equivalent spectrum level is

$$L_{th} = 20 \log f - 296.5 \text{ db//watt/cm}^2 \text{ cps} \qquad (810)$$

or

$$L_{th} = 20 \log f - 174.6 \text{ db//}\mu\text{b}^2\text{/cps} \qquad (810a)$$

The thermal equivalent spectrum level increases with frequency. Ordinary ambient sea noise, caused chiefly by the action of wind and waves, decreases with frequency. In the frequency range above a few hundred cps, ambient sea noise is described by the Knudsen curves which relate the spectrum level to the frequency and the sea state. These curves are based on measurements made during World War II. Although some more recent measurements have shown considerable variability in ambient noise levels, the Knudsen curves are still widely used. Dr. H. O. Benecke, while at the Naval Air Development Center, fit the Knudsen curves with an empirical formula

$$L_{amb} = -54 + 30 \log (n_s + 1) - 17 \log f \text{ db//}\mu\text{b}^2\text{/cps} \qquad (811)$$

where n_s denotes the sea state and f is the frequency in kc/sec. (The frequency in (810) is in cps.) Comparison of (811) with (810a) shows that the ambient noise is equal to the thermal noise at 60°F when

$$37 \log f = 60.6 + 30 \log (n_s + 1) \qquad \text{(f in Kc.)}$$

or, when f has the values listed below

Sea State	Frequency (kc/sec)
0	43
2	105
4	160
6	210

In the range of frequencies of interest in most sonars the thermal noise level is well below the other noise encountered in the sea.

If the transducer has directivity, the effect of the thermal agitation relative to signals received along the maximum response axis is reduced. The general formula for the capture area is (793), which includes the directivity factor d. At any given frequency the capture area of a directional transducer is larger than that of an omnidirectional transducer by a factor d. The thermal equivalent intensity per unit band in this case is

$$J_{th} = \frac{4\pi f^2 k_o T_K d}{c^2} \qquad (812)$$

and the thermal equivalent spectrum level is

$$L_{th} = 20 \log f - 10 \log \frac{c^2}{4\pi k_o T_K} - D \qquad (813)$$

where D is the receiving directivity index of the transducer.

3. Thermal Equivalent Hydrophone Noise.

Thermal noise originating in the electrical circuit components of the hydrophone appears directly as noise at the output terminals. If we wish to express this noise in terms of an equivalent acoustic wave in the water, we must take into account the efficiency of the hydrophone. In the previous discussion we were concerned only with replacing the actual thermal noise in the water with an equivalent noise. Both the actual and equivalent noise sources were located at the mechanical input to the hydrophone and both suffered the same efficiency loss in the transformation from input acoustic to output electrical power. Now, however, the source of the noise is electrical, and we are replacing it with an equivalent acoustic source. In order that the

acoustic power be truly equivalent, we must multiply it by the efficiency before setting it equal to the internally generated noise power. Let η_{AE} denote the acousto-electric efficiency and let N_{AE} be the corresponding loss in decibels

$$N_{AE} = -10 \log \eta_{AE} \tag{814}$$

Then if J'_{th} is the thermal equivalent intensity per unit band corresponding to the internally generated thermal noise, or transducer equivalent intensity per unit band, we have

$$J'_{th} A_c \eta_{AE} = U_{th}$$

or

$$J'_{th} = \frac{4\pi f^2 k_o T_k d}{\eta_{AE} c^2} \tag{815}$$

and the transducer equivalent spectrum level is

$$L'_{th} = 10 \log J'_{th}$$

$$= 20 \log f + N_{AE} - D - 10 \log \frac{c^2}{4\pi k_o T_K} \tag{816}$$

For satisfactory performance it is desirable that the overall thermal equivalent spectrum level (including the thermal effects of both the water and the internal electrical components) be at least 10 db below the ambient noise level.

D. Calibration of Transducers

This section will be limited to a brief discussion of some of the principles involved in transducer calibration. A description of the hardware and operational procedures employed in transducer calibration stations is beyond the scope of these notes.

Calibration measurements are made for a number of different purposes, such as to determine beam patterns, frequency response, and sensitivity. For many of these purposes, such as the determination of beam patterns, only relative measurements are required. However, the determination of sensitivity requires absolute measurements - that is to say, measurements of the relationship between the electrical input and acoustical output of a projector, and the relationship between the acoustical input and

electrical output of a hydrophone. In routine operations, such measurements are made relative to a calibrated standard transducer. Thus, although the routine measurements are in this sense relative measurements, they depend ultimately upon the absolute calibration of the standard. Absolute measurements are made by a method which is based on the reciprocity theorem, which we shall discuss shortly. Before doing this, however, we shall introduce a little background material.

1. Transducer Sensitivity.

The transmitting sensitivity of a projector is the ratio of the free-field pressure produced at a specified distance from the transducer to the input electrical current.

$$S_T = \frac{\text{Free-field rms pressure at distance } r}{\text{Input rms current}} \qquad (821)$$

Because of the spreading of sound waves the sensitivity is a function of the distance r. In these notes we shall make the following assumptions:

(1) For a directional transducer we shall consider pressure to be measured on the beam axis.

(2) We shall consider only the far-field response.

The term "free-field" signifies that the pressure is due to the outgoing wave alone, the field being free of reflections from any bounding surfaces.

For the purpose of standardization, the sensitivity should be referred to a standard distance, such as 1 yard. Actual measurements, of course, must be made in the far field (or suitable corrections must be applied to near-field measurements), and the conversion to the 1 yard reference distance is made by application of the spherical spreading law.

In the CGS system the transmitting sensitivity is expressed in μb/ampere, and in the MKS system in (newtons/m^2)/ampere, where

$$1 \text{ newton}/m^2 = 10 \text{ μb}$$

The sensitivity is also frequently expressed in decibels. Since the power is proportional to the square of the pressure and to the square of the current, the sensitivity is 20 log S_T, in db referred to 1 μb/amp or 1 (newton/m^2)/amp.

The receiving sensitivity of a hydrophone is the ratio of the open-circuit output voltage to the free-field pressure at the location of the hydrophone

$$S_R = \frac{\text{Output open-circuit rms voltage}}{\text{Free-field rms pressure}} \tag{822}$$

Here we assume that the pressure is due to a plane wave arriving in the direction of the beam axis. Also, it should be noted that in general the presence of the hydrophone alters the sound field. The pressure appearing in the definition (822) is understood to be the pressure which would exist if the hydrophone were not there. Expressed in decibels the receiving sensitivity is 20 log S_R, in db referred to a volt per microbar or a volt per newton/m^2, depending on the system of units.

2. Pressure Due to a Simple Source; Extended Source

At this point we shall extend our earlier discussion of spherical waves. The pressure and particle velocity of a spherical wave was expressed in terms of an arbitrary constant A in equations (136) and (138). Let us now assume that the spherical wave is generated by a small pulsating sphere of mean radius a, such that

$$a << \lambda \tag{823}$$

Let the rate of change of the radius, that is, the radial velocity, be

$$u = U e^{jwt} \tag{824}$$

At the surface of the sphere, the particle velocity of the water must be equal to the velocity of the spherical surface. Therefore, when r = a, equations (138) and (824) yield

$$U = \frac{A}{\rho c a}\left(1 - \frac{j}{ka}\right) e^{-jka}$$

$$= -\frac{A}{\rho c k a^2}(j - ka)(1 - jka + \cdots) \tag{825}$$

But from (823) we see that, since $k = \frac{2\pi}{\lambda}$,

$$ka << 1 \tag{823a}$$

If we neglect ka in comparison with 1, we obtain

$$A = j\rho c k a^2 U \tag{826}$$

Let us now express U in terms of the rate of change of the volume of the sphere. The amplitude of the pulsation is the integral of (824), so that the volume of the sphere is

$$V = \frac{4\pi}{3}\left(a + \frac{U}{jw}\, e^{jwt}\right)^3 \tag{827}$$

Since the amplitude $\frac{U}{w}$ is assumed to be small compared to a, we may retain only the first power of U in the expansion of (827)

$$V = \frac{4\pi a^3}{3} + \frac{4\pi a^2 U}{jw}\, e^{jwt} \tag{827a}$$

Let

$$Q\, e^{jwt} = \frac{dV}{dt}$$

Then

$$Q = 4\pi a^2 U \tag{828}$$

The quantity Q is called the volume velocity of the sphere. Note that (828) could have been derived directly by observing that for small amplitudes the volume velocity is the product of the linear velocity and the area of the spherical surface. Substitution of (828) into (826) yields

$$A = \frac{j\rho ck}{4\pi} Q = \frac{j\rho c}{2\lambda} Q \tag{829}$$

The acoustic pressure at a distance r (136) is then

$$p = \frac{j\rho cQ}{2\lambda r}\, e^{j(wt-kr)} \tag{830}$$

The small sphere which generates the pressure field (830) is called a simple source and the volume velocity Q is called the strength of the source.

If we now consider an extended source consisting of a radiating surface of finite area, we may break up the surface into elements dS, as we did in deriving beam patterns, and treat each element as a simple source. The general expression for the pressure is then

$$p = \frac{j\rho c}{2\lambda} \iint_S \frac{e^{j(wt-kr')}}{r'} U_n\, dS \tag{831}$$

where r' is the distance from the element dS to the point at which the pressure is computed and U_n is the component of the velocity amplitude normal to the

surface. If we restrict our attention to a point on the beam axis in the far field, then r' has a constant value r for all elements of the surface, and

$$p = \frac{j\rho c}{2\lambda r} e^{j(wt-kr)} \iint_S U_n \, dS \tag{832}$$

In this case the pressure along the axis is equal to that produced by a simple source of strength

$$Q_s = \iint_S U_n \, dS \tag{833}$$

3. The Reciprocity Theorem

The relationship between the transmitting and receiving sensitivities of a reversible transducer can be determined from an application of the reciprocity theorem stated by Lord Rayleigh in his classical work, The Theory of Sound. The theorem states that in a linear network, if we observe the output in branch B caused by an input in branch A, and if, on the other hand, we observe the output in branch A caused by an input in branch B, then the ratios of the outputs to their respective inputs will be equal.

To apply this theorem to the calibration of a transducer, the network under consideration consists of the transducer plus the acoustic medium out to the point P at distance r. Let branch A be the electrical terminals of the transducer and branch B be located in the medium at the point P. The inputs and outputs are as follows:

Input A: Electrical current

Output B: Free-field pressure at P

Input B: Simple source at P

Output A: Open-circuit voltage at terminals

Let the following quantities refer to rms values:

I = input current

p' = free-field pressure at P due to I

Q = strength of simple source at P

p = free-field pressure at transducer, due to Q

E = open circuit voltage due to p (and hence to Q)

Then the sensitivities are

$$S_T = \frac{p'}{I} \tag{834}$$

$$S_R = \frac{E}{p} \tag{835}$$

and the reciprocity thereon states

$$\frac{p'}{I} = \frac{E}{Q} \quad \text{(MKS system)} \tag{836}$$

Although we shall not prove the theorem, a few words of explanation may be useful. When the transducer is used as a projector, the power is applied at the electrical terminals. At the other end the source is removed and the free-field pressure is the "open-circuit" output. When the transducer is used as a hydrophone, the input power is applied acoustically at the simple source Q; the electrical power source is removed, and the output is the open-circuit voltage. We have previously considered an electro-acoustic network in which force and linear velocity were the analogs of voltage and current. Here the corresponding quantities are pressure and volume velocity. A check of the dimensions in (836) will show that both products EI and $p'Q$ have the dimensions of power. A more careful check of the units reveals the following.

The product EI (volts x amperes) is expressed in watts.

The product $p'Q$ has the dimensions of force/area times volume/sec, which is force x velocity. In the MKS system

$$1 \text{ newton} \times 1 \text{ meter/sec} = 1 \text{ watt}$$

In the CGS system

$$1 \text{ dyne} \times 1 \text{ cm/sec} = 1 \text{ erg/sec} = 10^{-7} \text{ watt}$$

Therefore, the reciprocity relation, as expressed in (836) is valid only in the MKS system. In the CGS system it must be replaced by

$$\frac{p'}{I} = 10^7 \frac{E}{Q} \quad \text{(CGS system)} \tag{836a}$$

From here on we shall assume the CGS system and shall use (836a).

We are now in a position to relate the sensitivities S_T and S_R. The pressure generated at the transducer by the source Q is given by (830). If we consider p', I, E, and Q to be rms values, then, since the absolute value of the complex exponential is 1, or

$$\left| j\, e^{j(wt-kr)} \right| = 1$$

the rms pressure is

$$p = \frac{\rho c Q}{2\lambda r}$$

so that

$$\frac{p'}{I} = \frac{10^7\, \rho c}{2\lambda r} \quad \frac{E}{p} \tag{837}$$

Insertion of (834) and (835) yields

$$S_T = \frac{10^7\, \rho c}{2\lambda r} S_R$$

or

$$S_R = J\, S_T \tag{838}$$

where

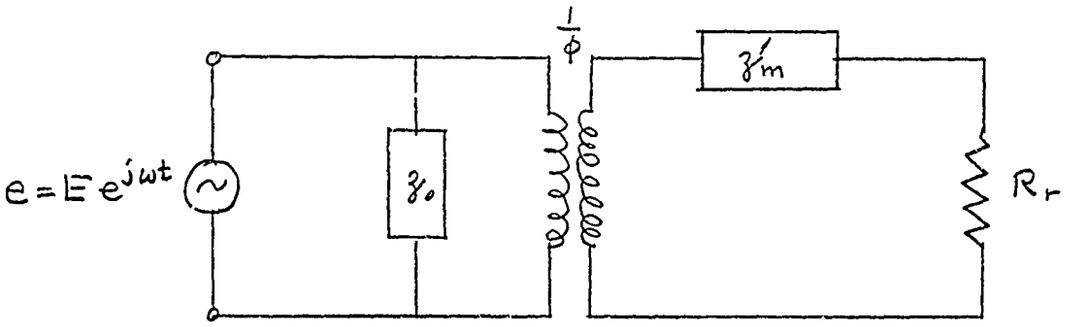
$$J = \frac{2 \times 10^{-7}\, \lambda r}{\rho c} \quad \frac{cm^4\, sec}{gm} \tag{839}$$

The constant J is called the reciprocity parameter.

It is seen that the reciprocity parameter depends only upon the wavelength λ, the distance r at which the transmitting sensitivity is measured, and the specific acoustic impedance ρc of the medium. For all transducers for which this relation holds, the ratio of the sensitivities is independent of the transducer parameters.

4. Direct Calculation of J for Quartz Crystal Transducer.

As a check on the general theorem stated above, let us calculate the transmitting and receiving sensitivities of the ideal quarter-wave quartz crystal transducer which was analyzed previously. The equivalent circuit of the transducer as a projector is reproduced below.



The impedance z_o is the blocked impedance consisting of the clamped capacitance C_o shunted by the leakage resistance R_o. The mechanical impedance z'_m includes all the motional characteristics except the radiation resistance R_r. The previously defined mechanical impedance z_m includes R_r, so that

$$z_m = z'_m + R_r \tag{840}$$

The relation between the velocity amplitude U of the crystal face and the amplitude of the input voltage E is given by (628) which may be written as

$$U = \frac{\phi E}{z'_m + R_r} \tag{841}$$

The relation between the input current and voltage is given by (648), which may be written as

$$I = \left(\frac{1}{z_o} + \frac{\phi^2}{z'_m + R_r}\right) E \tag{842}$$

Inserting (842) into (841) we obtain

$$U = \frac{\phi I z_o}{\phi^2 z_o + z'_m + R_r} \tag{843}$$

To compute the pressure at a point at a distance r on the beam axis, we make the following assumptions.

(1) The area S of the transducer is a plane and its dimensions are large compared to a wavelength.

(2) The point at which we desire to know the pressure is in the far field, so that the distance r is constant for all points on S.

(3) The transducer is mounted in an infinite baffle, so that all radiation is confined to the forward hemisphere.

With these assumptions the pressure is equal to twice the value given by (832). If P denotes the amplitude, so that

$$p = P e^{jwt}$$

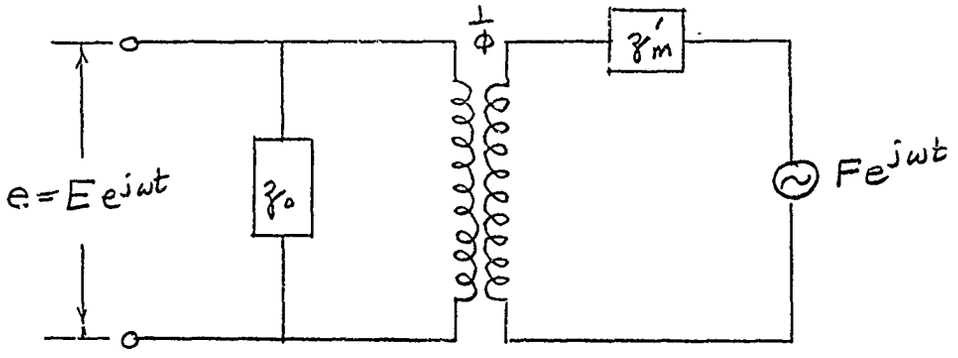
then

$$P = \frac{j\rho c\, US}{\lambda r} e^{-j\, kr} \tag{844}$$

The transmitting sensitivity is obtained from (843) and (844)

$$S_T = \left|\frac{P}{I}\right| = \frac{\rho c S \phi}{\lambda r} \left|\frac{z_o}{\phi^2 z_c + z'_m + R_r}\right| \tag{845}$$

The equivalent circuit for the measurement of receiving sensitivity is as follows



where F is the amplitude of the force exerted by the water on the transducer face,

$$F = P'S \tag{846}$$

and P' is the pressure at the transducer face. P' is not the free-field pressure, because of the disturbance caused by the presence of the transducer. Looking in at the acoustic input terminals, we see that the circuit is a series combination of a mechanical impedance z'_m and an equivalent impedance $\phi^2 z_o$. Hence the relationship between P' and the velocity amplitude U of the face is

$$F'S = (\phi^2 z_o + z'_m) U \tag{847}$$

and the output open circuit voltage amplitude is

$$E = \frac{P'S - z'_m U}{\phi}$$

$$\approx \frac{\phi z_o P'S}{\phi^2 z_o + z'_m} \tag{848}$$

We must now relate P' to the free-field pressure P. In accordance with our previous assumptions we may treat the incoming wave as a plane wave arriving in a direction normal to the surface. In order to satisfy the phase relations in the transducer there must be a reflected wave, which we assume also to be planar.

The free-field pressure P is the pressure of the incident wave. The driving pressure P' is the sum of the pressures of the incident and reflected waves. The velocity of the surface is the (algebraic) sum of the particle velocities of the two waves. If the subscripts i and r refer to incident and reflected, then

$$P_i = \rho c U_i = P \tag{849}$$

$$P_r = \rho c U_r \tag{849a}$$

and

$$P' = P_i + P_r \tag{850}$$

$$U = U_i - U_r \tag{850a}$$

Combining these equations, we find

$$P' = P + \rho c U_r$$

$$\rho c U = P - \rho c U_r$$

Hence

$$P = \tfrac{1}{2}(P' + \rho c U) \tag{851}$$

Insertion of (847) into (851) yields

$$P = \frac{P'}{2}\left(1 + \frac{\rho c S}{\phi^2 z_o + z'_m}\right)$$

or, since

$$R_r = \rho c S$$

$$P' = 2P \frac{\phi^2 z_o + z'_m}{\phi^2 z_o + z'_m + R_r} \tag{852}$$

The receiving sensitivity is obtained by substituting (852) into (848)

$$S_R = \left|\frac{E}{P}\right| = 2\phi S \left|\frac{z_o}{\phi^2 z_o + z'_m + R_r}\right| \tag{853}$$

Comparison of (853) and (845) shows

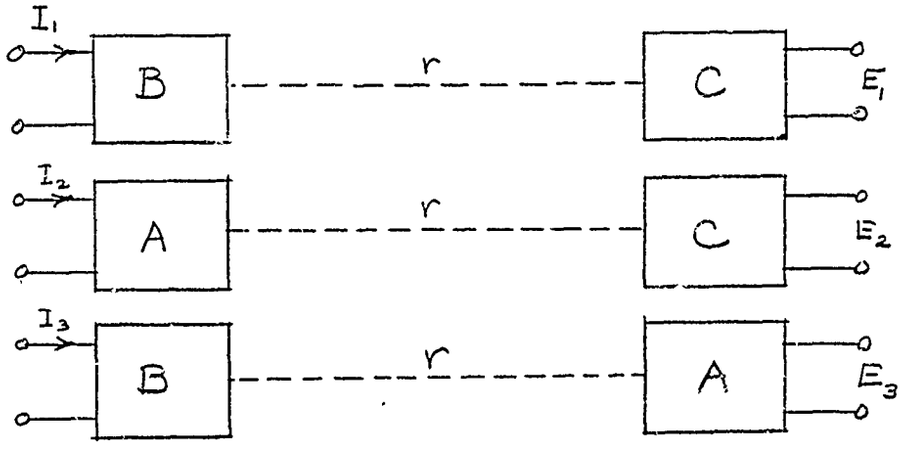
$$S_R = \frac{2\lambda r}{\rho c} S_T$$

which (except for the factor 10^{-7} which we have ignored) is the same as (838). The ideal quartz crystal obeys the reciprocity theorem.

5. Absolute Calibration Using Reciprocity Theorem.

To measure the sensitivity of a transducer by use of the reciprocity theorem, it is necessary to employ three transducers, which we shall label A, B, and C. Transducer A is the transducer being calibrated. It must obey the reciprocity relation; in order to do so, it must be linear and have the same efficiency when used as a projector as when used as a hydrophone. Transducer B is used only as a projector; its receiving characteristics are unimportant. Transducer C is used only as a hydrophone; its transmitting characteristics are unimportant.

The calibration consists of three measurements as indicated in the following sketch and table.



Measurement No.	Projector	Hydrophone	Input Current	Output Voltage
1	B	C	I_1	E_1
2	A	C	I_2	E_2
3	B	A	I_3	E_3

The receiving sensitivity of transducer C is the same in measurements 1 and 2, and if the same separation between transducers is employed in measurements 1 and 3, the transmitting sensitivity of B is the same for both measurements. (If the separations are different, an appropriate correction must be made, based on the fact that the transmitting sensitivity is inversely proportional to the separation distance.)

Let the sensitivities of the three transducers be identified by subscripts A, B, and C. Now, in each measurement the free-field pressure produced by the projector is identical to the free-field pressure measured by the hydrophone. In other words, the pressure p' in the transmitting sensitivity of the projector is the same as the pressure p in the receiving sensitivity of the hydrophone. Applying equations (834) and (835) to the three measurements, we obtain the following relations

$$S_{TB}\, S_{RC} = \frac{E_1}{I_1} \tag{854a}$$

$$S_{TA}\, S_{RC} = \frac{E_2}{I_2} \tag{854b}$$

$$S_{TB}\, S_{RA} = \frac{E_3}{I_3} \tag{854c}$$

Furthermore, the reciprocity relation (838) applies to transducer A

$$S_{RA} = J\, S_{TA} \tag{855}$$

The above four equations may now be solved to give the following results

$$S_{TA} = \sqrt{\frac{I_1\, E_2\, E_3}{E_1\, I_2\, I_3\, J}} \tag{856a}$$

$$S_{RA} = \sqrt{\frac{I_1\, E_2\, E_3\, J}{E_1\, I_2\, I_3}} \tag{856b}$$

Also, as a by-product, the sensitivities S_{TB} and S_{RC} may also be evaluated

$$S_{TB} = \sqrt{\frac{E_1\, I_2\, E_3}{I_1 E_2\, I_3\, J}} \tag{857}$$

$$S_{RC} = \sqrt{\frac{E_1\, E_2\, I_3\, J}{I_1\, I_2\, E_3}} \tag{858}$$

6. Requirements for Free-Field Calibration.

For accurate calibration it is important that the sound field at the receiving hydrophone be free from reflections from the bounding surfaces of the body of water in which the measurements are made. Unless special precautions are taken or special methods are used, the only way of guaranteeing free-field conditions is to make the dimensions of the test facility so large that the spreading loss of the reflected paths reduces the reflected intensity to acceptable levels. For very precise measurements the reflected signals should be at least 40 db below the direct-path signal. However, for most measurements a reduction of 20 db is adequate.

When the calibration is done in lakes or other large bodies of water, the limiting dimension is usually the water depth. For minimum interference from both top and bottom, the two transducers (the transducer being calibrated and the reference) should be located half-way down. If h is the water depth, there is then a clearance of $\frac{1}{2}h$ above and below. If the transducers are separated by a distance r, the requirement that the spreading loss of the surface-or-bottom-reflected path be at least 20 db higher than that of the direct path is

$$20 \log \frac{h}{r} \geq 20$$

or

$$h \geq 10\, r \tag{858}$$

Since the minimum acceptable depth depends upon the separation r, we must investigate the requirements placed upon r. We have already seen that to achieve far-field conditions the separation must meet the requirement

$$r \geq \frac{d^2}{\lambda} \tag{859}$$

where d is the maximum linear dimension of the active portion of the transducer and λ is the wavelength. The inequality (859) is based on the requirement that the far-field pressure distribution be applicable along the axis of a directive transducer. If we consider the situation which exists when the orientation is at right angles to the beam, we see that the separation between the nearest

portion and the reference transducer is $r - \frac{1}{2}d$ and the separation between the farthest portion and the reference transducer is $r + \frac{1}{2}d$. The difference in intensity levels for such an orientation is

$$\Delta L = 20 \log \left(\frac{r + \frac{1}{2}d}{r - \frac{1}{2}d}\right)$$

A rule of thumb which is commonly used for this condition is

$$r \geq 5d \tag{860}$$

It is seen that this rule leads to a maximum intensity difference between the nearest and farthest elements of about 1.75 db.

From these two conditions and from (858) we find that the minimum water depth is determined by

$$h \geq \frac{10d^2}{\lambda} \tag{861}$$

$$h \geq 50d \tag{862}$$

For large, highly directive modern transducers the depth requirements are quite severe. Depths of hundreds of feet are frequently required. The Navy has two inland deep-water calibration stations. One is located on Lake Pend Oreille, Idaho, which has a maximum depth in excess of 1000 feet. The other station is located in the eastern part of the country on Seneca Lake, New York, with a depth of about 600 feet.

7. Special Calibration Techniques.

Since the two large facilities have a limited workload capability, it is necessary to conduct most calibration measurements in smaller local facilities. Most laboratories engaged in underwater acoustic work have small outdoor bodies of water which are suitable for a large fraction of the calibration measurements. In addition, it is useful to perform measurements in indoor tanks.

Anechoic Linings

When measurements must be made in small volumes, the problem of reflections can be alleviated by a number of special techniques. The most obvious idea is to make the walls non-reflective. Considerable success has

been achieved with a material known as insulcrete, consisting of sawdust and concrete. The boundaries are lined with wedges of this material. One practical problem is that for effective anechoic linings at low frequencies the wedges must be very large. Several rubber-base materials have good absorption characteristics at low pressures but are not very effective at high pressures.

The Navy Electronics Laboratory has recently developed a new type of calibration pool, Transdec, in which boundary reflection is controlled. The shape of the pool is that of a canted ellipsoid of revolution. Reflections from the walls of the pool and from the surface of the water terminate in an absorptive sound trap which surrounds the lip of the pool. By this technique the interference due to reflected energy is sharply reduced and the far-field test region is an appreciable portion of the total volume of the pool.

Special Methods Involving Transducer Directivity

There is a class of techniques which takes advantage of transducer directivity in order to reduce the effect of reflected energy. An example of such a technique is the use of a vertical line for the reference transducer, the line being so placed that the reflected waves from the surface and bottom arrive in the direction of a null in the beam pattern of the line. The application of these techniques must be worked out separately for each individual case.

Pulse Measurements

An appreciable reduction in the required dimensions of the test volume can be achieved by employing pulsed operation. In this method a short pulse is emitted and received before the boundary reflections arrive. When this method is used, however, it is necessary that the received pulse be of sufficient duration to reach steady state before any reflected signals arrive. For a transducer of small dimensions the time to build up to 95 percent of the steady state value is approximately Q/f, where Q is the overall Q of the calibration system including the transducer under test. This result

may be obtained from the fact that the transient build-up of the amplitude of a series resonant circuit is proportional to

$$1 - e^{-\frac{Rt}{2L}}$$

while

$$Q = \frac{2\pi f_o L}{R}$$

Hence, at 95 percent of steady state

$$1 - e^{-\frac{\pi f_o t}{Q}} = 0.95$$

or

$$\frac{\pi f_o t}{Q} = -\log_e 0.05$$

$$t = \frac{3.0\,Q}{\pi f_o} \approx \frac{Q}{f_o} \tag{863}$$

If the transducer has appreciable size, an additional delay is caused by interaction between elements, for which an allowance should be made equal to the time required for the wave to sweep at least once over the maximum dimension d of the transducer. Furthermore, in order to obtain a satisfactory measurement it is advisable to continue the pulse for an additional time equivalent to at least three cycles at the operating frequency. Adding all these time intervals and multiplying by the speed of sound, we obtain the result that the difference in length Δr between the reflected path and the direct path should be at least

$$\Delta r = (Q + 3)\,\lambda + d \tag{864}$$

or

$$\Delta r = (Q + n + 3)\,\lambda \tag{864a}$$

where n is the number of wavelengths in the maximum dimension, that is,

$$n = \frac{d}{\lambda} \tag{865}$$

In facilities where the depth h is the limiting dimension, this criterion may be expressed as

$$\Delta r = \sqrt{h^2 + r^2} - r$$

or

$$h = \sqrt{2r\,\Delta r + (\Delta r)^2} \qquad (866)$$

As a numerical example, suppose the transducer is two wavelengths long (n = 2) and the Q of the system is 7. The criteria (859) and (860) give

$$r \geq \frac{(2\lambda)^2}{\lambda} = 4\lambda$$

$$r \geq 5(2\lambda) = 10\lambda$$

Also, $$\Delta r = (7 + 2 + 3)\lambda = 12\lambda$$

The required depth is then

$$h \geq \sqrt{2(10\lambda)(12\lambda) + (12\lambda)^2}$$

$$= \sqrt{384}\lambda \approx 20\lambda$$

The minimum depth as a function of frequency is shown in the following table.

Frequency (kc)	Min. Depth (ft.)
1	100
2	50
5	20
10	10
20	5

Noise Techniques

A radically different technique having limited though very useful application to small hydrophones is the so-called noise method, in which the calibration tank is flooded with wideband noise and the frequency response of the hydrophone is obtained by sweeping through the desired range of frequencies with a bandpass filter. The calibration is obtained by comparing the results against similar data taken on a reference hydrophone. This method is used at the Naval Air Development Center for calibrating sonobuoy hydrophones over a range of frequencies extending down to 200 cps. It is rapid, inexpensive, and repeatable to within about ± 1 db, and is, in fact,

the only economically feasible method of calibrating large quantities of cheap hydrophones. Experience has shown that a remarkably uniform sound field can be generated, the variation being less than 0.5 db over a range of 10 inches. The noise method is of course subject to error in the vicinity of resonant frequencies and its application is therefore limited to relatively low-Q devices.

8. Near-Field Calibration Techniques

The development, since World War II, of larger and larger transducer arrays has provided increasing impetus to efforts to develop and perfect near-field calibration techniques. If, through suitable analysis of measurements made in the immediate vicinity of the transducer, one were able to predict the far-field pattern, it would then be possible to carry out the calibration procedure in a volume not much larger than that occupied by the transducer itself.

Considerable effort has been applied to this problem, both at the Underwater Sound Reference Laboratory, Orlando, Florida, and at the Defense Research Laboratory of the University of Texas. Several methods have been explored and techniques have been developed to the point where routine calibrations can be run successfully on many types of transducers.

In this section we shall discuss the DRL method. It is based fundamentally on Green's theorem, which states that if $\phi(x, y, z)$ and $\psi(x, y, z)$ are continuous functions of the spatial coordinates, then

$$\iint_S \left(\phi \frac{\partial \psi}{\partial n} - \psi \frac{\partial \phi}{\partial n}\right) dS = \iiint_V \left(\phi \nabla^2 \psi - \psi \nabla^2 \phi\right) dV \qquad (867)$$

where the surface integral is taken over a closed surface S and the volume integral is taken over the entire volume enclosed by S. The derivatives $\frac{\partial \psi}{\partial n}$ and $\frac{\partial \phi}{\partial n}$ are taken along the direction normal to the surface element dS, and ∇^2 is the Laplacian operator (111).

In applying this theorem to the near-field calibration problem, we surround the transducer with an imaginary surface S, which is the surface over which the near-field measurements are taken. At the point P in the

far field, at which we desire to determine the pressure, we place (mathematically) a simple source. We assume in the present discussion that the pressure field is sinusoidal, and to simplify the writing we shall factor out the complex exponential e^{jwt} which appears in every term.

Let ϕ represent the pressure due to the transducer, that is, due to the sources located within the closed surface S,

$$\phi(x, y, z) = p(x, y, z) \tag{868}$$

and let ψ be

$$\psi(x, y, z) = \frac{e^{jkr}}{r} \tag{869}$$

where r is the radial distance measured from the point P at which the simple source is located, and k is the wave number.

$$k = \frac{2\pi}{\lambda} = \frac{w}{c}$$

The function ψ is proportional to the pressure which would be created by a simple source located at the point P.

We are concerned in (867) with the region outside the closed surface S. Since this region extends all the way to infinity, there are two bounding surfaces - the surface S surrounding the transducer, and the surface of the sphere at infinity. However, both ϕ and ψ vanish at infinity, and the integral of $\phi \frac{\partial \psi}{\partial n} - \psi \frac{\partial \phi}{\partial n}$ over the infinite sphere is zero.

It will be noted that both the pressure p and the function $\frac{e^{jkr}}{r}$ satisfy the wave equation

$$\nabla^2 \phi + k^2 \phi = 0 \tag{870}$$

at all points in space except at a sound source. It may be shown that if the region V contains no sources, so that both ϕ and ψ obey the wave equation (870), then the volume integral on the right side of (867) vanishes. By our hypothesis the transducer is completely enclosed within the surface S and is therefore not included in the volume V. The only source we have to

worry about is the simple source at P. Let us exclude this source from the region of integration by surrounding P with a small sphere of radius a. The volume integral is now zero and in its place we have a surface integral over the small sphere. Equation (867) is now transformed to the following

$$\iint_S \left[p \frac{\partial}{\partial n} \left(\frac{e^{jkr}}{r} \right) - \left(\frac{e^{jkr}}{r} \right) \frac{\partial p}{\partial n} \right] dS + \iint_{\substack{\text{sphere} \\ r=a}} \left[p \frac{\partial}{\partial n} \left(\frac{e^{jkr}}{r} \right) - \left(\frac{e^{jkr}}{r} \right) \frac{\partial p}{\partial n} \right] dS = 0 \quad (871)$$

Strictly speaking, since we are integrating over the region outside the closed surfaces, the direction n in the derivatives in (871) should be measured toward the insides of the surfaces. However, the equation will remain valid if we reverse the direction of n in both integrals, so that it represents the outward normal to each surface. Considering the small sphere, the normal n has the direction of the radius r, so that

$$\frac{\partial}{\partial n} \left(\frac{e^{jkr}}{r} \right)_{r=a} = \frac{d}{dr} \left(\frac{e^{jkr}}{r} \right)_{r=a}$$

$$= -\frac{1}{a^2} (1 - jka) e^{jka}$$

If a becomes very small, this approaches the value

$$\frac{\partial}{\partial n} \left(\frac{e^{jkr}}{r} \right)_{r=a} \longrightarrow -\frac{1}{a^2}$$

Also, as a approaches zero, the pressure p approaches the constant value p(P), corresponding to the point P at the center of the sphere. Furthermore the factor $\frac{e^{jkr}}{r}$ has a constant value over the surface of the sphere, while the pressure gradient $\frac{\partial p}{\partial n}$ averages to zero as the radius of the sphere becomes very small. The integral over the small sphere therefore approaches the value

$$\iint_{\substack{\text{sphere} \\ r=a}} \left[p \frac{\partial}{\partial n} \left(\frac{e^{jkr}}{r} \right) - \left(\frac{e^{jkr}}{r} \right) \frac{\partial p}{\partial n} \right] dS \longrightarrow -\frac{p(P)}{a^2} \iint_{\substack{\text{sphere} \\ r=a}} dS$$

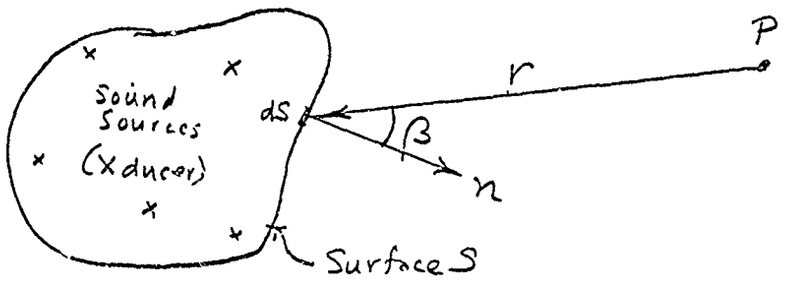
$$= -4\pi p(P)$$

The far-field pressure is obtained by substituting this result into (871).

$$p(P) = \frac{1}{4\pi} \iint_S \left[p \frac{\partial}{\partial n} \left(\frac{e^{jkr}}{r} \right) - \left(\frac{e^{jkr}}{r} \right) \frac{\partial p}{\partial n} \right] dS \qquad (872)$$

where r is the distance from the far-field point P to the element dS of the imaginary surface S surrounding the transducer, the derivatives are taken along the outward normal to S, and the integration is carried out over the entire closed surface. This is the basic equation for near-field calibration.

It is seen that in order to implement (872) exactly, it is necessary to measure both the amplitude and phase of not only the pressure but also the normal component of the pressure gradient, at all points over the surface. In the real world it is virtually impossible to make all the measurements indicated. For practical measurements the following approximations are made.



(1) The pressure gradient is approximated by

$$\frac{\partial p}{\partial n} = j k p \qquad (873)$$

This would be the correct expression if the pressure were a plane wave traveling in a direction normal to the surface S. It should be a reasonably valid approximation for a transducer of moderate curvature and a close-fitting surface over which the pressure does not change rapidly.

(2) Since the distance r to the far-field is normally many wavelengths,

$$kr >> 1$$

the following approximation is valid

$$\frac{\partial}{\partial n}\left(\frac{e^{jkr}}{r}\right) = \frac{d}{dr}\left(\frac{e^{jkr}}{r}\right)\frac{\partial r}{\partial n}$$

$$= jk\left(1 - \frac{1}{jkr}\right)\left(\frac{e^{jkr}}{r}\right)\frac{\partial r}{\partial n}$$

$$\approx jk\left(\frac{e^{jkr}}{r}\right)\frac{\partial r}{\partial n}$$

From the above sketch it can be seen that

$$\frac{\partial r}{\partial n} = -\cos\beta$$

where β is the supplement of the angle between r and n. Hence

$$\frac{\partial}{\partial n}\left(\frac{e^{jkr}}{r}\right) \approx -jk\left(\frac{e^{jkr}}{r}\right)\cos\beta \qquad (874)$$

With these approximations (852) becomes

$$p(P) = -\frac{jk}{4\pi}\iint_S (1+\cos\beta)\frac{e^{jkr}}{r}\, p\, dS \qquad (875)$$

In taking measurements on a projector, a small probe is moved over the surface S, and measurements of both the amplitude and phase of the pressure are made at discrete points. Although theoretically S is a closed surface, in practice it is not necessary to take measurements in regions where the pressure is very low. It has been found, for example, that for a planar array, it is sufficient to move the probe only over a plane in front of the array. For a cylindrical array the surface S is a cylinder concentric with the array, and the probe is moved in a circle about the axis of the cylinder. It has been found that three such passes are usually adequate, one in a plane through the center of the transducer and one in each of two parallel planes near the top and the bottom of the transducer. In calibrating transducers whose beam patterns are omnidirectional in azimuth, such as line transducers, measurements need to be taken in only one vertical plane. In general, the spacing between adjacent probe measurement positions should not exceed one half wavelength.

The integration indicated by (875) is carried out numerically on a digital computer which has been programmed to receive as inputs the data obtained from the probe measurements.

As a check on the near-field method, a number of transducers have been calibrated in the far field by conventional methods, and the results have been found to agree with the predictions from near-field measurements to within about ± 1 db. Successful near-field calibrations have been carried out in a tank whose diameter is less than three times the transducer diameter. Such measurements, of course, are pulse measurements; in fact, the data obtained at such close quarters must be based on measurements made near the leading edge of the pulse, before steady-state conditions are reached.

The function $\frac{e^{jkr}}{r}$ in (869) is one of a class of functions known as Green's functions. Any function which satisfies the wave equation (870) can be used in lieu of $\frac{e^{jkr}}{r}$ in (872) to compute the far-field pressure. It is possible in some cases to find a shape of the surface S and a matching Green's function such that the function vanishes over the surface, thereby reducing the integrand of (872) to one term and avoiding the necessity for measuring the pressure gradient $\frac{\partial p}{\partial n}$. However, it is difficult to measure the pressure over surfaces (such as spheres) for which the Green's function is easily determined, and it is also difficult to find Green's functions for surfaces over which p is easy to measure.*

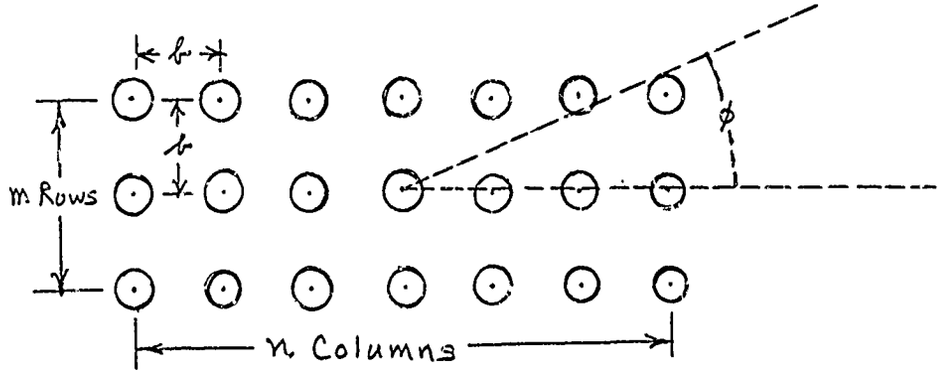
* Defense Research Laboratory Report No. DRL-A-196, The Determination of Farfield Characteristics of Large, Low-Frequency Transducers from Nearfield Measurements, by D. D. Baker, 15 March 1962.

E. Transducer Systems

Up to this point we have been considering the characteristics of individual transducer elements and of very simple arrays of omnidirectional elements. Many sonar systems both in fleet use and in experimental development have complicated transducer systems consisting of numbers of elements electrically coupled in various ways. In this section we shall consider briefly a few of the characteristics of these systems.

1. Directivity Patterns of Simple Multi-Element Systems.

When a transducer array consists of a number of individual transducer elements, each having its own individual directivity pattern, the overall pattern of the array is a function not only of the individual patterns, but also of the arrangement of the elements in the array. In the special case where (1) all the elements are identical and (2) all are oriented in the same direction in space, the overall pattern may be computed in a relatively simple manner from the product theorem, which states that the overall relative pressure response is equal to the product of the relative pressure response of one of the individual elements and the relative pressure response of a hypothetical array in which the actual elements are replaced by omnidirectional elements. As an example, consider a plane rectangular array of circular pistons, each having a radius a, and arranged in a regular pattern of n columns and m rows with a spacing b between adjacent elements. Let θ be the polar angle measured from the axis normal



to the plane of the array (which is also normal to the face of each of the pistons), and let ϕ be the angle in the plane of the array as indicated in the sketch, which determines the plane in which θ is measured. The relative pressure ratio due to a single piston is given by (728b)

$$\left[\frac{p(\theta)}{p_o}\right]_1 = \frac{2\, J_1 (ka \sin \theta)}{ka \sin \theta} \qquad (881)$$

The relative pressure ratio due to a rectangular array of omnidirectional elements is (by analogy with (730)) the product of the ratios for two multi-spot linear arrays at right angles to each other.

$$\left[\frac{p(\theta,\phi)}{p_o}\right]_2 = \frac{\sin (\frac{1}{2} nkb \sin \theta \cos \phi)}{n \sin (\frac{1}{2} kb \sin \theta \cos \phi)} \cdot \frac{\sin (\frac{1}{2} mkb \sin \theta \sin \phi)}{m \sin (\frac{1}{2} kb \sin \theta \sin \phi)} \qquad (882)$$

The product theorem states that the relative pressure response of the complete array is

$$\frac{p(\theta, \phi)}{p_o} = \left[\frac{p(\theta)}{p_o}\right]_1 \left[\frac{p(\theta, \phi)}{p_o}\right]_2 \qquad (883)$$

This result may be justified from the following argument. In any given direction (θ, ϕ) the individual relative pressure response is the same for all elements. If the array consisted of omnidirectional elements, each having a relative pressure response of unity in all directions, the response of the array would be $\left[\frac{p(\theta, \phi)}{p_o}\right]_2$ of (882). Since the response of each individual element is $\left[\frac{p(\theta)}{p_o}\right]_1$ of (881), the overall relative pressure response of the complete array is the product of these two functions.

To compute the directivity factor of such an array one must integrate the square of (883) over the sphere surrounding the array. This is a job which is best done numerically on a digital computer. In some instances the resultant directivity factor turns out to be the product of the individual directivity factors

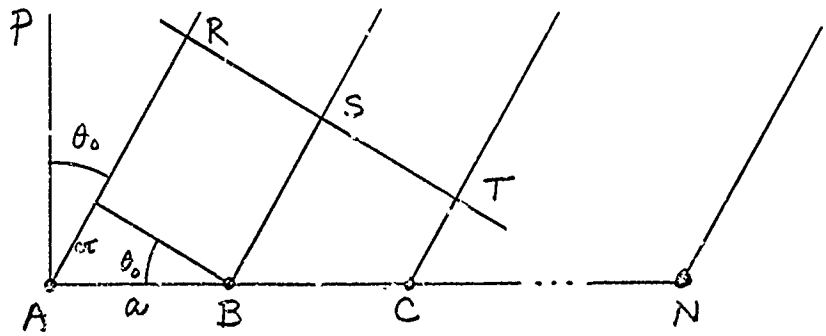
2. Electrically Steered Arrays

To introduce this subject let us imagine a transducer which is in the form of a large rectangular array of omnidirectional elements. If all these elements transmit sound waves in phase (or if, when used for receiving, the outputs are all connected together with no phase delays), then this array will have the relative pressure response (or relative voltage response) indicated by (882), with the axis of the beam oriented normal to the plane of the rectangle.

Suppose now that the rectangular transducer surface is oriented in a vertical plane and that it is desired to sweep the beam in azimuth or to tilt it up or down. An obvious way to do this would be to rotate the whole array mechanically about a vertical or horizontal axis. However, if the array is very large and heavy, as indeed some are, such a process might be quite difficult and impractical. Fortunately, there is another way to do this, which does not require any mechanical motion. By inserting electrical delays in the proper manner, thereby shifting the phase of some elements relative to others, it is possible to steer the array electrically in any desired direction.

a. Electrically Steered Linear Arrays.

To illustrate the point, consider a multispot linear array consisting of n uniformly spaced omnidirectional elements, the spacing being a. Let the elements be A, B, C, ···, N, as indicated in the diagram below. Considering the array as a projector, let us delay the output of B relative to A by a time τ. We likewise delay C relative to B, D relative to C, etc., by the same time τ. As may be deduced from the sketch, the



transmitted wave will not travel in the direction AP normal to the line, but in a direction AR making an angle θ_o with AP, such that

$$\sin \theta_o = \frac{c\tau}{a} \tag{884}$$

A typical wave front is RST, in which

$$AR - BS = c\tau = a \sin \theta_o$$

$$BS - C\dot{T} = c\tau \doteq a \sin \theta_o$$

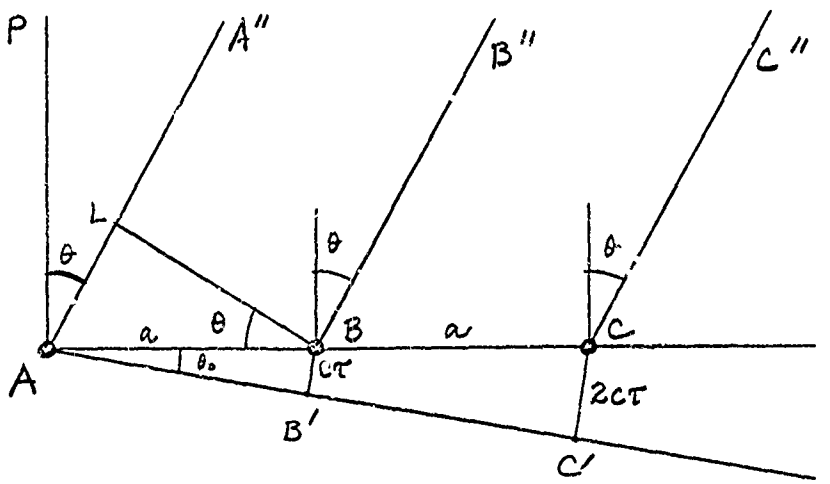
etc.

The effect of inserting the time delay is thus to steer the beam axis through the angle θ_o given by (884). By varying the time delay τ we can steer the beam in any direction we wish. Also, b reversing the direction of the delay, that is, by delaying A relative to B, etc., instead of B relative to A, etc., we can steer the beam in the opposite direction.

The series of networks by which these delays are applied is called a delay line or lag line, and the entire system including the switching mechanism by which the amount of the delay is adjusted, is called a compensator. The angle θ_o is called the compensation angle or steer angle. The electrical phase angle ψ_c which corresponds to the applied delay is

$$\psi_o = \frac{2\pi c\tau}{\lambda} = \frac{2\pi a}{\lambda} \sin \theta_o \tag{885}$$

Let us next examine the directivity pattern of the electrically steered array. In the accompanying sketch A, B, C, etc. represent



the transducer elements, separated by a distance a as before. We seek the relative response in the direction making an angle θ with the normal to the array, AP. To do this we proceed as we did previously for the uncompensated array by adding the pressures of the waves traveling along AA'', BB'', CC'', DD'', etc. The only difference between the present case and the previous analysis of a multispot array is the intentional time delay inserted between successive elements.

To facilitate the analysis, let us draw a line A B' C' ... making an angle θ_o with the array ABC... We then drop perpendiculars BB', CC', etc., from the array to this line. As may be seen from (884) the segment BB' is equal to the distance which would be traveled by a sound wave during the delay time τ. Also, the segment CC' is equal to the distance which would be traveled by a sound wave during the time 2τ, which is the time delay between C and A. It is evident that we may consider B', C', etc., as being hypothetical sound sources which emit waves in phase with A. These waves travel along the lines B'B, C'C, etc., arriving at B, C, etc. in time to account for the inserted time delays.

To compare the phase of the wave AA'' with that of BB'', all we need note is that by the time A has traveled a distance

$$AL = a \sin \theta$$

B has traveled an effective distance

$$BB' = c\tau - a \sin \theta_o$$

Hence the difference in effective distance traveled is

$$AL - BB' = a (\sin \theta - \sin \theta_o)$$

and the difference in phase is

$$\Delta\psi = \frac{2\pi a}{\lambda} (\sin \theta - \sin \theta_o)$$

This same phase difference occurs between BB'' and CC'', and between CC'' and DD'', etc.

For an uncompensated array the time delay τ is zero, so that θ_o is zero. Hence the phase difference between AA'' and BB'' is

$$\Delta\psi = \frac{2\pi a}{\lambda} \sin \theta \qquad \text{(uncompensated array)}$$

Therefore the effect of the compensation has been to change sin θ to sin θ - sin θ_o. The formula (715b) for the relative pressure response is applicable to a compensated array provided we redefine the angle ψ. Thus

$$\frac{p(\theta)}{p_o} = \frac{\sin n\psi}{n \sin \psi} \qquad (715b)$$

where

$$\psi = \frac{\pi a}{\lambda} (\sin \theta - \sin \theta_o) \qquad (886)$$

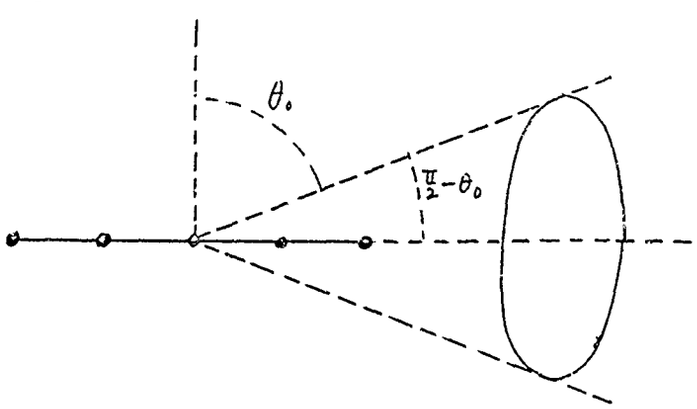
$$\sin \theta_o = \frac{c\tau}{a}$$

a = separation between adjacent elements of the array

τ = time delay between adjacent elements

It will be recalled that the beam pattern of an unsteered linear array is a figure of revolution obtained by rotating the two-dimensional pattern about the line of the array. In the two-dimensional pattern the beam axis is a line through the center of the array, perpendicular to the array. In the three-dimensional unsteered pattern, the locus of the maximum response axis is spread out over the equatorial plane. The side lobes are conically shaped regions.

In the electrically steered array the maximum response axis of the two-dimensional pattern is shifted away from the normal through the angle θ_o. The locus in three dimensions is thus a cone, similar to the side lobes of the unsteered array. The pattern is no longer symmetric about the maximum response axis.



One price which must be paid for electrical steering is a broadening of the beam. The larger the angle θ_o through which the beam is steered, the broader the beam becomes. The beam is broadest when the delay is such that $\sin \theta_o = 1$, or $\theta_o = 90°$, and the beam axis is pointed along the direction of the array line. This is called the end-fire position. (Although the beam is broadest in this position, the directivity index of an end-fire array is generally higher than that of the same array without electrical steering). An unusual end-fire condition occurs in the case where the element spacing is one-half wavelength. Examination of (886) shows that when

$$a = \tfrac{1}{2}\lambda \text{ and } \theta_o = \frac{\pi}{2}$$

the phase angle ψ is

$$\psi = -\frac{\pi}{2}(1 - \sin\theta)$$

The response, of course, is a maximum in the direction of the beam axis, $\theta = \frac{\pi}{2}$, where $\psi = 0$ and

$$\frac{\sin n\psi}{n \sin \psi} = 1$$

But it is also a maximum in the opposite direction, $\theta = -\frac{\pi}{2}$, since $\psi = -\pi$ when $\sin\theta = -1$. An end-fire array therefore has a high relative response in the backward direction at frequencies such that the element spacing is an integral number of half wavelengths.

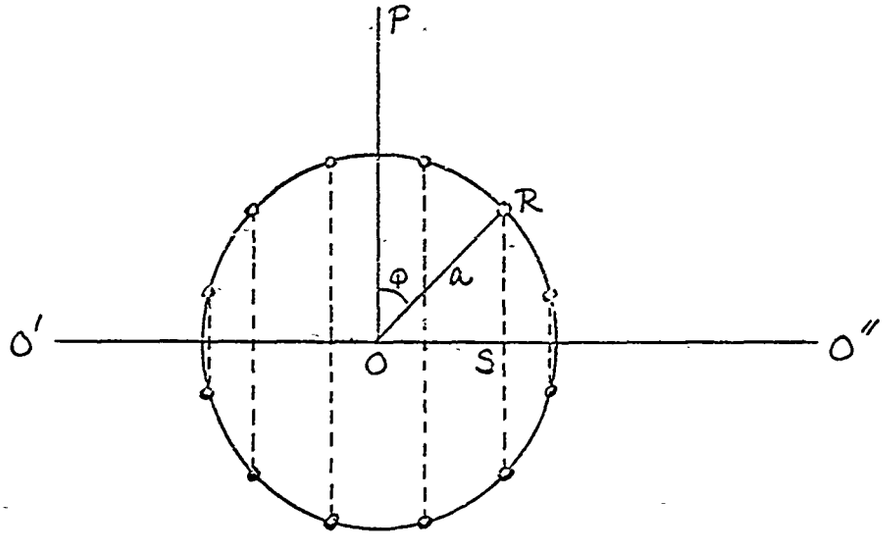
Electrically steered linear arrays are sometimes mounted horizontally and used to determine bearings of targets in azimuth. So long as the sound waves from the target arrive at the array while traveling in a horizontal plane, the bearing determination is accurate. However, if the target is located appreciably below or above the array, so that the waves strike the array at an appreciable angle with the horizontal, the conical nature of the beam pattern introduces significant errors into the bearing measurements.

b. Electrically Steered Circular Array.

A number of important transducer arrays are designed in the form of a circular pattern of vertical staves. Each stave is itself a linear multispot array, and the entire assembly covers the surface of a cylinder. For the present discussion we shall ignore the directivity in the vertical plane and shall consider the array to be a circular arrangement of omnidirectional elements.

If such an array is unsteered, its beam pattern will be similar to that of a continuous ring transducer, discussed earlier. The beam axis will lie along a line through the center of the circle perpendicular to the plane of the circle, and the beam pattern will be symmetric on either side of this plane. The pattern of a continuous circular ring is omnidirectional in the plane of the circle. The corresponding pattern of a circular multispot array is also omnidirectional except for a slight scallop effect due to the finite spacing between elements.

In order to generate a beam in the plane of the array it is necessary to apply a suitable set of delays to the various elements. The delays are the same regardless of whether we consider the array as a projector or as a hydrophone system. In this discussion we shall consider it as a projector. Since there is no a priori preferred direction, we may arbitrarily choose any direction in the plane of the circle as a reference. It is clear that the beam pattern will differ slightly depending upon whether the reference direction passes through a transducer element, or bisects the angle between two adjacent elements, or occupies some intermediate position. For the purpose of the present discussion it will suffice to consider the simple case in which the reference direction passes midway between two elements. The geometry of the situation is shown in the following sketch. The center of the circle is at O and the reference direction in which the array is being steered is OP. Let ϕ be the angle in the plane of the circle which measures the positions of the elements relative to the reference direction OP. If all of the elements are used, the maximum



delay must be applied to the foremost elements (near $\phi = 0$) and the minimum is applied to the rearmost elements (near $\phi = 180°$). Although a negative delay is a physical impossibility, there is no loss in generality if we apply mathematical delays of such an amount as to reduce the phases of all elements to a common value of zero along the line O'O'' at right angles to OP. Mathematically, this calls for negative delays to be applied to the elements in the rear half of the array. Physically there is no change in the array performance if a constant amount is added to all of the delays, sufficient to make all the delays positive.

Reference to the above figure will show that the mathematical delay which must be applied to element R at an angle ϕ is

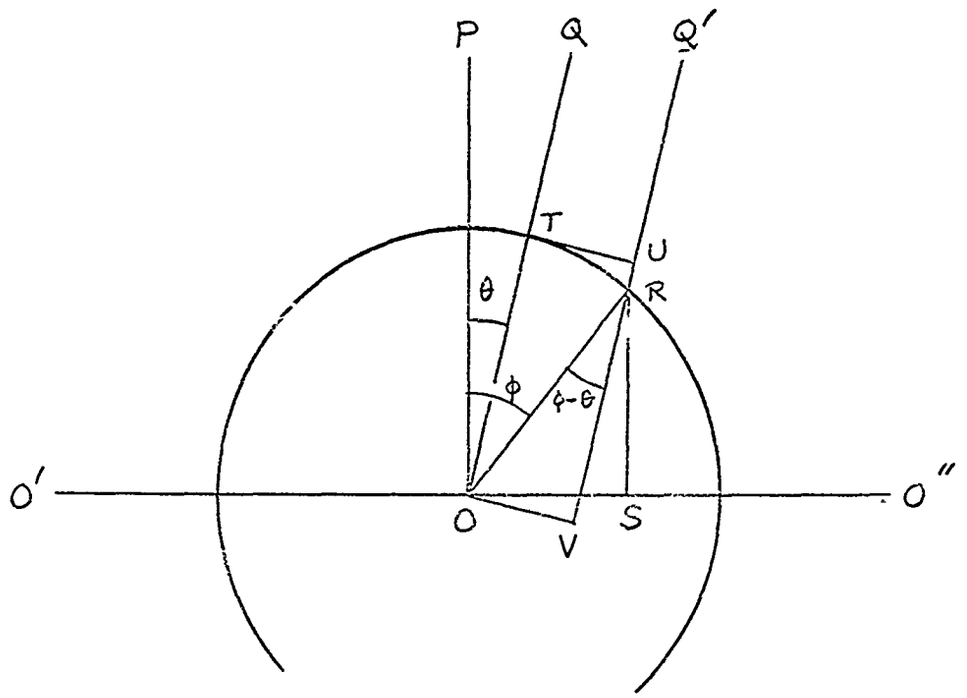
$$\tau = \frac{a \cos \phi}{c} \tag{887}$$

where a is the radius of the circle and c is the speed of sound.
If corresponding delays are applied to all the elements (the delays being proportional to their perpendicular distance from the line O'O''), then as

far as the reference direction OP is concerned, the circular array will be equivalent to a linear array in which all the elements are located at their projections on the line O'O".

It should be pointed out that in forming beams with most circular arrays in practical use, only those elements in the forward semi-circle are used. In fact, the number of active elements is usually restricted to those in a limited arc of about 120 degrees in the front.

Although the compensated circular array is equivalent to a plane array so far as the reference direction is concerned, its behavior in other directions is considerably different. To determine the beam pattern in the plane of the circle, consider the direction OQ making an angle θ with OP, as indicated in the following sketch. We are concerned with the resultant



pressure at a distant point in the direction OQ, due to the individual contributions from all of the elements of the array. According to the usual far-field assumptions, we are concerned only with the relative phases of the pressure waves. It will be convenient to use the distance from O to the far-field point as the reference and to measure all phases in terms of the difference in path length from this reference distance.

Considering an element located at R, we note that the effect of the delay inserted by the compensator is to move this element to its projection S on the line O'O''. The effective path length is therefore from S to R and thence to the distant point along RQ'. If we draw TU tangent to the circle at T, then TU is perpendicular to both OQ and RQ', and the path length to the distant point from T is the same as from U. The difference between the reference path length from O and the effective path length for the element at R is

$$OT - (SR + RU)$$

and the phase angle is obtained by multiplying the difference in path length by the wave number k ($= \frac{2\pi}{\lambda}$). Assuming the radius of the array circle to be a, we find

$$SR = a \cos \phi$$

$$RU = VU - VR = OT - VR$$

$$= a - a \cos (\phi - \theta)$$

The phase angle relative to the origin O is then

$$\psi = k\left[OT - (SR + RU)\right]$$

$$\psi = ka\left[\cos (\phi - \theta) - \cos \phi\right] \qquad (888)$$

Let N = total number of elements in the array

and n = number of elements active in beam

(assumed to be an even number).

Then the angular spacing between adjacent elements is $\frac{2\pi a}{N}$ and the resultant pressure is proportional to

$$\sum_{i=1}^{\frac{n}{2}} \left\{ e^{jka\left[\cos (\phi_i - \theta) - \cos \phi_i\right]} + e^{jka\left[\cos (\phi_i + \theta) - \cos \phi_i\right]} \right\}$$

When θ is zero, the above expression simplifies to n. Hence the relative pressure response is

$$\frac{p(\theta)}{p_o} = \frac{1}{n} \sum_{i=1}^{\frac{n}{2}} \left\{ e^{jka\left[\cos(\phi_i - \theta) - \cos\phi_i\right]} + e^{jka\left[\cos(\phi_i + \theta) - \cos\phi_i\right]} \right. \tag{889}$$

where
$$\phi_i = \frac{(2i-1)\pi a}{N} \tag{890}$$

The computation of beam patterns for specific numerical cases is best done with the aid of a digital computer. If all of the elements in the entire circular array are actively used, a certain amount of simplification of (889) can be accomplished through symmetry considerations. One of the resulting simplifications is the cancellation of the imaginary terms, leading to a phase relationship of 0 or 180 degrees relative to the pressure along the axis. In general, however, when only a portion of the elements, comprising the forward sector, are used, the relative pressure response remains complex, indicating a variable phase relationship as a function of the deviation angle θ. An interesting consequence of this behavior is the absence of true nulls in the beam pattern, since the real and imaginary components of (889) do not both go to zero at the same value of θ.

The preceding analysis was carried out for the special case where the reference direction passes midway between two adjacent elements. Slight differences in patterns are observed at other positions. However, if the number of elements is more than a dozen or so, these local variations are relatively insignificant.

The circular array is admirably suited for electrical steering in azimuth, since it has no strongly preferred direction of symmetry. The directivity pattern varies only slightly as the beam is rotated in azimuth, and it repeats itself every time the axis turns through one element spacing. This means that with very little distortion the beam can be steered to any azimuth orientation. This behavior contrasts sharply with that of the steered linear array.

If each of the omnidirectional elements around the periphery of the circle is now replaced with a vertical stave which is itself

a linear array, the beam pattern in the plane of the circle is unaffected, but greater directivity is achieved in the vertical plane. The geometry of the pattern in three dimensions is rather complicated and will not be analyzed in these notes. It will be noted that the steering of such an array in the vertical plane as well as in azimuth is a rather complicated engineering design problem. When the beam is steered more than about 30 degrees in elevation a noticeable broadening of the beam occurs in the vertical direction.

c. Volumetric Arrays

The arrays thus far considered have consisted of simple arrangements of transducer elements along a straight line, in a plane, or around the circumference of a circle. It is possible to achieve the same amount of directivity in an array of somewhat smaller overall dimensions by means of a so-called volumetric arrangement. In a volumetric array, transducer elements are placed at various pre-calculated positions within the volume of the array, in addition to those located around the outside. Computations must be carefully worked out to determine the proper amount of compensation required for beam steering. An example of a volumetric array might be a circular (actually cylindrical) array in which elements are spaced around one or more smaller concentric circles in addition to the outer circle. In general the design of such a transducer array is sufficiently complicated to warrant the use of a digital computer to determine the required delays, the beam pattern, and the directivity index.

d. Delay Lines and Compensators

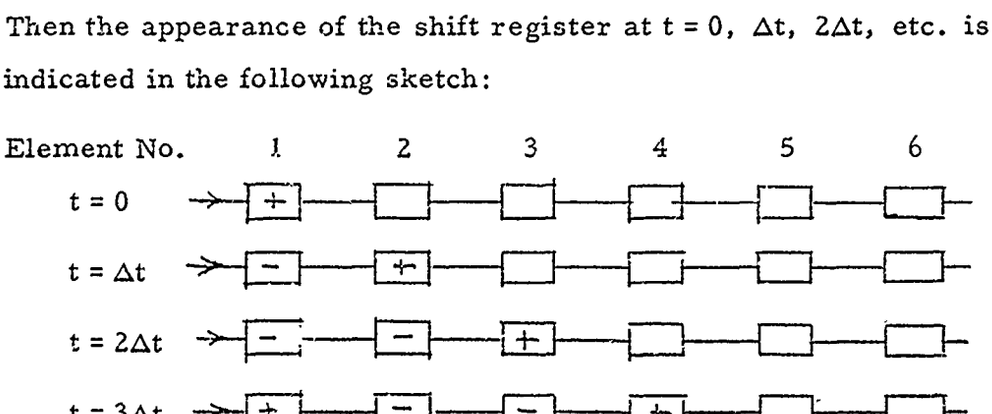
In some transducer applications it is desired to form a finite number of fixed beams by applying suitable delays to the various elements of the array. We shall discuss this subject with reference to the outputs of the hydrophone elements of a receiving array, although the theory also applies in principle also to the input signals to the elements of a projector. Given a number of elements in prescribed physical locations in the array, it is possible to figure out the required delays to be applied to

the outputs of the elements such that when the delayed outputs are connected in phase a beam is formed in a desired direction. By subjecting the outputs of the hydrophones to a number of different delays and properly connecting the outputs it is possible to form a number of fixed beams simultaneously.

Traditionally, delay lines have been constructed from conventional circuit elements (inductors and capacitors). In recent years a somewhat different technique has been developed, called DIMUS (digital multibeam steering). In this method the outputs of all the hydrophones are sampled at a suitable rate (depending upon the bandwidth of the signal) and are then infinitely clipped, leaving only polarity information (no amplitude information). The output of each hydrophone is then fed into a shift register, which is a device containing a large number of flip-flop elements.* The input digital information is transmitted down the shift register from one element to the next at a rate equal to the sampling rate. Suppose that at $t = 0$ a "+" is fed from a hydrophone into the first element of its shift register. After one sampling period Δt this "+" is transferred to the second element, and a new digit is fed from the hydrophone (and clipper) into the first element. Suppose that the sequence of digits (that is, polarities) from the hydrophone is as follows:

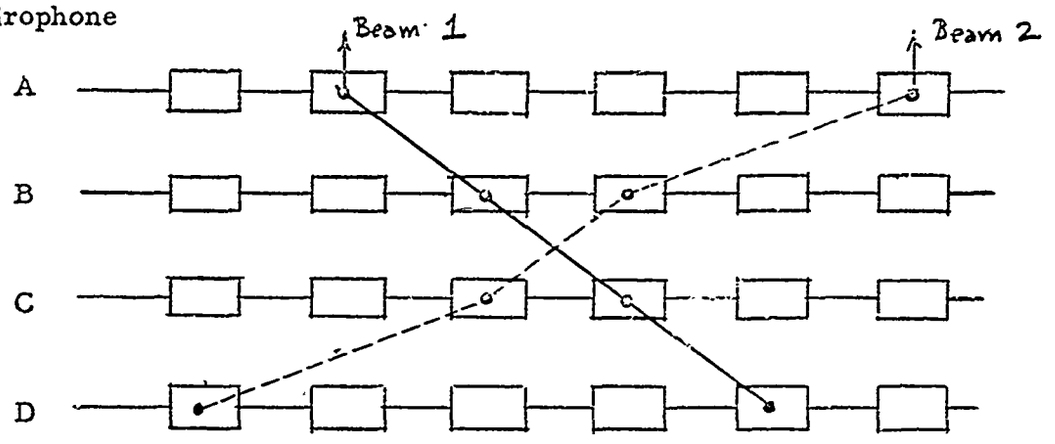
+, -, -, +, -, +, etc.

Then the appearance of the shift register at $t = 0$, Δt, $2\Delta t$, etc. is as indicated in the following sketch:



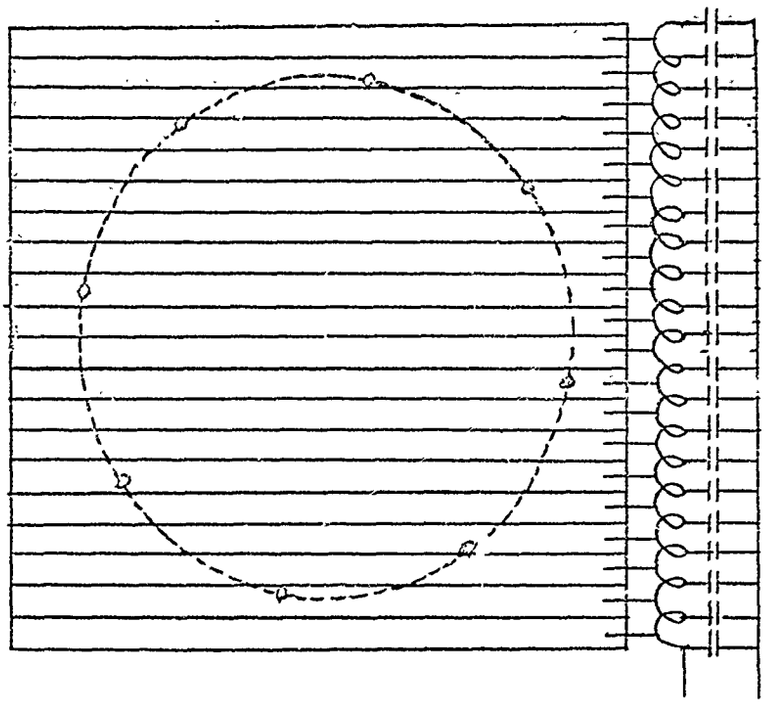
*Usually called stages.

The output of each hydrophone element is fed into a separate shift register. Any desired delay up to the maximum available from the shift registers can be obtained by tapping off from the proper element of the shift register. A beam can be formed by taking a tap off each of the shift registers at the proper point corresponding to the required delay. The outputs are then added together. A number of beams can be formed simultaneously by connecting together the outputs of several combinations of shift register elements. The diagram below shows schematically the formation of two beams from four hydrophone elements and their associated shift registers.



Another device which is particularly useful for training a single beam is the universal compensator. This consists of a flat plate made up of a number of conducting bars connected to successive points along a delay line, as indicated in the sketch below. The time for an electrical signal to move upward from the bottom to the top of the delay line is the same as the time required for an acoustic wave to travel across the transducer array.

On top of the fixed plate there rests a rotatable commutator arrangement, built as a scale model of the actual array, with a wiper contact corresponding to each transducer element of the array, and connected electrically to that element. The maximum delay corresponds to



the topmost portion of the fixed plate shown in the sketch. Hence the vertical direction (pointing upward on the paper) corresponds to the direction of the sonar beam. As the commutator is rotated over the surface of the fixed plate, the sonar beam is rotated electrically relative to the fixed hydrophones installed in the array.

In the example shown above the array consists of a symmetric circular arrangement of eight hydrophones. It is not necessary that the array be symmetric. Compensators can be built to work with irregular arrays tailored to meet the individual requirements of particular installations.

e. Frequency-Dependent Compensation

So long as the delay introduced is independent of the acoustic frequency, the compensation angle of the major lobe is also independent of frequency. There are, however, compensator systems in which the delay is inversely proportional to the frequency. In such transducers the position of the major lobe varies with the frequency.

If the transmitted signal consists of a continuous spectrum, it is possible to echo-range simultaneously over a large sector, since each bearing in that sector corresponds to the location of the maximum response axis for some particular frequency. Thus, when an echo is received, the bearing of the target can be determined from the frequency of the echo.

3. Bearing Deviation Indicators.

In principle the bearing of a source of sound or of an echo returned from a target can be determined by steering the receiving array until the axis of the main lobe points in the direction from which the sound is coming. This type of procedure is commonly employed in making various sorts of physical measurements in which the maximum of some function is sought. When it is applied to sonar, however, serious problems arise from the fact that the received signal is subject to very large fluctuations. These fluctuations, as we have learned previously, are due to a number of causes and include both short-term variations in transmission loss and longer term interference effects such as the Lloyd mirror effect. For these reasons it is in general impractical to determine the bearing of a sound source by steering the beam until a maximum response is obtained. In echo ranging the difficulty is even worse.

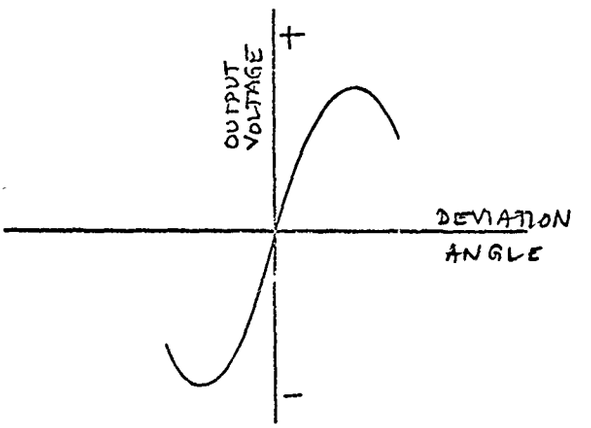
For these reasons it has been necessary to develop a device called a bearing-deviation indicator which is specially designed to yield bearing indication. Several types are in use, two of which are described below.

a. Amplitude - Difference Method.

To illustrate the principle, let us assume that the transducer system consists of two similar transducers mounted one above the other on a common vertical shaft. The two transducers are mounted in azimuth positions relative to each other such that their maximum response axes are separated by a small angle called the separation angle. The outputs of the two transducers are connected in series opposition, so that the system indicates the difference between the two transducer voltages.

Let the two halves of the transducer be designated as A and B. If the transducer is oriented such that the maximum response axis of transducer A coincides with the direction of the incoming signal, then A will give a larger indication than B. On the other hand, if the direction of the incoming signal coincides with the maximum response axis of B, then B will give a larger indication than A. If the transducer is oriented so that the direction of the incoming signal is midway between the axes of A and B, the output will be zero.

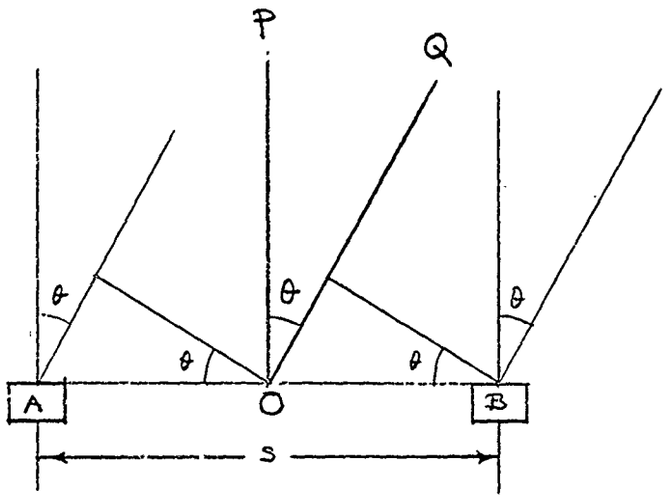
The behavior of the differential output as a function of the bearing deviation angle is shown in the sketch below. The output is zero for zero deviation and has a positive or negative value depending upon whether the deviation is to the right or left of the true bearing.



A more practical arrangement is one in which a single transducer is operated with two different electric compensators arranged so that the two maximum response bearings differ by a few degrees. The differential output of the two systems furnishes a source of bearing information similar to that described in the preceding paragraph.

b. Phase-Differential Method.

In this method the transducer is split into two halves A and B, the outputs of which are compared in phase. For sound waves coming in the direction OP which is at right angles to the line joining A and B, the waves arrive at A and B simultaneously and there is no difference



in phase. If, on the other hand, the sound is coming from the direction OQ, then the distance traveled to B is less than the distance traveled to A by an amount s sin θ. Hence there is a difference of phase

$$\Delta\psi = \frac{2\pi}{\lambda} s \sin\theta$$

where s is the separation between A and B. By noting whether $\Delta\psi$ is positive or negative, one can tell whether the source is to the right or to the left of OP.

The advantage of these bearing-deviation indicators over the simple method of training a single beam to the maximum response bearing lies in the fact that in each case the two outputs which are compared are obtained from acoustic transmissions which have simultaneously traversed the same acoustic path, thus eliminating most of the problems of variability which cause trouble in the search for the bearing of maximum response.

INDICATORS AND RECORDERS; DETECTION OF SIGNALS

We have discussed the processes by which acoustic signals in the water are transformed to electrical signals in the receiving equipment. In order to be useful, these signals must be further transformed to a form in which they may be recognized by a human operator. Instruments which perform this function are called indicators and recorders. An indicator exhibits a transient response which is perceptible only while the signal is being received. A recorder makes a permanent record of certain of the pertinent characteristics of the signal.

A. Types of Indicators and Recorders.

1. Audible Presentations.

In this type the signal is presented to the operator in the form of audio waves in air, normally by means of headphones. Such instruments are obviously indicators. Most recorders are of the visible rather than the audible type, but audible records can be made with a tape recorder.

The audible presentation is the most logical way of presenting acoustic signals to the operator, and in many applications it has a great advantage over the visible type. In direct listening, where it is desired to recognize the sound of a particular object such as a ship, against the general background interference, the ear of a trained operator possesses a unique capability. For underwater voice communications, audible presentation is clearly the only practical type. In echo-ranging against a target which is moving toward or away from the sonar, the returning echo experiences a change of frequency called the Doppler shift. The human ear is very sensitive to this change of frequency and has an exceptionally good capability of picking such an echo out of a reverberation background.

2. Visual Presentations. Most visual presentations are of either of two types--the paper-recording type or the cathode ray type.

a. Variable Intensity Recorders.

The first widely used visual recorder was the Sound Range Recorder, developed by the British during the early part of World War II. The

paper used in this recorder is chemically treated and is in the form of a long strip wound on a roll from which it is drawn at constant speed. The paper passes between a fixed metal plate and a metal stylus which moves across the paper at right angles to the direction in which the paper moves. When a ping is sent out, the stylus begins to move across at constant speed, so that its position at any instant is proportional to the round trip distance of the sound wave out to the point from which the acoustic energy is scattered back to the transducer. The electric current which flows between the stylus and the plate causes a darkening of the paper in proportion to the received acoustic power. The presence of a target echo would thus appear as a dark spot on the trace. This type of indication is called variable-intensity ication.

When an echo is present, the range to the target can be found by measuring the distance from the beginning of the trace to the dark spot representing the echo. One advantage of this type of presentation is that when a number of pings are sent out in sequence, the dark spots representing the echoes on successive traces tend to form a line, enabling the eye to carry out a sort of visual integration. Whereas a single spot may have been too weak to detect, the presence of a line of spots enhances the recognition.

Other recorders of this type have been devised to indicate other information. For example, there are recorders used for passive listening, in which the position of the stylus is proportional to bearing rather than range. If a ship or other source of noise is present on any given bearing, the stylus will produce a dark spot at the corresponding point on its sweep. After a large number of sweeps, the record of the ship will appear as a line on the paper.

b. Variable-Displacement Recorders.

Recorders in which the displacement of the stylus is proportional to the intensity of the received acoustic radiation (or is some other function of the intensity) are called variable-displacement recorders. A recorder of this type is used in an important airborne application. This type is not suitable for the visual integration associated with the repetition of the signal. Its principal advantage is in applications where the signal can be recorded but once. It provides a more quantitative and objective record of the received acoustic radiation, enabling a more reliable estimate to be

made of the presence of the signal.

One disadvantage of the variable displacement type of presentation is that one of the two dimensions of the presentation must be devoted to intensity, thereby restricting the amount of information which can be presented.

c. Variable-Intensity Cathode Ray Presentations.

A common type is the plan position indicator, similar to the corresponding type of radar scope, in which range is measured radially outward from the center and azimuth is measured by the polar angle about the center point. The brightening of the screen at any point is proportional to the amount of acoustic energy received from the particular range and bearing. An echo therefore appears as a bright spot. In a conventional radar, pulses are continually sent out and the back-scattered electromagnetic energy is received as the antenna sweeps around in azimuth. The appearance on the scope is that of a bright radial line which sweeps around the circle, leaving bright spots wherever returned energy is received. The pattern gradually fades away and is renewed by the next sweep. Because of the slow speed of sound in water, this type of operation is not practical in a sonar P.P.I. Instead, it is desirable to employ a scanning type of sonar in which the outgoing pulse is transmitted omnidirectionally in azimuth and the receiving equipment is designed to scan through 360°.

d. Variable-Displacement Cathode Ray Presentations.

In this type the horizontal sweep is controlled by either range or azimuth and the vertical displacement is proportional to the acoustic intensity. Such a scope is called an A-scope.

One of the principal functions of visible presentations is to provide quantitative data, such as range, bearing, Doppler speed, etc. Although the ear is a marvelous instrument for detecting and classifying sounds, auxiliary instruments are needed for making precise quantitative measurements. The continuing trend toward the development of more sophisticated signal processing systems in modern sonars is placing greater and greater emphasis upon visual presentations.

B. Signal Detection.

1. General Remarks Concerning the Observation of Acoustic Signals.

The type of display with which an acoustic system is provided

depends upon the purpose of the acoustic observation or measurement to be made. Large sonar systems are usually designed in a number of different modes and are therefore provided with a number of different displays--both indicators and recorders, and both aural and visual presentations.

Consider, for example, a typical sequence of operations involved in anti-submarine warfare. The first event which must take place is detection. Second, the target must be classified, that is, a decision must be made as to whether the detected target is or is not a submarine. Third, the target must be tracked for a sufficient period of time to permit the attacking vehicle to get into a favorable attack position. Fourth, the target must be localized, so that its position is known with sufficient accuracy to permit a weapon to be launched. Each of these operations imposes its own requirements on the nature of the sonar information to be presented and the manner in which it is displayed.

The principal requirement for detection is the ability to identify the presence of a weak signal in the midst of ever-present background interference. Since the ear is a very sensitive detector of threshold signals, aural presentations are valuable for the detection phase. Another type of presentation which is particularly useful for passive sonars is the variable-intensity record of the bearing (or some other characteristic of the target signal) versus time, as has been mentioned previously. The visual integration performed by the eye in sighting along the lines formed by repeated sweeps is of significant value in enhancing the detectability of weak signals. Another technique, which is widely used in active sonars, is to transmit special types of pulses designed to operate in conjunction with signal processing systems. The detectability of a returned signal is enhanced if it is correlated with a stored replica of the transmitted pulse.

In classifying a target we search for characteristics of the target signal which distinguish it from the signals from other objects. The ear is extremely valuable for this purpose, especially in passive sonars where special characteristics of the signal, such as the rhythmic beat of propellers, may frequently be identified. Another important characteristic of the signal generated by a target is its spectrum. In the case of active sonars, clues as to the identity of a target can be obtained from various characteristics of the echo, such as its duration, Doppler shift, and detailed structure. For

example, if a suspected echo has an extremely short duration, it is very likely not an echo but a noise spike, since even though the original pulse might have been quite short, the echo is lengthened by the finite dimensions of the target. Also, if the transmitted pulse is sufficiently short, the returned echo will tend to reveal highlights corresponding to the major reflecting surfaces of the target. Doppler information is useful in several ways, such as in ping-to-ping correlation, where the change in range between pings should equal the product of the Doppler speed and the elapsed time.

In tracking operations the principal object is to keep the target within range of the sonar at all times. A bearing deviation indicator is useful for tracking with passive sonar, since it provides continuous information on the bearing of the target. Passive range information is more difficult to come by. Information obtained from the intensity of the received signal can at best provide only the crudest idea of the range, since the source level of the target is in general not known, and since the propagation loss depends on a number of variable factors and is therefore not a reliable measure of range. There are, however, a few phenomena available from which a certain amount of range information can be obtained, though we shall not discuss them. Historically submarines have widely used a method of tracking involving a combination of bearing information and maneuvering of the tracking vehicle, the so-called bearings-only method.

Active tracking has the obvious advantage of accurate range information, but the disadvantage of fully alerting the target. An important requirement in active tracking is a high data rate. Ideally such tracking is best accomplished when the tracking vehicle moves continuously through the water with the target, so that a continual flow of data is available. If sonobuoys are used, the buoys must be dropped at advanced positions in anticipation of the target. The problem is most severe in the case of tracking by a single dipped sonar, where the tracking vehicle is blind (more accurately, deaf!) during the interval between successive dips, and the target is most likely aware of this fact. These are essentially operational problems which are beyond the scope of these notes.

The highest accuracy in range and bearing is required in the final localization prior to weapon launch. This is an operation which clearly requires active sonar (or possibly other, non-acoustic sensors) and which demands

the minimum time late.

Of all the functions discussed above, the detection function places the highest requirement upon the ability of the sonar operator to pick weak signals out of the background interference. Classification normally requires a higher signal-to-noise ratio, since more information concerning the target is needed; tracking and localization require a still higher signal-to-noise ratio. For this reason the most sensitive displays are required for the detection phase, and in the discussion which follows we shall be concerned chiefly with the problem of detection. We shall first discuss the conventional approach based on aural detection of signals, and then a modern approach based on statistical decision theory.

2. Signal-to-Noise Ratio; Signal Differential; Observed Differential.

Let W be the input bandwidth of the sonar system and let S and N denote the signal power and the noise power respectively in the band W at the output of the hydrophone array, which is the input to the detection system. Then the input signal-to-noise ratio is S/N, and the input signal differential is the decibel equivalent of the signal-to-noise ratio,

$$\Delta L_{S/N} = L_S - L_N = 10 \log \frac{S}{N} \tag{901}$$

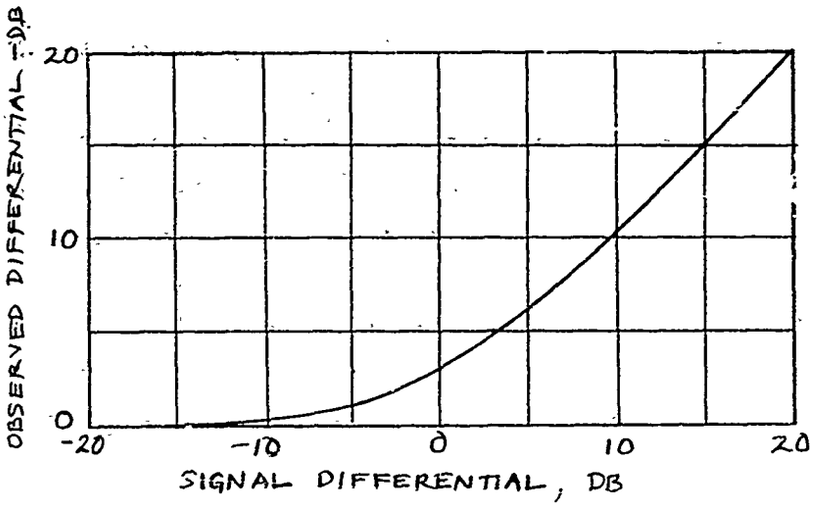
It should be noted that the signal can never be observed alone without the noise being present. If the signal is present, the observed power is S+N; if the signal is not present, the observed power is N. Hence the observed differential is

$$\begin{aligned} \Delta L_{obs} &= 10 \log \frac{S+N}{N} \\ &= 10 \log \left(1 + 10^{0.1 \Delta L_{S/N}} \right) \end{aligned} \tag{902}$$

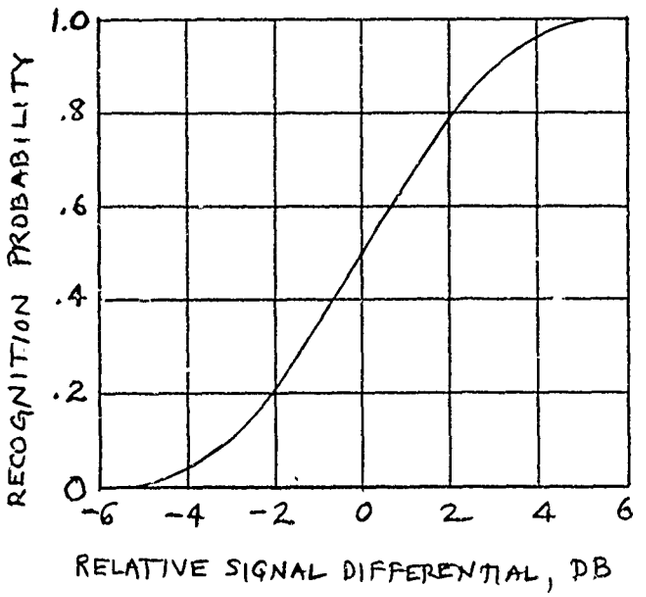
The relationship between the observed differential and the signal differential, expressed by (902), is shown in the following graph. Note that when the signal power is equal to the noise power, the signal differential is zero and the observed differential is 3 db.

3. Recognition Differential.

It has long been known from tests of human operators on both aural and visual presentations that detection is practically certain when the signal-to-noise ratio is high and practically impossible when the signal-to-



noise ratio is low, and that there is a relatively narrow transition region between these two extremes in which the recognition probability changes from 0 to 100 percent. This transition region is illustrated in the following graph in which the probability of recognition is plotted against a variable labeled "relative signal differential". The curve is based on the average response of a large number of human operators tested under controlled laboratory conditions. It is clearly this region of transition which is of major interest in the investigation of the dectability of signals.



The zero point of the relative differential scale has been set arbitrarily at the point where the probability is 50 percent. The actual value of the signal differential at this point is called the recognition differential. It may be expressed mathematically as follows:

$$M = L_{50} - L_N \tag{903}$$

where M = recognition differential, db,

L_{50} = signal level required for 50% probability of recognition,

L_N = level of interference, both levels being expressed in db referred to the same reference.

The numerical value of M depends upon the nature of both the signal and the interference, upon the type of indicating system, and upon the bandwidth of the system. It should be noted that a large positive value of M means that a large signal is required for recognition. It therefore signifies poor recognition. On the other hand, a small value of M signifies good recognition. We shall shortly consider this subject in more detail, chiefly in connection with echo-ranging. For the present a few general remarks will suffice.

When the signal and the interference are alike in character, as is sometimes the case in direct listening, the only indication that a signal is present is obtained from the magnitude of the total response of the system. If, for example, the signal begins at a very low level and increases gradually, it is apt to go unnoticed until it becomes very large. Under these conditions the recognition differential is high. If, however, the signal either starts or ends abruptly, so that a sudden change occurs in the output (i.e., in the observed differential), the signal is more easily recognized. Under these conconditions the observed differential corresponding to 50% probability of aural recognition is approximately 3 db, which corresponds to a recognition differential of 0 db.

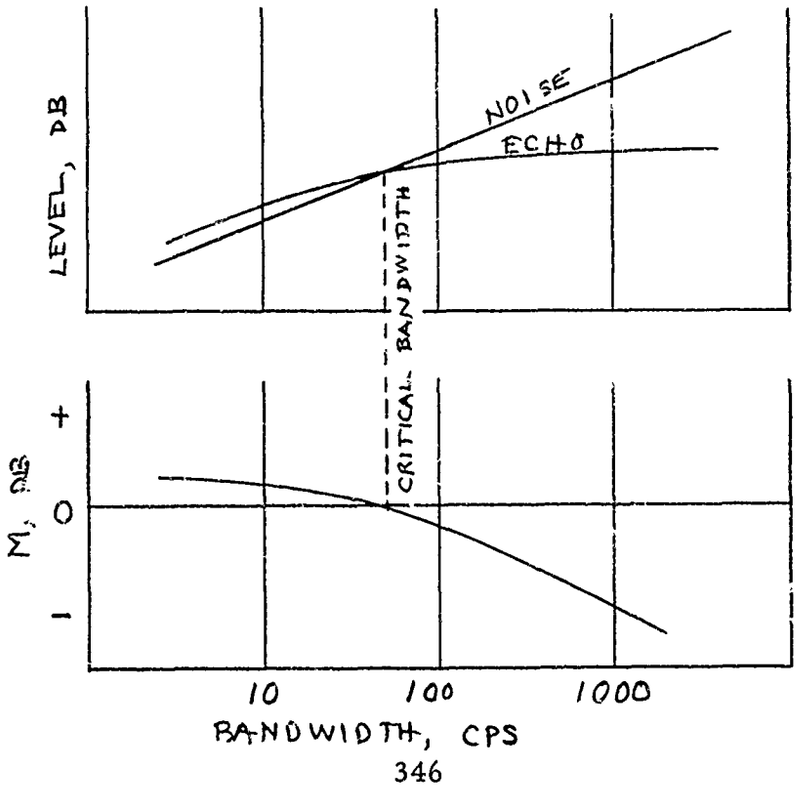
When the signal possesses certain readily distinguishable characteristics, it may be easily recognized even though it is continuous, without any sudden starting or stopping. In such cases the recognition differential may be very small or even negative, depending upon the nature of the signal.

In echo-ranging with conventional sonar the signal is an echo, which is a short pulse having a fluctuating amplitude, due to the characteristics of the target and the fluctuations of sound transmission in the sea. The background against which the signal must be recognized consists of both reverber-

ation and wideband noise. Under some conditions the noise is dominant and under other conditions the reverberation is dominant. The recognition differential in the two cases is different.

Aural Recognition Differential for Signals in Ambient Noise Background. The ear is most sensitive to frequencies in the vicinity of 800 cps and for this reason it is customary to reduce the sonar frequency (which may be as high as 30 kc) to about 800 cps by means of a heterodyne circuit. Another important characteristic of the human ear is its ability to filter out broadband noise when listening for a pure tone. Tests have shown that only a small range of frequencies in the vicinit of the tone contribute a masking effect, even though a much wider band is presented to the ear. This range of frequencies is called the critical bandwidth of the ear and amounts to about 50 cps. All frequencies outside this bandwidth are automatically rejected.

The numerical value of the aural recognition differential for pulsed CW active sonars depends upon the bandwidth of the system and upon the duration of the pulse, or ping length. The effect of bandwidth for a ping of relatively long duration is illustrated in the two sketches below. The upper sketch shows qualitatively as a function of the bandwidth the total noise level in the band



and the required signal (echo) level for 50% recognition. The lower sketch shows the recognition differential, which is the difference between these two levels. It will be seen that while the noise level rises steadily with increasing bandwidth (due to the 10 logW term) the required signal level for 50% recognition rises at first and then levels off.

At the critical bandwidth of the ear the signal and noise levels are equal and the recognition differential is zero, indicating that the ear can just pick a tone out of the background when the total power in each is the same.

When the bandwidth exceeds the critical bandwidth, the additional noise does not produce any further masking since the ear, in effect, filters out this noise. Therefore even though there is more noise present, no increase in echo level is required. The negative value of recognition differential merely indicates that more noise is present, not that less signal strength is needed.

When the bandwidth is smaller than the critical bandwidth, the ear begins to lose the sensation of tone and both echo and noise are heard as a single blended sound, and the recognition of the echo must be made principally on the basis of loudness alone. Because of this behavior of the ear it is desirable that the bandwidth employed in aural presentation systems be at least as wide as the bandwidth of the ear.

In defining the recognition differential of pulsed CW aural systems it is often convenient to include the bandwidth of the ear in the definition, so that the recognition differential is defined as the difference between the signal level and the noise spectrum level. Let M' denote this modified definition. Then

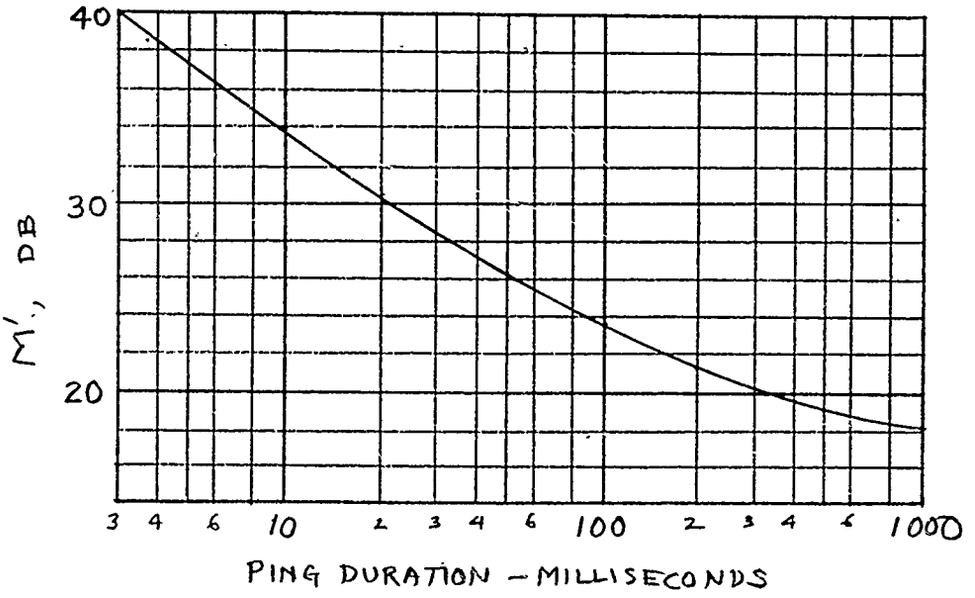
$$M' = L_{50} - L_{N_o} \tag{904}$$

where L_{N_o} is the spectrum level of the noise at the signal frequency. M' is the amount which must be added to the noise spectrum level to obtain the required signal level for 50% probability of recognition. If M denotes the recognition differential based on the total noise in the critical band W_c, the relationship between M and M' is

$$M' = M + 10 \log W_c \approx M + 17 \text{ db} \tag{905}$$

Effect of Ping Length on Recognition Differential.

Experience has shown that it is more difficult to recognize short pulses of single-tone sound in a background of wideband noise than to recognize longer pulses. A typical curve of recognition differential (based on noise spectrum level) vs. ping length is shown in the accompanying graph. It is seen that for pings shorter than about 200 ms (milliseconds),



the recognition differential depends very strongly upon ping length. As the ping length is increased above 200 ms, the recognition differential curve levels off and reaches a steady value for pings of about 1 second duration. The ear responds to a 1 second pulse as well as it responds to a steady tone.

The experimental curve shown above can be fitted fairly well with the following equation

$$M' = 45.5 - 11.7 \log \frac{\tau}{1+0.0036\tau} \text{ db} \qquad (906)$$

where τ is the ping length in milliseconds. The asymptotic value of M' obtained from (906) for very long ping lengths is 17 db. The linear portion of the curve for ping lengths less than about 100 ms can be fitted with the simple equation

$$M' = 43 - 10 \log \tau \qquad (906a)$$

The recognition differential for the case where the background interference consists of reverberation will be discussed in a later section.

4. Statistical Detection Theory.

The steadily increasing complexity of ASW operations has led to a developing trend toward automation in order to relieve human operators of as many burdens as possible. An example of this trend is the interest in automatic detection systems which monitor the output of the sonar and set off an alarm whenever a target is detected. The concept of recognition differential, which was developed with reference to signal detection by human beings, is largely subjective and empirical and is not adequate for automatic detection systems. What is needed is a completely objective theory of the detection of signals in the presence of random noise. A considerable body of theory has been developed by a number of workers in the fields of radar and communications over the past two decades and is gradually being applied to sonar systems. One of the most useful approaches has been that of Birdsall and Peterson*, which seeks to define the basic mode of operation and to evaluate the performance of a so-called "optimum receiver" on the basis of what is known about the characteristics of the signal and the statistical properties of the background interference. An optimum receiver is an ideal device which operates in such a way upon the input presented to it as to yield the theoretically best possible performance consistent with the input signal-to-noise ratio. The theory does not evaluate the performance of any particular hardware, but rather the optimum performance against which practical, non-optimum receivers can be rated, so that one may know how much could be gained by further improvement.

The theory was originally worked out for a few types of signals in an ideal background of white Gaussian noise. Although it has been extended somewhat in scope, the mathematical complications become severe as more realism is incorporated into the model, and progress is slow. However, two of the original cases are of considerable importance and will be discussed below. Inasmuch as an adequate presentation of the required background of probability and statistics is beyond the scope of these notes, the derivation of the results will be somewhat sketchy.

- - - - - - - - - - - - -

*W. W. Peterson and T. G. Birdsall, "The Theory of Signal Detectability," Engineering Research Institute, University of Michigan, Technical Report 13, June 1953.

a. The Decision Process.

The detection of signals in the presence of a background of random noise is essentially a decision process. Given an input consisting of noise which may or may not contain the desired signal, the decision maker, whether it be a human being or a piece of electronic equipment, is required to decide whether a signal is present or not. Since the noise is a random process, the problem is statistical in nature and there is always a finite probability of making an error. There are two possible types of input--either the signal is present or it is not--and two possible decisions--"yes", it is there, or "no", it is not. Therefore on any trial one of the following four possible events must occur:

(1a). Correct detection -- signal present; decision "yes"
(1b). False rest -- signal present; decision "no"

(2a). False alarm -- signal not present; decision "yes"
(2b). Correct rest -- signal not present; decision "no"

It is seen that (1a) and (1b) are mutually exclusive, exhaustive events. Given that a signal is present, one or the other must occur, but if one occurs, the other cannot occur. Therefore the sum of the probability of saying "yes" and the probability of saying "no" is unity, so that if one of the two probabilities is known, the other may be computed from it. The same is true of (2a) and (2b) when the signal is not present. Thus, there are really two probabilities involved in the detection process--the probability of a (correct) detection P(D) and the probability of a false alarm P(FA).

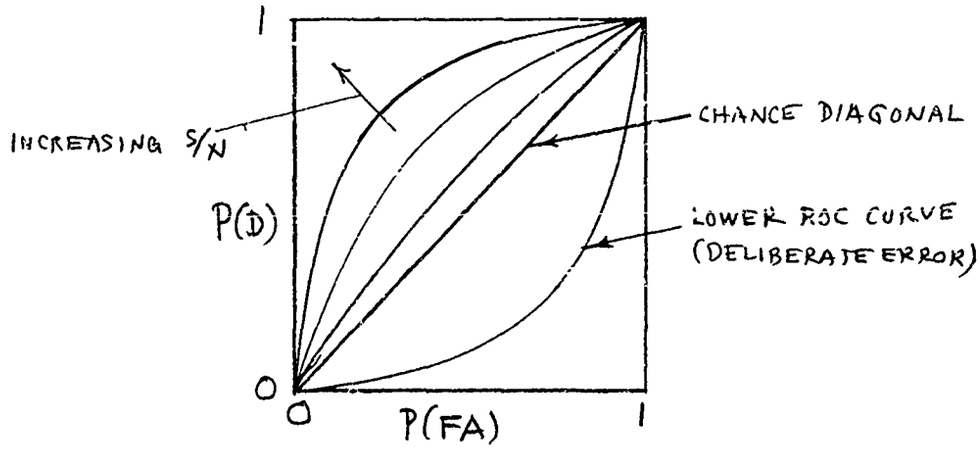
b. ROC Curves.

To see how these two probabilities are involved, let us assume that the receiver measures some property of the input supplied to it, such that the probability of detection is a monotonically increasing function of the measured value of this property (i.e., when the measured value increases, P(D) also increases). The receiver also contains a decision device which operates in such a manner that when the measured value exceeds a certain preset threshold the decision is "yes", while if the threshold is not exceeded, the decision is "no".

We now see that the probabilities P(D) and P(FA) are functions of both the threshold setting and the input signal-to-noise ratio. Consider first the effect of the threshold setting. If the threshold is set

extremely high, it will rarely be exceeded, regardless of whether a signal is present or not; hence P(D) and P(FA) are very small. At the other extreme, if the threshold is set very low it will be exceeded virtually all the time and P(D) and P(FA) are both near unity. Intermediate values of the probabilities occur for intermediate settings of the threshold. Consider next the effect of the input signal-to-noise ratio. If the signal-to-noise ratio is exceedingly small, there is practically no difference whether the signal is present or not, and both probabilities are virtually identical, regardless of the threshold setting. If, on the other hand, the signal-to-noise ratio is large, the threshold will be exceeded more frequently when a signal is present than when it is not; hence P(D) exceeds P(FA). The larger the signal-to-noise ratio, the greater will be the difference between these probabilities and the more unambiguous the basis for decision.

If the probability of detection is plotted against the false alarm probability and the decision threshold is varied for fixed values of the input signal-to-noise ratio, the curves of the type illustrated below are obtained. These are called ROC (Receiver Operating Characteristic) curves. The



straight line corresponding to zero signal-to-noise ratio is called the chance diagonal. This represents the poorest possible performance short of deliberately making the wrong choice; it is the performance which would be obtained simply by flipping a coin. When the signal-to-noise ratio is small but finite, the ROC curve lies above the chance diagonal, indicating that at all threshold settings (except the limiting values of $-\infty$ and $+\infty$) the detection probability is higher than the false alarm probability. As the signal-to-noise ratio is increased, the ROC curves move upward away from the chance diagonal, indicating

more favorable performance. The ideal goal of the upper left-hand corner P(D)=1; P(FA)=0 cannot be achieved unless the signal-to-noise ratio is infinite.

It should be noted that for any given signal-to-noise ratio the performance of all possible receivers lies between the appropriate ROC curve and its image curve symmetric about the center of the square. The region below the chance diagonal is of course of no interest, since it represents purposeful errors. (It is interesting, however, to observe that it is impossible to do any worse than the lower curve, because not enough information is available to enable a more consistently wrong choice to be made!) An optimum receiver operates on the upper ROC curve, whereas all practical, non-optimum receivers operate somewhere in the region between the ROC curve and the chance diagonal. All such receivers can theoretically be improved to the point where the ROC curve is reached, but not beyond.

In summary, the particular ROC curve on which an optimum receiver operates is determined by the input signal-to-noise ratio. The point on the curve at which the receiver operates is determined by the threshold setting. If, for example, the optimum receiver were adjusted to yield a constant false alarm probability of 0.001, the operating points on all ROC curves would lie on a vertical line

$$P(FA) = 0.001$$

and the probability of detection would be an increasing function of the signal-to-noise ratio.

c. Likelihood Ratio.

It has been stated that in order to make a decision, a receiver must measure "some property" of the input which is suitably related to the probability of detection. Birdsall and Peterson show that the property which should be measured by an optimum receiver is the likelihood ratio, which is the ratio of the conditional probability density of the observed input to the receiver under the hypothesis that the signal is present, to the conditional probability density of the same under the hypothesis that only noise is present.

An explanation of what is meant by these probability densities may be helpful at this point. A typical input to the receiver might consist of a continuously varying random noise voltage to which is added a

continuously varying signal voltage. We shall assume that the wave lasts for a time T and is band-limited to a bandwidth W. Since the noise is a random process, any discrete sample of the noise voltage, taken at any instant, is a random variable. If x_i denotes the i^{th} sample, the probability density $p(x_i)$ is a function such that $p(x_i)dx_i$ represents the probability that the voltage has a value between x_i and $x_i + dx_i$.

According to the sampling theorem it can be shown that if the wave is sampled at regular intervals of 1/2W second, it can be completely reconstructed from the 2WT samples, so that the wave may be described interchangeably either as a continuous function of time $\mathbf{x}(t)$ or as a sequence of random variables, $\mathbf{x}_1, x_2, \cdots, x_n$, where

$$n = 2WT \tag{907}$$

Furthermore, when sampled at this rate the samples are all statistically independent, so that the joint probability density function of all n samples is the product of the n individual density functions,

$$p(\mathbf{x}) = \prod_{i=1}^{n} p(x_i) \tag{908}$$

where x stands symbolically for the combination of all n values $x_1, \cdots, x_n$.

Suppose now a wave $\mathbf{x}(t)$ is presented to the receiver. It is not known whether the wave consists of signal-plus-noise or of noise alone. Let $p_N(x)$ denote the conditional probability density function under the hypothesis that only noise is present, and $p_{SN}(x)$ denote the conditional probability density function under the hypothesis that the wave consists of signal plus noise. Each of these functions is a product of individual functions like (908) except, of course, that the functions have different values depending upon which hypothesis is assumed. The likelihood-ratio $\mathcal{l}(x)$ is defined as

$$\mathcal{l}(x) = \frac{p_{SN}(x)}{p_N(x)} \tag{909}$$

An optimum receiver, if it is to yield the optimum performance predicted by the theory, must be able to compute the likelihood ratio corresponding to each input wave presented to it. The receiver is provided with a threshold β such that the decision is "yes" (alarm) whenever $\mathcal{l}(x) \geq \beta$ and "no" whenever $\mathcal{l}(x) < \beta$.

d. Explicit Results (White Gaussian Noise).

In order to obtain a practical solution it is necessary to have explicit expressions for the probability density functions. In the following paragraphs we shall derive the results for the simple cases of two types of signals in white Gaussian noise. "White" means that the spectrum level of the noise is a constant, independent of frequency over the band W. "Gaussian" means that each of the individual samples has a Gaussian (or normal) distribution,

$$p_N(x_i) = \frac{1}{\sqrt{2\pi N}} e^{-\frac{x_i^2}{2N}} \tag{910}$$

where N is the variance of x_i, which is proportional to the noise power. (The noise voltage is assumed to have no d.c. component, so that its mean is zero.)

The joint probability density function under the hypothesis of noise only is, from (908),

$$p_N(x) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi N}} e^{-\frac{x_i^2}{2N}}$$

$$= \frac{1}{(2\pi N)^{n/2}} e^{-\frac{1}{2N}\sum_{i=1}^{n} x_i^2} \tag{911}$$

Except under the simplest assumptions the computation of the likelihood ratio becomes very difficult. We shall limit our discussion to two simple but important cases: (1) signal known exactly and (2) signal consisting of white Gaussian noise.

Case 1: Signal Known Exactly.

Let x_i denote the voltage of the i^{th} sample as before, and let s_i denote the contribution of the signal and n_i be the contribution of the noise, so that

$$x_i = s_i + n_i \tag{912}$$

Now, if the signal is known exactly, each of the s_i is a known constant, and only the noise n_i contributes to the randomness of x_i. Hence the probability density function under the hypothesis that a signal is present is

$$p_{SN}(x) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi N}} e^{-\frac{n_i^2}{2N}}$$

$$p_{SN}(x) = \frac{1}{(2\pi N)^{n/2}} e^{-\frac{1}{2N}\sum_{i=1}^{n}(x_i - s_i)^2} \tag{913}$$

The likelihood ratio is the ratio of (913) to (911),

$$\ell(x) = \frac{\frac{1}{(2\pi N)^{n/2}} e^{-\frac{1}{2N}\sum_{i}(x_i - s_i)^2}}{\frac{1}{(2\pi N)^{n/2}} e^{-\frac{1}{2N}\sum_{i} x_i^2}}$$

or

$$\ell(x) = e^{-\frac{1}{2N}\sum_{i=1}^{n} s_i^2} e^{+\frac{1}{N}\sum_{i=1}^{n} s_i x_i} \tag{914}$$

From (914) we see what type of operation an optimum receiver should perform. It must compute the sum appearing in the exponent of the second factor, which, from the sampling theorem, is

$$\frac{1}{N}\sum_{i=1}^{2WT} s_i x_i = \frac{2W}{N}\int_{0}^{T} s(t)x(t)dt \tag{915}$$

This is the correlation function of the input $x(t)$ and the known signal $s(t)$.

The first factor of (914), according to the Case 1 assumption, is a known constant. In fact, each term s_1^2 represents the square of a sample of the signal voltage, which is proportional to the signal power. (If we assume that the voltage is applied across a resistance of one ohm, the square of the voltage may be considered to be equal to the power.) Since there are $n = 2WT$ samples, the average signal power S is

$$S = \frac{1}{2WT}\sum_{i=1}^{n} s_i^2 \tag{916}$$

and the signal energy is

$$E = S\,T = \frac{1}{2W}\sum_{i=1}^{n} s_i^2 \tag{917}$$

Hence the exponent in the first factor of (914) is

$$\frac{1}{2N}\sum_{i=1}^{n} s_i^2 = \frac{WE}{N} = \frac{E}{N_o} \tag{918}$$

where

$$N_o = \frac{N}{W} \tag{919}$$

is the noise power per unit bandwidth.

In view of the above, we see that establishing a threshold β for $\ell(x)$ is equivalent to establishing a threshold α for the sum (915), such that

$$\beta = e^{-(E/N_o + \alpha)} \tag{920}$$

The probability of a false alarm is the probability that

$$\frac{1}{N}\sum_i s_i x_i \geq \alpha$$

under the hypothesis that the x_i contain noise only. To evaluate this probability, we note that each of the x_i is a Gaussian random variable with zero mean and variance N. Since the s_i are known constants, the variance of $s_i x_i/N$ is s_i^2/N. Furthermore, because of the statistical independence of the x_i, the sum $\frac{1}{N}\sum_{i=1}^{n} s_i x_i$ is a Gaussian random variable with zero mean and variance

$$\frac{1}{N}\sum_{i=1}^{n} s_i^2 = \frac{2E}{N_o} \tag{921}$$

The false alarm probability is therefore

$$P(FA) = \sqrt{\frac{N_o}{4\pi E}} \int_{\alpha}^{\infty} e^{-\frac{N_o}{4E} y^2} dy \tag{922}$$

where

$$y = \frac{1}{N}\sum_{i=1}^{n} x_i^2 \tag{923}$$

$\frac{1}{N}\sum_i s_i x_i$ The probability of detection is the probability that the above sum exceeds α under the hypothesis that the x_i contain a signal in addition to noise. In this case x_i is a Gaussian random variable with mean s_i and variance N, and $s_i x_i/N$ is a Gaussian random variable with a mean of s_i^2/N and vari-

ance s_i^2/N. The sum therefore has both a mean and a variance of $2E/N_0$, and the probability of detection turns out to be

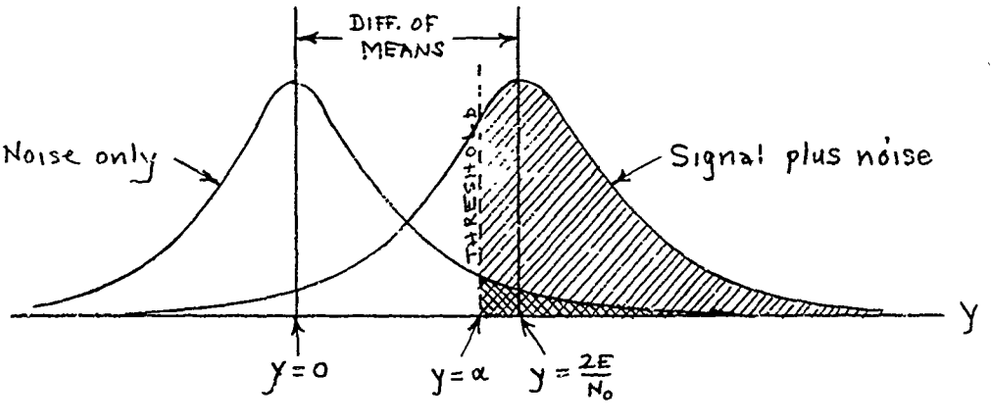
$$P(D) = \sqrt{\frac{N_0}{4\pi E}} \int_{\alpha}^{\infty} e^{-\frac{N_0}{4E}(y - \frac{2E}{N_0})^2} dy$$

$$= \sqrt{\frac{N_0}{4\pi E}} \int_{\alpha - \frac{2E}{N_0}}^{\infty} e^{-\frac{N_0 y'^2}{4E}} dy' \qquad (9-4)$$

where

$$y' = y - 2E/N_0$$

We can get a little better insight into the statistical nature of the decision process if we plot the probability density functions for both of these distributions. Both distributions are Gaussian and have the same variance, $2E/N_0$. The distribution for noise alone has a zero mean, whereas the distribution for signal plus noise has a mean of $2E/N_0$. Hence a plot of the two functions would look pretty much like the sketch shown below. The



dashed line drawn at $y = \alpha$ is the decision threshold. If the measured value of y lies to the left of α, the decision is "no"; if it lies to the right, the decision is "yes". If only noise is present, the probability of a false alarm is represented by the cross-hatched area under the tail of the "noise only" curve. If a signal is present in the noise, the probability of detection is represented by the area under the "signal plus noise" curve to the right of the line $y = \alpha$.

The separation of the maxima of the two curves (the means of the two distributions) is $2E/N_0$. It is quite clear that the ability to distinguish

the signal from the noiseis directly related to this separation. On the other hand, the spreading out of the curves, due to the variance, causes overlapping and makes the decision more difficult. The effective output signal-to-noise ratio of the system, which we shall denote by the symbol d, may be defined as follows:

$$d = \left[\frac{\text{DIFFERENCE OF MEANS}}{\text{STANDARD DEVIATION}}\right]^2 = \frac{(\text{DIFFERENCE OF MEANS})^2}{\text{VARIANCE}} \tag{925}$$

Applying this formula to the Case 1 probability distributions shown above, we obtain

$$d = \frac{(2E/N_o - 0)^2}{2E/N_o} = \frac{2E}{N_o} \tag{926}$$

or, by substitution from (917) and (919),

$$d = 2WT\,\frac{S}{N} \tag{926a}$$

If we interpret the parameter d as the output signal-to-noise ratio, we may say that the optimum receiver in Case 1 has a signal processing gain of 2WT, which represents the number of independent samples in the input wave. We may bypass the concept of output signal-to-noise ratio, however, and treat d simply as a parameter relating the input signal-to-noise ratio to the performance of the system as expressed by the detection probability and false alarm probability. Using the transformation

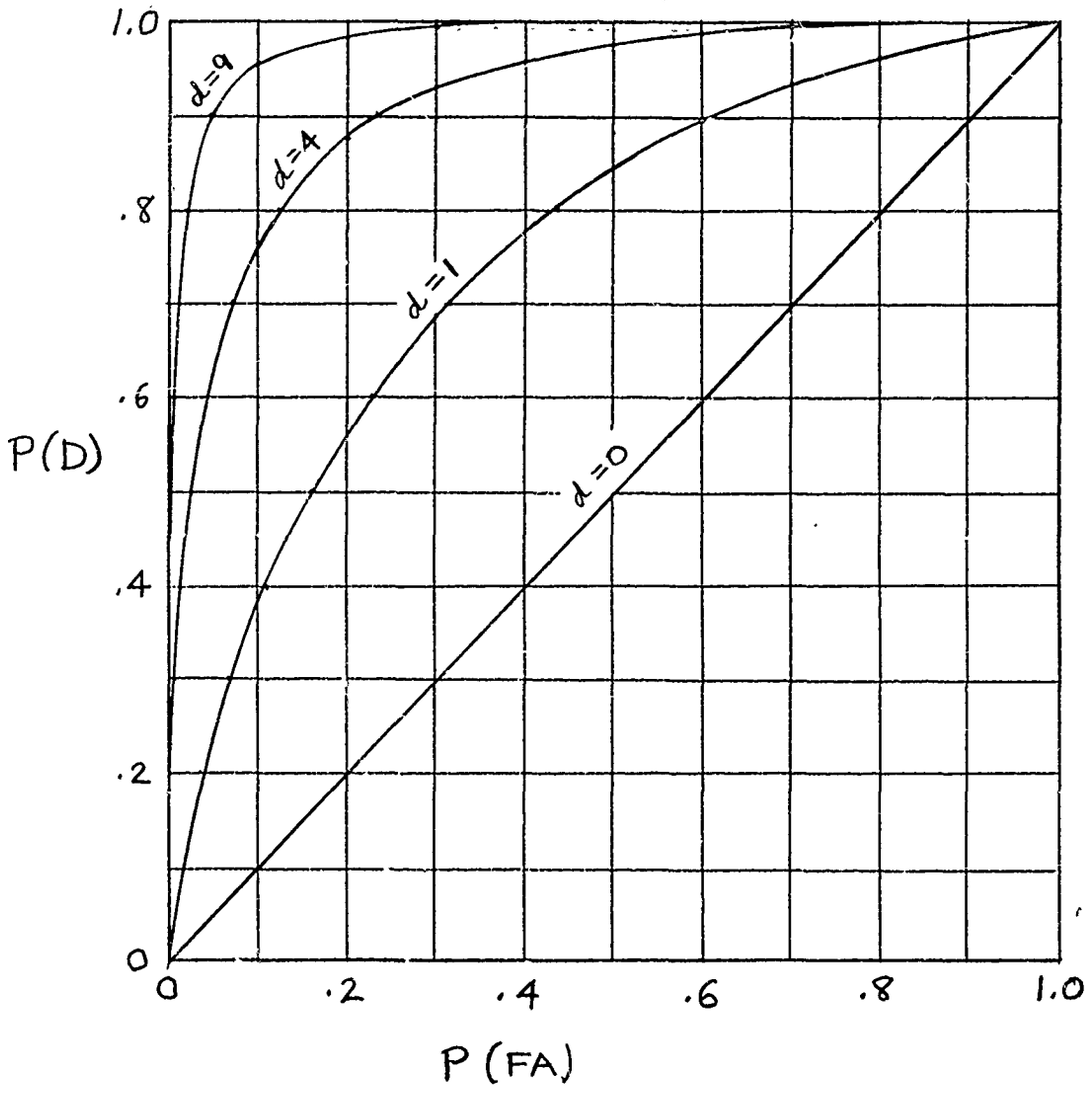
$$\eta = \frac{N_o y}{2E} = \frac{y}{d}$$

we may rewrite (922) and (924) in the following form

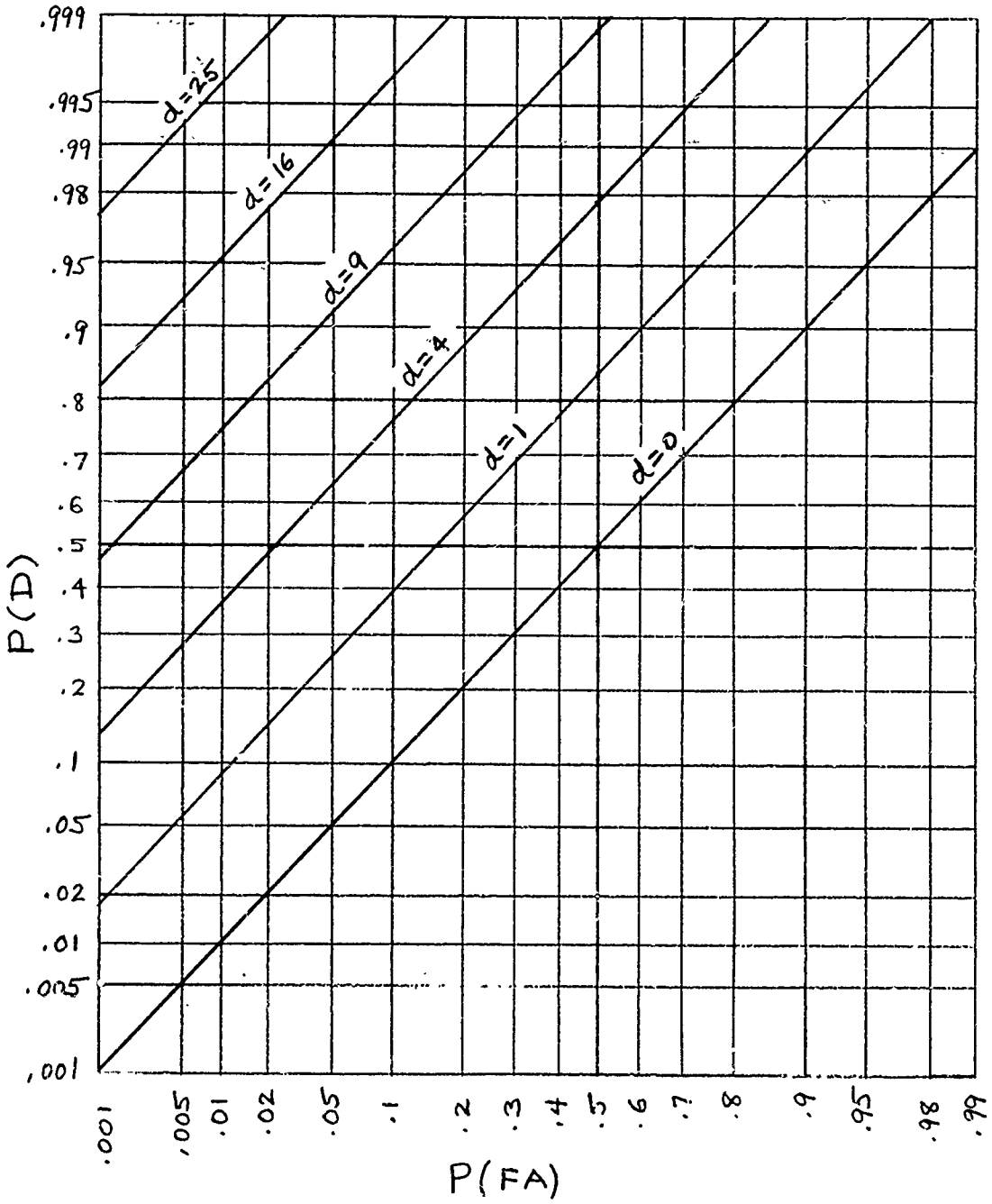
$$P(FA) = \frac{1}{\sqrt{2\pi}} \int_{\alpha/\sqrt{d}}^{\infty} e^{-\eta^2/2}\, d\eta \tag{922a}$$

$$P(D) = \frac{1}{\sqrt{2\pi}} \int_{(\alpha-d)/\sqrt{d}}^{\infty} e^{-\eta^2/2}\, d\eta \tag{924a}$$

Here we have two parametric equations in α, from which the ROC curves for Case 1 may be computed for any desired value of the parameter d. Sample curves are plotted on accompanying graphs in both Cartesian coordinates and Gaussian probability coordinates.



NORMAL ROC CURVES



NORMAL ROC CURVES

Equation (926a) may be turned around and solved for S/N to evaluate the input signal-to-noise ratio required to achieve a given level of performance

$$\frac{S}{N} = \frac{d}{2WT} \tag{927}$$

The decibel equivalent of this signal-to-noise ratio

$$10 \log \frac{S}{N} = 10 \log \frac{d}{2WT} \tag{927a}$$

is the required input signal differential and has a significance similar to that of the recognition differential. To establish the required level of performance we may specify both a minimum detection probability and a maximum false alarm probability. These two probabilities are the coordinates of a point on the ROC diagram. Out of the entire family of ROC curves there is one which goes through this point. If the value of d corresponding to this curve is inserted into (927a), together with the bandwidth and duration of the wave, the required signal differential at the input to the receiver may be computed. (It should be remembered, however, that (927a) refers to the case in which the signal is known exactly, which represents a different physical situation from that which occurs in aural detection.)

Signal processing of the type suggested by the preceding analysis is called <u>coherent</u> processing because the incoming wave is correlated with a known signal. This type of processing is applicable to active sonars in which an exact replica of the transmitted pulse is stored in a memory and is used as the reference in the correlation process. There are two well-known methods of implementing the mathematical procedure indicated in (915). One method involves sampling the incoming wave and the stored reference at the proper rate and performing the multiplication and summation indicated on the left side of (915) in digital fashion. The simplest way of doing this is to clip the waves, throwing away amplitude information, but retaining the polarity. The multiplication and summation can then be done with flip-flop circuits. This method involves a loss of about 2 db in the clipping.

A second method involves the use of a <u>matched filter</u>. A matched filter is one whose impulse response function is matched to the waveform of the transmitted pulse. The impulse response h(t) is the output obtained when a unit impulse (or delta function) is applied to the input. It may be shown that

if an input $x(t)$ is applied to the filter, beginning at $t = 0$, the output may be expressed as the convolution integral

$$y(t) = \int_0^t h(t-\tau)\, x(\tau)\, d\tau \tag{928}$$

Suppose now that the transmitted signal is $s(t)$, $0 \leq t \leq T$. The impulse response function for a matched filter designed for this signal is

$$\begin{aligned} h(t) &= s(T-t), \qquad 0 \leq t \leq T \\ h(t) &= 0, \qquad \text{otherwise} \end{aligned} \tag{929}$$

The output of this matched filter when the input is $x(t)$ is

$$y(t) = \int_{0_T}^{t} s(T-t+\tau)\, x(\tau)\, d\tau, \qquad t \leq T \tag{930a}$$

$$y(t) = \int_{t-T}^{t} s(t-t+\tau)\, x(\tau)\, d\tau, \qquad t \geq T \tag{930b}$$

It is seen from (930a) and (930b) that the output of the filter builds up with time until a maximum value is reached at $t = T$, and then begins to drop. At this time the output is

$$y(T) = \int_0^T s(t)\, x(t)\, dt \tag{930}$$

which is the desired correlation function (915).

Case 2: Signal Consists of White Gaussian Noise.

In Case 1 the maximum amount of information is available concerning the signal. In Case 2 we assume that nothing is known except the statistical properties. The signal consists of white Gaussian noise filtered to the same bandwidth as the background noise. The average signal power is S. The probability density function under the hypothesis of noise only is the same as in Case 1 (911). If the signal is present, each of the samples x_i of the input wave includes both a sample s_i of the signal and a sample n_i of the noise, as indicated by (912), but this time each s_i is a Gaussian random variable with zero mean and variance S, the noise samples being also Gaussian random variables as in Case 1. Now, the sum of two Gaussian random variables is itself a Gaussian random variable whose mean is the sum of the individual means (zero in this case) and whose variance is the sum of the two variances (S+N in this case). Hence the probability density function is

$$p_{SN}(x) = \frac{1}{[2\pi(N+S)]^{n/2}}\, e^{-\frac{1}{2(N+S)} \sum_{i=1}^{n} x_i^2} \tag{931}$$

$$\ell(x) = \frac{(2\pi N)^{n/2}}{[2\pi(N+S)]^{n/2}} \frac{e^{-\frac{1}{2(N+S)}\sum_{i=1}^{n} x_i^2}}{e^{-\frac{1}{2N}\sum_{i=1}^{n} x_i^2}}$$

or

$$\ell(x) = \left(\frac{N}{N+S}\right)^{n/2} e^{\frac{S}{2N(N+S)}\sum_{i=1}^{n} x_i^2} \tag{932}$$

Here we see that the optimum receiver must calculate the sum of the x_i^2. This suggests a square-law detector and averager, which measures the total energy in the input wave. All other factors in the likelihood ratio are known constants. Hence the criterion for decision,

$$\ell(x) \geq \beta$$

can be expressed in an equivalent form in terms of α, such that

$$\frac{1}{N}\sum_{i=1}^{n} x_i^2 \geq \alpha^2 \tag{933}$$

if we define the parameter α as follows

$$\beta = \left(\frac{N}{N+S}\right)^{n/2} e^{\frac{S\alpha^2}{2(N+S)}} \tag{934}$$

The probability of a false alarm, originally defined as the conditional probability of $\ell(x) \geq \beta$ under the hypothesis of noise only, may now be expressed equivalently as the conditional probability of (933) under the same hypothesis. But when only noise is present, each of the x_i is a Gaussian random variable with a mean of zero and a variance N, so that the variance of $x_i/\sqrt{N}$ is unity. The sum (933) therefore has a chi-square distribution with n degrees of freedom. Similarly, when a signal is present, each of the x_i has a Gaussian distribution with zero mean and variance N+S. For this case (933) must be expressed in the form

$$\frac{1}{N+S}\sum_{i=1}^{n} x_i^2 \geq \frac{N}{N+S}\alpha^2 \tag{935}$$

The sum on the left side of (935) now has a chi-square distribution with n degrees of freedom.

It can be shown that when the number of degrees of freedom is large, a chi-square distribution approaches a Gaussian distribution. If we let

$$y^2 = \frac{1}{N}\sum_{i=1}^{n} x_i^2 \tag{936}$$

the Gaussian approximations to the chi-square distributions have the following characteristics:

CONDITION	MEAN OF y	VARIANCE OF y
Noise only	$\sqrt{n-\frac{1}{2}}$	$\frac{1}{\sqrt{2}}$
Signal plus noise	$\sqrt{\frac{N+S}{N}}\,(n-\frac{1}{2})$	$\sqrt{\frac{N+S}{2N}}$

The false alarm and detection probabilities are

$$P(FA) = \frac{1}{\sqrt{2\pi}} \int_{y_1}^{\infty} e^{-y^2/2}\, dy \tag{937}$$

where

$$y_1 = \sqrt{2}\,\alpha - \sqrt{2n-1} \tag{938}$$

and

$$P(D) = \frac{1}{\sqrt{2\pi}} \int_{y_2}^{\infty} e^{-y^2/2}\, dy \tag{939}$$

where

$$y_2 = \sqrt{\frac{2N}{N+S}}\,\alpha - \sqrt{2n-1} \tag{940}$$

In attempting to evaluate the parameter d for this case by substituting the appropriate values into (925), we encounter the problem that the variances of the two distributions (noise and signal-plus-noise) are not the same. However, in evaluating detection probabilities we are interested primarily in the detection of signals near the treshold, and this occurs when the signal-to-noise ratio is small, i.e., when

$$S \ll N$$

Under these conditions the two variances are very nearly equal, and d is approximately

$$d = \frac{\left[\sqrt{\frac{N+S}{N}(n-\frac{1}{2})} - \sqrt{n-\frac{1}{2}}\right]^2}{1/2}$$

$$= (2n-1)\left[\sqrt{1+\frac{S}{N}} - 1\right]^2 \tag{941}$$

If

$$n = 2WT \gg 1$$

and

$$S/N \ll 1$$

equation (941) reduces approximately to

$$d = \frac{1}{2} n \left(\frac{S}{N}\right)^2 = WT \left(\frac{S}{N}\right)^2 \tag{941a}$$

These results show that when the input signal-to-noise ratio is low, we may use the same ROC curves as we did in Case 1, provided we re-define the parameter d. In this case d is proportional to the square of the input signal-to-noise ratio, indicating a greater difficulty in detecting very weak signals. Solution of (941a) for the signal-to-noise ratio in terms of d yields

$$\frac{S}{N} = \sqrt{\frac{d}{WT}} \tag{942}$$

or

$$10 \log S/N = 5 \log (d/WT) \tag{943}$$

It will be noted that since the signal in this case is not known, we are unable to match the incoming wave to a replica of it. Signal processing of this type, in which only the average properties of the signal are known, but not its detailed structure, is called incoherent processing. Comparing coherent with incoherent processing we note that in coherent processing the required input signal differential goes as -10 log WT, whereas in incoherent processing the input signal differential goes as -5 log WT, where W and T are the bandwidth and duration of the signal. It should be noted also that we have considered only one ideal case of each type of processing. Signals with other characteristics have been examined, as well as other types of processers (e.g., half-wave rectifier in place of square-law detector), and the results can usually be approximated by either the coherent or the incoherent formula, except that the parameter d is multiplied by an efficiency factor less than unity.

In the preceding analysis it was assumed that the square of the input voltage is averaged over the duration of the signal, as indicated by the sum of the x_i^2 in (936). By the sampling theorem this sum is equivalent to an integral of $x^2(t)$ over the same interval of time, the total energy in the signal being

$$E = \frac{1}{2W} \sum_{i=1}^{n} x_i^2 = \int_0^T x^2(t)\, dt \tag{944}$$

A common way of mechanizing this mathematical procedure is to feed the input into a square-law detector and to pass the output of the detector through a

filter having an integration time T. In practice, however, it is not always feasible to match the integration time of the filter to the duration of the signal. Failure to achieve a match degrades performance and results in a higher required input signal-to-noise ratio.

Let T' be the filter integration time. Suppose first that $T' < T$. In this case only a portion of the sample T' seconds long contributes to the output of the receiver. The number of independent samples in (936) is only 2WT' instead of 2WT. Hence the required input signal differential (943) is

$$10 \log \frac{S}{N} = 5 \log \frac{d}{WT'}$$

$$= 5 \log \frac{d}{WT} + 5 \log \frac{T}{T'} \qquad (945)$$

Suppose next that the integration time is too long, so that $T' > T$. In this case (936) becomes

$$y^2 = \frac{1}{N} \sum_{i=1}^{2WT'} x_i^2$$

When a signal is present, 2WT of the 2WT' samples contain the signal, whereas the remaining $2W(T'-T)$ samples contain only noise. Hence the mean of the distribution under the hypothesis signal plus noise is (neglecting $\frac{1}{2}$ in comparison with n)

$$\sqrt{2WT \frac{N+S}{N} + 2W(T'-T)}$$

The value of d, which is given by (941) for the ideal case $T' = T$, becomes

$$d = 2 \left[\sqrt{2WT' + 2WT \frac{S}{N}} - \sqrt{2WT'} \right]^2$$

$$= 4WT' \left[\sqrt{1 + \frac{TS}{T'N}} - 1 \right]^2$$

which, when $S \ll N$, is approximately

$$d = WT \left(\frac{T}{T'} \right) \left(\frac{S}{N} \right)^2$$

The required input signal-to-noise ratio is

$$\frac{S}{N} = \sqrt{\frac{d}{WT} \frac{T'}{T}}$$

and the signal differential is

$$10 \log \frac{S}{N} = 5 \log \frac{d}{WT} + 5 \log \frac{T'}{T} \qquad (945a)$$

It is seen that in both (945) and (945a) the second term on the right is positive, indicating an increase in the required signal differential. The effect of the mismatch may therefore be written as a single equation

$$10 \log \frac{S}{N} = 5 \log \frac{d}{WT} + \left| 5 \log \frac{T'}{T} \right| \qquad (945b)$$

e. Relationship to Recognition Differential.

From the preceding discussion it is clear that the deficiency in the concept of recognition differential is the absence of any reference to false alarm probability. This does not mean, of course, that false alarm probability is not present in the aural detection process, merely that it is not present explicitly. Measured values of the recognition differential are obtained from tests of human operators who set their own threshold of detection through a psychological desire not to make mistakes in either direction--false rest or false alarm.

Urick and Stocklin* have made an interesting comparison of the theory for incoherent processing with the observed values of the recognition differential. On the assumption that the process of detection by the ear is similar to that of the optimum receiver for Case 2, they applied equation (945b) with the following numerical values:

W = 50 cps (critical bandwidth of the ear at 800 cps)
T' = 1 second (approximate integration time of the ear)
P(D) = 0.5
P(FA) = 0.05

If the probability of detection is 0.5, the lower limit of integration in (924a) is zero, so that

$$\alpha = d$$

If the probability of a false alarm is 0.05, the lower limit in (922a) is

$$\frac{\alpha}{\sqrt{d}} = \sqrt{d} = 1.645$$

or

$$d = 2.7$$

*R. J. Urick and P. L. Stocklin, "A Simple Prediction Method for the Signal Detectability of Acoustic Systems," NOL White Oak Report NOLTR 61-164, 20 Nov. 1961.

From (945b) the signal differential relative to the noise spectrum level is

$$10 \log \frac{S}{N_o} = 10 \log \frac{S}{N/W}$$

$$= 5 \log \frac{Wd}{T} + \left| 5 \log \frac{T'}{T} \right|$$

$$= 10.7 - 10 \log T, \qquad T \leq 1 \text{ sec.}$$

$$= 10.7, \qquad T \geq 1 \text{ sec.}$$

This result is in fairly good agreement with measured values of the recognition differential. The authors point out that the agreement would be better if a larger value were used for the critical bandwidth of the ear, as has been suggested by recent measurements. It will be noted also that the agreement would be better if a smaller value were used for the false alarm probability.

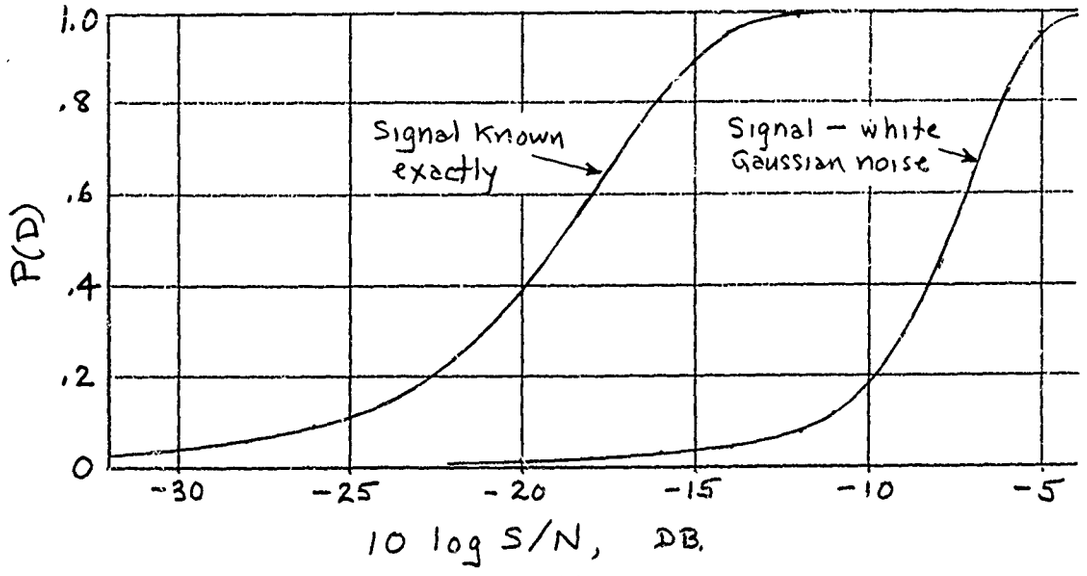
The preceding example illustrates the fact that the signal differential given by (927a) for Case 1 and (945b) for Case 2 may be treated as recognition differentials. Once the maximum allowable false alarm probability has been specified, the value of d for P(D) = 0.5 is determined and this, together with the WT product of the signal, determines the value of the signal differential.

Also, if values are selected for P(FA) and WT, the probability of detection may be computed as a function of the input signal differential. The accompanying graph shows curves for both Case 1 and Case 2 for

$$P(FA) = 0.01$$

and

$$2WT = 400$$



ECHO-RANGING

In this section it will be our purpose to analyze the performance of echo-ranging sonar, chiefly from the point of view of predicting detection ranges. Since the performance of a sonar depends upon whether the background interference is predominantly noise or reverberation, we shall consider the two cases separately. We shall begin by computing the echo level, which is applicable to both cases.

A. Echo Level.

The echo level is computed in four steps, as was mentioned in the introduction to the discussion of decibels and indicated in equation (203). The steps are as follows:

(1) Source Level. The acoustic energy transmitted by the source is expressed in terms of the intensity of the outgoing wave at a standard reference distance of one yard. In the sonar equation this is expressed in decibels as the source level S_o. The relation between the source level and the radiated acoustic power of an omnidirectional transducer is expressed by (238) and (239). For a directive transducer we must include the directivity index defined by (752) and (753). The source level along the beam axis is

$$S_o = 10 \log P + D + 71.7 \text{ db}//\mu b^2 \qquad (1001)$$

where

P = radiated acoustic power in watts

and

D = transmitting directivity index

The acoustic power may be expressed in terms of the input electrical power P_{in} and the electroacoustic efficiency η_{EA} of the transducer

$$P = \eta_{EA} P_{in} \tag{1002}$$

or

$$10 \log P = 10 \log P_{in} - N_{eff} \tag{1002a}$$

where N_{eff} is the efficiency loss

$$N_{eff} = -10 \log \eta_{EA} \tag{1003}$$

(2) The propagation loss N_w from the one yard point to the target.

(3) The target strength. When the outgoing wave strikes the target, it is scattered in all directions. In re-radiating the sound, the target acts as a source. The returning wave is treated in the same manner as the outgoing wave. The target strength is a measure, in decibels, of the ratio of the reference intensity of the returning echo at one yard from the target to the incident intensity of the outgoing wave as it strikes the target. Let

I = intensity of outgoing wave at target

I_1' = index intensity of returning echo

Then the target strength is

$$T = 10 \log \frac{I_1'}{I} \tag{1004}$$

(4) The propagation loss of the returning echo.

The echo level L_E is the intensity level of the returning wave as it reaches the transducer. The complete process described above may be expressed by the simple formula

$$L_E = S_o + T - 2N_w \tag{1005}$$

1. Target Strength.

The source level and propagation loss have been discussed above. Let us now consider the target strength.

a. Target Strength of a Sphere.

To introduce the subject, let us first consider the echo formed by a hypothetical target consisting of a perfectly scattering sphere of radius a. In all cases of practical interest the target is far enough away from

the transducer that the incident radiation may be considered as a plane wave. The projected area of the sphere which intercepts the incident wave is πa^2. If the incident intensity is I, the intercepted power is $\pi a^2 I$. Let us assume that the sphere scatters all this power uniformly in all directions, acting like an omnidirectional source. The index intensity I_1' is the ratio of the total scattered power to the area of a sphere of radius 1 yard, as indicated by equation (237). Thus,

$$I_1' = \frac{\pi a^2 I}{4\pi r_1^2} \tag{1006}$$

Substituting (1006) into (1004), we obtain for the target strength of the sphere,

$$T = 10 \log \frac{\pi a^2}{4\pi r_1^2} \tag{1007}$$

or

$$T = 10 \log \left(\frac{a}{2r_1}\right)^2 \tag{1007a}$$

The target strength of such an ideal sphere is tabulated below as a function of the radius

Radius a (yds)	Target Strength T (db)
1	-6
2	0
4	6
10	14
20	20
40	26
100	34

These computed values agree remarkably well with observed values for submarines, as we shall see.

b. Other Targets.

If the sphere is not a perfect scatterer we may take this imperfection into account by multiplying the effective intercepting area by a factor α, so that (1007) becomes

$$T = 10 \log \frac{\pi a^2 \alpha}{4\pi r_1^2} \tag{1007b}$$

This equation may be generalized to targets of other shapes by replacing the numerator $\pi a^2 \alpha$ with a symbol σ representing the effective scattering area of the target.

$$T = 10 \log \frac{\sigma}{4\pi r_1^2} \qquad (1008)$$

The effective scattering area σ is more commonly used in radar work; the target strength T is more commonly used in sonar work.

c. Experimental Measurements.

A series of tests were run several years ago on spheres of about one yard diameter, using sound waves of various frequencies. The results of the measurements, although they showed considerable variability, nevertheless tended to confirm the validity of the theoretical formula. The observed target strengths ranged from the theoretical value to about 6 db higher.

Target strength measurements of submarines in general show a strong dependence on aspect. As is to be expected, the smallest values are found for bow and stern aspect and the largest values for beam aspect. The following table lists generally accepted values for typical submarines.

Submarine Aspect	Approximate Target Strength db
Bow, Stern	10
Beam	25
Random	15

The above values should be considered merely as nominal values. Whenever measurements are made as a function of aspect, the experimental curves usually show strong spikes and depressions, indicating wide fluctuations in the observed values.

It is of interest to compare the observed results with a rough theoretical calculation based on formula (1008). A typical beam aspect submarine looks somewhat like an ellipse about 100 yards long and

10 yards wide. The area of the ellipse is 250π square yards, and assuming perfect reflection the target strength would be

$$T = \frac{250\pi}{4\pi} = 18 \text{ db}$$

For the bow and stern aspect the projected area is roughly 25π, so that the target strength would be

$$T = \frac{25\pi}{4\pi} = 8 \text{ db}$$

Considering the roughness of the computations, the agreement with observed values is quite good.

Target strength appears to be practically independent of frequency so long as the dimensions of the target are large compared with the wavelength of sound.

In order to obtain experimental values of target strength, it is necessary to measure the other three quantities (L_E, S_o, and N_w) in equation (1005) and to calculate T from this equation. The practical difficulties encountered in obtaining reliable results stem from the problems involved in measuring these three quantities. One of the serious difficulties is the large fluctuations which continually occur in the echo level.

2. Effect of Deviation Loss

The echo level as defined by equation (1005) applies only to targets situated on the maximum response axis of the transducer. If the target is in some other direction (θ, ϕ) away from the axis, the detection capability of the sonar is reduced. Two effects must be considered: (1) The intensity level of the transmitted sound is reduced because of the transmitting deviation loss. The echo level is therefore lower by the same amount. (2) The returning echo is received in a direction of reduced receiving response. So far as the response of the sonar is concerned the echo level must therefore be reduced by an amount equal to the receiving deviation loss. The effective echo level, as sensed by the transducer, is

$$L_E(\theta, \phi) = S_o - N(\theta, \phi) - N'(\theta, \phi) + T - 2\, N_w \qquad (1009)$$

where

$N(\theta, \phi)$ = transmitting deviation loss

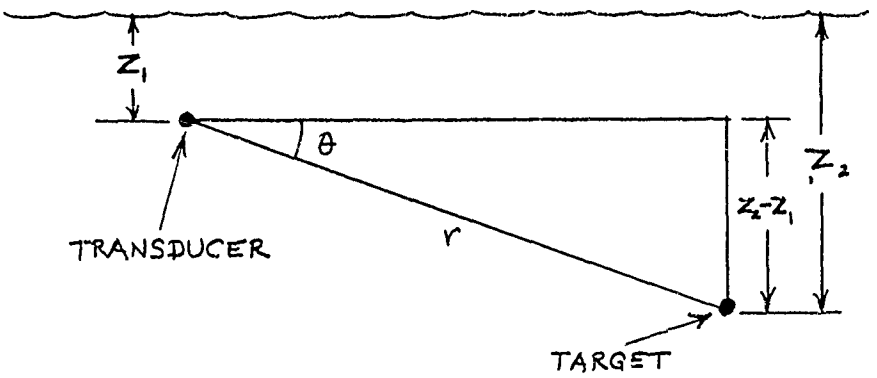
$N'(\theta, \phi)$ = receiving deviation loss

as given by (706) and (706a).

Deviation loss may occur in either of two ways. First, since most transducers possess directivity in the vertical plane it is possible for a target to "get under the beam" or "over the beam." This may occur whenever the transducer is shallow and the target is deep, or vice versa. Suppose the depths of the transducer and target are Z_1 and Z_2, respectively, and the slant range is r. Neglecting the ray curvature due to refraction, the angle off the beam in the vertical plane is given by

$$\sin \theta = \frac{Z_2 - Z_1}{r} \tag{1010}$$

Since appreciable deviation angles occur only at very short ranges, the assumption of straight-line ray propagation does not introduce appreciable error.



Deviation loss may also occur in azimuth when the transducer generates a beam with horizontal directivity. This should not normally be a serious problem, however, because in a properly designed sonar there should be sufficient overlap between successive horizontal beams or between successive positions of a horizontally stepped beam (used as a searchlight), so that all azimuths are covered with a deviation loss of not more than 3 db.

3. Note on Positive Values of Target Strength

The fact that the target strengths of submarines are positive indicates that the index intensity of the scattered wave is higher than the intensity of the incident wave. This sometimes causes confusion because on first thought it appears to violate the principle of conservation of energy. However, a more careful consideration of the situation shows that this is not the case. The numerical value of the index intensity depends upon the standard reference distance at which it is defined. As explained previously, the index intensity is really only a mathematical fiction. At large distances from a source of finite dimensions the waves appear to have come from a point. For the purpose of describing the energy transmission, it is standard practice to refer propagation loss measurements to a distance of one yard from a hypothetical equivalent source. The index intensity does not therefore represent the actual intensity of a real wave. It will be noted that if a reference distance of 100 yards had been selected instead of 1 yard, the target strength and the propagation loss would each have been 40 db lower.

4. Bi-static Echo-Ranging

In bi-static echo-ranging operations the projector and hydrophone array are separated from each other. There are two differences between bi-static and mono-static echo-ranging. First, the propagation loss is not the same for the two directions of travel. Second, the target strength is different. The bi-static target strength involves two angles relative to any fixed direction in the target. The bi-static target strength is a maximum when the incident and reflected angles, relative to the beam of the target, are equal. Measurement of bi-static target strength is considerably more complicated than conventional target strength measurements, since two angles are involved.

B. Noise-Limited Ranges

1. Noise

Interfering noise arises from a number of sources, including the ambient noise of the sea, noise caused by the vehicle which supports the

sonar, and noise caused by the motion of the sonar through the water. Noise originating within the sonar, its installation, or its supporting vehicle, is called self-noise. The overall noise which interferes with the echo is the resultant of all the individual contributions.

Since the symbols N and N_o have been used in a previous section to denote the noise power and power per unit band, respectively, we shall use the symbols L_N and L_{N_o} to denote the corresponding levels, the band level and the spectrum level. If the band width of the receiving system is W and the spectrum level is evaluated at the geometric mean frequency in the band, then from (223)

$$L_N = L_{N_o} + 10 \log W \qquad (1011)$$

a. Ambient Sea Noise.

Reference has already been made to the Knudsen curves which relate the average ambient noise to the acoustic frequency and the sea state. These curves can be approximated reasonably well by equation (811) which is reproduced below

$$L_{amb} = -54 + 30 \log(\eta_s + 1) - 17 \log f \text{ db//}\mu b^2/\text{cps} \qquad (811)$$

where

L_{amb} = spectrum level of ambient noise in db//μb^2/cps

η_s = sea state

f = frequency in kc

The Knudsen curves are still the most widely used reference for ambient sea noise. They are applicable only to frequencies above about 500 cps. Subsequent measurements have shown a significant departure from the Knudsen curves at low frequencies. Extensive measurements have been made by G. M. Wenz* of the Navy Electronics Laboratory, and others. Wenz's curves agree with the Knudsen curves reasonably well at the higher

* G. M. Wenz, Journal of the Acoust. Soc. of Am., Vol. 34, pp 1936-1956 (1962)

frequencies, but below about 1000 cps they begin to taper off from the constant slope of -5 db per octave. The spectrum level reaches a maximum at 400 to 500 cps and then drops at lower frequencies. At very low frequencies, below about 20 cps, the noise begins to rise again, showing a strong frequency dependence of about -8 to -10 db/octave.

Another factor which contributes significantly to the ambient noise at low frequencies is shipping. Because of the low attenuation, sound generated by ships travels great distances at very low frequencies. In areas of high shipping density this man-made noise becomes widely diffused and is sometimes the dominant interference. On the other hand, at frequencies in the kilocycle range ship noise is a local phenomenon which is troublesome only in the vicinity of individual vessels

Effect of Directivity.

We have seen previously that where the noise field is isotropic the effective noise as sensed by a directive transducer is less than ambient by an amount equal to the directivity index. If L_N' denotes the effective noise spectrum level as sensed by the transducer, then (783) may be rewritten as

$$L'_{N_o} = L_{N_o} - D \qquad (783a)$$

If the noise field has directional properties, the effective noise spectrum level must be computed from (779) and (785).

b. Vehicle Self-Noise

Sonars which are mounted to the hulls of large ships are generally bothered by a rather high background noise. This noise increases with ship speed and with sea state, and under most operating conditions greatly exceeds the ambient sea noise. It is common practice in destroyer sonar installations to measure the overall noise as sensed by the installed equipment. Such a measurement automatically includes the ambient sea noise, the machinery and flow noise of the ship, and the directivity of the transducer.

The noise generated by a helicopter is an important factor in the performance of helicopter dipped sonar equipments. Although a considerable amount of work has been done in studying the effect of vehicle noise on dipped sonar transducers, it is still not definitely known whether the noise sensed by the transducer is caused directly by the engines and rotor blades or by a disturbance of the water surface set up by the down-wash. It has been established that the noise in the water directly underneath the helicopter decreases with depth below the surface. It approximately follows the inverse square law, such as would be expected if the cause of the noise were a point source on the surface of the water.

It has also been found that the helicopter noise has directional characteristics. Directly under the vehicle the noise appears to come from a predominantly vertical direction, as might well be expected. Because of this directional dependence the effective noise as sensed by the transducer is not given by (783a) but depends, instead, upon the response of the transducer in the vertical direction. This may be written mathematically as follows

$$L'_{NH} = L_{NH} - N_v \qquad (1012)$$

where

L'_{NH} = helicopter noise spectrum level as sensed by the directional transducer

L_{NH} = helicopter noise spectrum level as sensed by an omnidirectional hydrophone

N_v = deviation loss of the directional transducer in the vertical direction

If (1012) is rewritten in the following form

$$L'_{NH} = L_{NH} - (N_v - D) - D$$

and compared with (783a), it is seen that the quantity

$$L_{NH} - (N_v - D')$$

enters in the same way as the isotropic (i.e., omnidirectional) spectrum level L_{N_0}. This quantity may be termed the equivalent omnidirectional noise of the helicopter. Any noise which is directed at right angles to the sonar

beam may be treated as an omnidirectional noise if it is first corrected by subtracting the term $N_V - D$.

c. Noise of Towed Sonar

Transducers intended to be towed through the water are mounted in hydrodynamically streamlined housings, often called "fish." At low speeds (below 15 knots or so) the noise is usually below the ambient sea noise. At higher speeds the noise increases rapidly. Although considerable effort has been directed to the study of flow noise, the development of high-speed towed sonars still appears to present difficult problems.

d. Resultant Noise

When noise from more than one source is present at the input to a transducer, the overall effect is determined by their resultant. Consider the case of two sources. For rough work it is customary simply to consider the one which produces the higher level and to ignore the other. The maximum error involved in this procedure occurs when the two levels are equal and amounts to 3 db. If more accuracy is desired, the two may be combined on the basis of random phases by adding intensities. (See equation (185)). However, before combining them, one must be sure they are all expressed on a comparable basis. For example, they must all be expressed either as spectrum levels or as intensity levels having the same bandwidth. Also, they must all be expressed as equivalent omnidirectional noise, in which case the receiving directivity index D is subtracted from the resultant, or else they must be expressed as effective values as sensed by the transducer, such as given by (783a) or (1012).

To combine the noise spectrum levels we add the intensities per unit band. Let J_1 and J_2 be the intensities per unit band corresponding to spectrum levels L_{N_1} and L_{N_2} for the noise from two different sources, so that

$$L_{N_1} = 10 \log J_1 \qquad (1013a)$$

and

$$L_{N_2} = 10 \log J_2 \qquad (1013b)$$

The resultant intensity per unit band is

$$J = J_1 + J_2 \tag{1014}$$

and the resultant spectrum level is

$$L_{N_o} = 10 \log J \tag{1015}$$

We seek a relation between L_{N_o}, and L_{N_1} and L_{N_2}. Solving (1013a) and (1013b) for J_1 and J_2 gives

$$J_1 = 10^{0.1\, L_{N_1}} \quad \text{and} \quad J_2 = 10^{0.1\, L_{N_2}}$$

Therefore

$$L_{N_o} = 10 \log \left(10^{0.1\, L_{N_1}} + 10^{0.1\, L_{N_2}}\right) \tag{1016}$$

Equation (1016) may be expressed in a form more suitable for computation. Let us define L_{N_1} as the larger of the two:

$$L_{N_1} \geq L_{N_2} \tag{1017}$$

Then $$L_{N_2} = L_{N_1} - (L_{N_1} - L_{N_2})$$

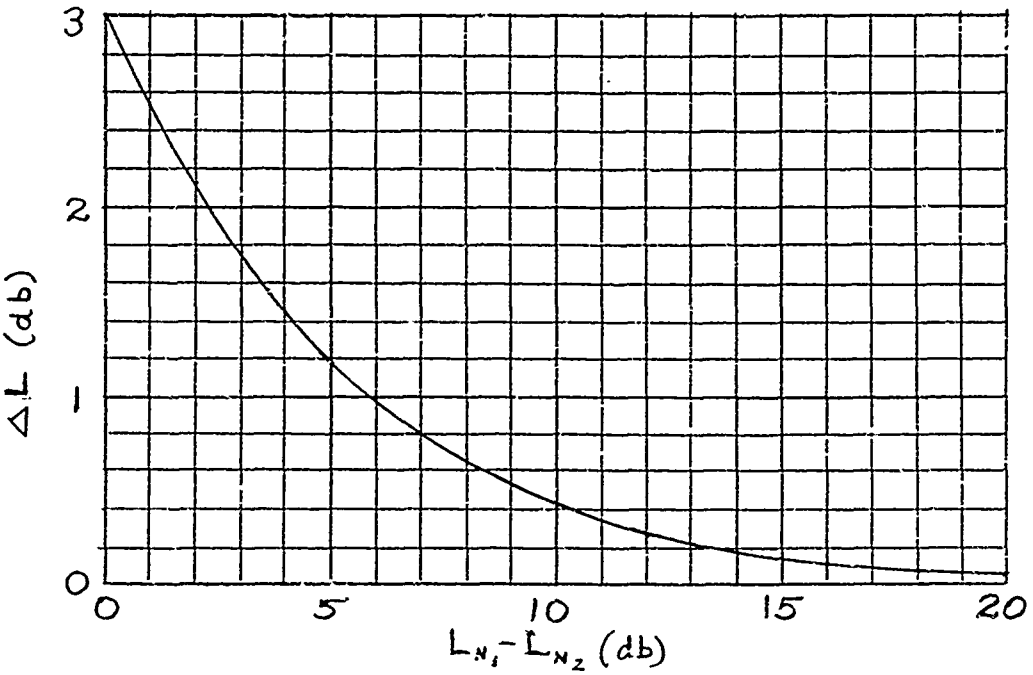
and $$L_{N_o} = 10 \log \left[10^{0.1\, L_{N_1}} + 10^{0.1\, L_{N_1}}\, 10^{-0.1\, (L_{N_1} - L_{N_2})}\right]$$

$$L_{N_o} = L_{N_1} + 10 \log \left[1 + 10^{-0.1\, (L_{N_1} - L_{N_2})}\right]$$

Let $$\Delta L = 10 \log \left[1 + 10^{-0.1\, (L_{N_1} - L_{N_2})}\right] \tag{1018}$$

Then $$L_{N_o} = L_{N_1} + \Delta L \tag{1019}$$

The accompanying graph shows ΔL as a function of $L_{N_1} - L_{N_2}$. To compute the resultant of two noises L_{N_1} and L_{N_2}, of which L_{N_1} is the larger (algebraically), form the difference $L_{N_1} - L_{N_2}$, and find ΔL from the graph. Add this to L_{N_1}. Note that to the nearest 0.1 db L_{N_2} may be neglected if it is more than 20 db smaller than L_{N_1}.



2. The Sonar Equation

The sonar equation, as conventionally written, is a relation among the various sonar parameters, applicable to the situation in which the probability of detection is 50 percent. It can be seen from equation (1005) that for any given source level and target strength the echo level varies with the two-way propagation loss. The propagation loss, in turn, varies with the distance from the source. At short ranges very strong echoes are received, but as the range increases, the echo level decreases until it becomes lost in the noise. The nominal detection range of a sonar is the range at which the detection probability is 0.5 and the echo level at this point is called the minimum detectable level (MDL). For an omnidirectional transducer the minimum detectable level is L_{50} of equation (903). If the noise field is isotropic (an assumption which we shall make throughout this discussion), the effective noise level sensed by a directive transducer is reduced by an amount equal to the directivity index. Hence the general expression is

$$L_{50} = L_N - D + M \qquad (1020)$$

The levels expressed in (1020) are band levels in the input band of the receiving system.

At the 50 percent point the echo level L_E of (1005) is equal to the MDL, L_{50} of (1020). The sonar equation is obtained by equating these two levels.

$$S_o + T - 2 N_w = L_N - D + M \tag{1021}$$

3. Figure of Merit.

The figure of merit (FOM) of a sonar is defined as the difference between the source level and the minimum detectable signal,

$$FOM = S_o - L_{50} \tag{1022}$$

If (1020) is substituted, the figure of merit can be expressed in the form

$$FOM = S_o + D - M - L_N \tag{1022a}$$

Here it is seen that the source level, directivity index, and recognition differential are properties of the sonar system (including the operator) and hence are a measure of the quality of the system. Furthermore, in some sonars, especially surface ship sonars, the dominant background noise is generally self-noise and may therefore be logically considered a property of the sonar system. In quieter sonars such as sonobuoys, where the background is chiefly ambient noise, the figure of merit depends upon environmental conditions as well as upon the sonar itself.

The relationship between the figure of merit and the maximum allowable propagation loss is obtained by substituting (1021) into (1022a)

$$2 N_w = FOM + T \tag{1023}$$

The allowable two-way loss is thus equal to the figure of merit plus the target strength.

Another measure of sonar performance which is used for surface ship and submarine sonars is the performance figure, which is defined as

$$PF = \text{Source Level} - \text{Measured Noise Level} \tag{1024}$$

Some ship sonars are provided with test instruments which measure the effective noise level as sensed by the transducer, and it is this noise level which is referred to above. The performance figure differs from the figure of merit in that it does not include the recognition differential. It is useful as an objective measure of the condition of the sonar.

4. Echo Excess; Detection Probability.

We have seen earlier, in the discussion of recognition differential for aural detection, that as the input signal differential is varied, there exists a transition region in which the probability of recognizing an echo varies from 0 to 1. A plot of this probability versus the signal differential reveals an approximately Gaussian distribution with a standard deviation of about 2.5 db. It is interesting to note also in connection with the statistical detection theory discussed previously that a plot of detection probability versus signal differential for a fixed false alarm probability also exhibits a distribution which is approximately Gaussian, except at low levels of probability. It can be shown that the Gaussian approximation to the distribution of the signal differential has a mean M and standard deviation σ as follows:

Case 1, Signal known exactly:

$$M = 10 \log \frac{d'^2}{2WT} \tag{1025a}$$

$$\sigma = \frac{20 \log_e 10}{d'} = \frac{8.7}{d'} \tag{1026a}$$

Case 2, Signal consisting of Gaussian noise:

$$M = 5 \log \frac{d'^2}{WT} \tag{1025b}$$

$$\sigma = \frac{10 \log_e 10}{d'} = \frac{4.3}{d'} \tag{1026b}$$

where d' is a parameter related to the false alarm probability as follows:

$$P(FA) = \frac{1}{\sqrt{2\pi}} \int_{d'}^{\infty} e^{-\frac{\eta^2}{2}} d\eta \tag{1027}$$

It will be noted that the σ's are independent of the WT product of the signal and that the σ for Case 1 is twice that for Case 2. The numerical values for a false alarm probability of 0.01 are about 3.7 and 1.9, respectively. The means M of the distributions are the input signal differentials required for 0.5 probability of detection and thus correspond to recognition differentials.

Up to this point we have considered only one of many factors which contribute to variability in the detection of echoes. In order to predict the performance of sonars under operational conditions we must include other sources of variability, the major ones being:

(1) fluctuations in acoustic propagation,

(2) target aspect,

(3) operator variability and the effect of random noise, discussed in the preceding paragraphs, and

(4) variability in the condition of the sonar equipment.

The values of the various parameters, such as source level, propagation loss, noise, etc., which are used to compute the echo level and the background interference level, are nominal average values. The actual values which exist during any particular echo-ranging operation are subject to unpredictable fluctuations and must therefore be considered on the basis of probability. When the nominal values of the various parameters are such that the computed echo level L_E is equal to the MDL L_{50}, the probability of detection is 0.5 (by definition). When L_E exceeds L_{50}, the probability is greater than 0.5; when L_E is less than L_{50}, the probability is less than 0.5. It is therefore useful to introduce the concept of <u>echo excess,</u> ΔL_E, which is defined as the difference between these two levels

$$\Delta L_E = L_E - L_{50} \qquad (1028)$$

Because of the random nature of the sonar parameters, the echo excess is a random variable. Unfortunately, the probability distributions of many of the parameters are not accurately known and hence one can only guess

at the distribution of the signal excess. It is generally assumed that the distribution is Gaussian with a mean of zero. It is also logical to assume that the contributions from the various inputs are statistically independent, so that the variance (that is, the square of the standard deviation) of the signal excess is equal to the sum of the individual variances. The probability of detection of a sonar ping may therefore be expressed as follows

$$P(D) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\Delta L_s}{\sigma}} e^{-\frac{\eta^2}{2}} d\eta \qquad (1029)$$

where σ is the standard deviation of the echo excess. If σ_i is the standard deviation of the i^{th} input to the signal excess, then

$$\sigma = \sqrt{\sum_i \sigma_i^2} \qquad (1030)$$

Let us now consider the contributions of the four items listed above.

(1) Although the variations in propagation loss do not obey the Gaussian distribution law extremely well, the best fit for the available data leads to a standard deviation of about 4.5 db for one-way propagation. The standard deviation for two-way propagation should be larger by a factor of $\sqrt{2}$, that is, about 6.5 db.

(2) Since the sonar operator normally does not know the aspect of the target, the target strength must be considered as a random variable. A reasonable estimate of the standard deviation is about 7 db.

(3) and (4). The operator variability and the equipment variability may be lumped together under the single category of figure of merit. We have already seen that the operator variability amounts to about 2.5 db. Experience has shown that among the various individual installations of any particular type of electronic equipment in service use there is a wide variability in equipment operating condition. Some installations will be in excellent condition, others in very poor condition, with all graduations in between. The

factors involved include such things as source level, directivity index, and self-noise. Because of meager data any estimate of the standard deviation of the figure of merit must be very rough. The figure of 7 db, which is sometimes applied to surface vessel sonar, appears to be reasonable.

The overall standard deviation of the echo excess is

$$\sigma = \sqrt{6.5^2 + 7^2 + 7^2} \approx 12 \text{ db} \qquad (1031)$$

The sonar equation (1021) may now be modified to apply to other probability levels than 0.5. According to (1029), to each value of probability P(D) there corresponds an echo excess ΔL_E, representing the amount by which the required echo level must exceed the MDL (equation (1028)). The modified sonar equation is therefore

$$S_O + T - 2N_W = L_N - D + M + \Delta L_E \qquad (1032)$$

5. Noise-Limited Ranges.

Once the values of the parameters S_O, T, L_N, D, M, and P(D) (hence also ΔL_E) have been specified, the propagation loss N_W may be computed from (1032). If the propagation characteristics of the water are known, this value of N_W may be converted to distance, thus determining the detection range corresponding to the selected value of probability. It will be noted that the above results are predicated on the assumption that the target is located on the axis of the beam. When this is not the case, the transmitting and receiving deviation loss $N(\theta, \phi)$ and $N'(\theta, \phi)$ (706) and (706a) must be subtracted from the source level.

C. The Doppler Shift

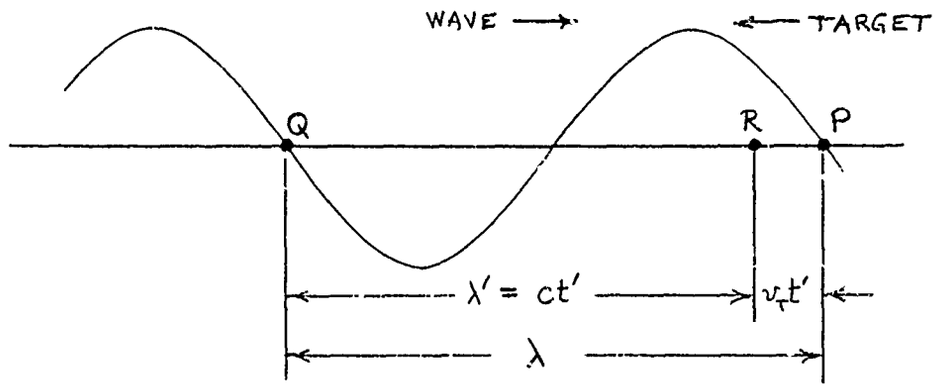
When a source of sound is moving toward the receiver, the frequency of the sound as observed by the receiver is higher than it would be if the source were stationary. Conversely, if the source is moving away from the receiver, the frequency is lower. The same type of effect is observed if the receiver is moving toward or away from the source. Furthermore, if the receiver acts as a target, reflecting sound back to the source, the change in frequency

observed at a hydrophone located at the source is approximately double that observed in one-way propagation.

The phenomenon which causes these frequency shifts is called the Doppler effect. The increase in frequency which occurs when the source and target are approaching each other is popularly called up-Doppler; the decrease which occurs when they are separating is called down-Doppler. In echo-ranging operations the frequency shift due to the motion of the transducer is called self-Doppler or own-Doppler.

1. Target Doppler.

Assume that the sonar is stationary; only the target is moving The only component of the target velocity which contributes to the Doppler shift is the radial component, i.e., the component along the line joining the source and target. Let v_T represent this component. Consider first the frequency shift as sensed by an observer moving with the target. The following sketch shows schematically the outgoing wave traveling toward the right. Assume that



at a given instant of time the target, traveling toward the left, is located at the point P. Consider now the point Q, one wavelength behind P. While this is moving with speed c, the target is moving with speed v_T . The target will therefore meet Q at some intermediate point R at some time t' later. During

this interval of time the wave has traveled a distance ct' and the target has traveled a distance $v_T t'$, and the sum of these distances is one wavelength. Hence

$$\lambda = (c + v_T)t'$$

But from the point of view of an observer on the target, a complete cycle has elapsed during this interval, so that the apparent wavelength is

$$\lambda' = ct'$$

The ratio of the two wavelengths is

$$\frac{\lambda}{\lambda'} = \frac{c + v_T}{c}$$

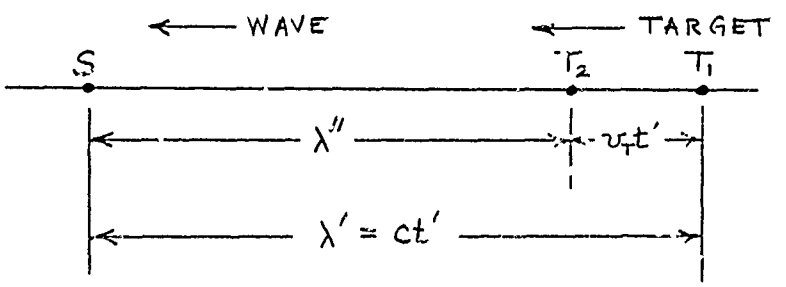
The wavelengths are related to their respective frequencies by the familiar formula (11)

$$\lambda f = c = \lambda' f'$$

The ratio of the apparent to the true frequency is therefore

$$\frac{f'}{f} = \frac{c + v_T}{c} \qquad (1033)$$

Consider now the return trip. In the following sketch both the target and the wave are both moving toward the left. Let the target be at T_1 at some given instant of time. During the time t' of one cycle as measured by an observer on the target, the wave will have traveled to S, a



distance of

$$\lambda' = ct'$$

and the target will have traveled to T_2, a distance of $v_T t'$. As seen by a stationary observer, the distance T_2S is one wavelength λ''. Hence

$$\lambda'' = (c - v_T)t'$$

and

$$\frac{\lambda'}{\lambda''} = \frac{f''}{f'} = \frac{c}{c - v_T} \tag{1034}$$

where f'' is the frequency as measured by a stationary observer at the sonar. Combining (1033) and (1034) we obtain

$$\frac{f''}{f} = \frac{c + v_T}{c - v_T} = \frac{1 + \frac{v_T}{c}}{1 - \frac{v_T}{c}} \tag{1035}$$

The Doppler shift Δf is the difference between the frequency f'' of the echo and the frequency f of the transmitted wave.

$$\Delta f = \frac{\frac{2fv_T}{c}}{1 - \frac{v_T}{c}} \tag{1036}$$

The speed of sound in water is of the order of 2800 to 3000 knots, whereas the speeds of sonar targets are well below 50 knots. Therefore in all practical situations

$$\frac{v_T}{c} << 1$$

and (1036) may be simplified approximately to

$$\Delta f = \frac{2fv_T}{c} \tag{1037}$$

2. Self-Doppler

Let v_S represent the component of the sonar speed in the direction of the target. The analysis in this case is identical to that of the target Doppler, except that the shifts associated with the outgoing pulse and the returning echo are interchanged. The resulting frequency shift is approximately

$$\Delta f = \frac{2fv_S}{c} \tag{1038}$$

3. Combined Doppler

If both the sonar and the target are moving relative to each other, the resultant Doppler shift is the sum of (1037) and (1038)

$$\Delta f = \frac{2f\,(v_S+v_T)}{c} \qquad (1039)$$

Although the formulas for target Doppler and self-Doppler are similar, there is a significant difference between the two effects in regard to reverberation. Reverberation refers to the sound energy which is scattered back to the sonar from objects in the ocean other than the target. Scattering objects are normally stationary, or at worst have very small Doppler speeds (as, for example, waves at the ocean surface). If the target has an appreciable Doppler speed, the frequency content of the target echo will be significantly different from that of the reverberation, and this has an important bearing on the detectability of the echo. If, on the other hand, the target Doppler speed is zero and only self-Doppler is present, the frequency content of the echo and reverberation are the same. It should be pointed out also that if the sonar has a large horizontal beamwidth the reverberation received when the sonar is moving will be spread out in frequency, due to different self-Doppler speeds in different parts of the beam.

If the frequency f is expressed in kilocycles, the Doppler shift Δf in cps, and the Doppler speed in knots, (1039) is approximately

$$\Delta f = 0.7\, f\,(v_S + v_T) \qquad (1039a)$$

4. Minimum Required Bandwidth

The Doppler shift sets a lower limit on the input bandwidth of the system, since the bandwidth must be large enough to allow all possible echoes to come through. Including both up-Doppler and down-Doppler, the minimum bandwidth is

$$W = 2\Delta f \qquad (1040)$$

where Δf is the Doppler shift corresponding to the maximum expected Doppler speeds.

D. Reverberation-Limited Ranges

Reverberation differs from noise in that it is caused by the sonar itself. Reverberation is the unwanted sound which is scattered back to the sonar by objects in the sea other than the target. There are two basic types of reverberation:

(1) Volume reverberation, which is caused by scatterers distributed throughout the volume of the water.

(2) Surface reverberation, which is the sound scattered back from the boundaries of the ocean, top and bottom. However, since the scattering characteristics of the upper surface are somewhat different from those of the bottom, the term surface reverberation is usually restricted to scattering from the water-air interface, and the scattering from the ocean bottom is referred to as bottom reverberation.

Reverberation differs from noise in several ways. In the first place its spectral characteristics are essentially the same as those of the transmitted pulse and are therefore different from those of noise. Secondly, the intensity of the reverberation varies with the range at which the scatterers are located, so that the reverberation level from any given pulse varies with time. Thirdly, the intensity of the reverberation is proportional to the intensity of the transmitted pulse, so that whenever a sonar system is reverberation-limited, no improvement in the detectability of a target can be achieved by increasing the source level.

1. Volume Reverberation

To simplify the analysis we shall begin by assuming an ideal ocean which is infinite in extent and in which sound waves spread out in accordance with the inverse square law with no attenuation. These over-simplifications will be corrected later. We shall begin with the sound as it is emitted by the transducer and shall follow it out to the scatterers and then back to the transducer.

Assume that the duration of the pulse is τ seconds. We are interested in the total scattered intensity which is returned to the transducer at any given instant. Consider the situation at a time $t + \tau$ seconds after the initiation of the pulse. The leading edge of the pulse will have traveled a round-trip distance of $c\ (t + \tau)$ during this time, which means that the scatterers causing the reverberation are located at a distance $\frac{1}{2} c\ (t + \tau)$ from the transducer. The trailing edge of the pulse has been traveling for a time t, so that it is returned by scatterers located at a distance $\frac{1}{2} ct$ from the transducer. Let

$$r = \tfrac{1}{2} ct \tag{1041}$$

Then the reverberation arriving at the transducer at time $t + \tau$ is caused by scatterers located within a spherical shell whose inner radius is r and whose thickness is $\frac{1}{2} c\tau$. In the analysis which follows we shall assume the thickness of the shell to be small compared with the radius, i.e.,

$$\tfrac{1}{2} c\tau << r$$

so that we may use the same value of propagation loss to all portions of the shell. In cases where the pulse is of long duration, suitable corrections must be made for the variations in propagation loss.

Let us now develop an expression for the reverberation level. Let the index intensity in the direction of the beam axis be I_1. The index intensity in any direction (θ, ϕ) is

$$I_1\ \eta(\theta, \phi)$$

where $\eta(\theta, \phi)$ is the relative transmitting response (702). Assuming inverse square spreading without attenuation, the intensity of the outgoing wave at the scattering shell is

$$I = \frac{I_1 \cdot r_1^{\,2}\ \eta(\theta, \phi)}{r^2} \tag{1042}$$

Assume next that the ocean is filled with scatterers, each of which has a scattering cross section σ yd^2, and that there are n_v of these scatterers per cubic yard. If the shielding effect (due to one scatterer being behind another) is negligible, the total scattering area per cubic yard is

$$m_v = \sigma n_v \tag{1043}$$

This quantity m_v is called the volume scattering coefficient and, for the units mentioned above, has the dimensions yd^{-1}. Consider now an infinitesimal element of area dA of the shell. Since the thickness of the shell is $\frac{1}{2}c\tau$ yards, the volume of this element is $\frac{1}{2}c\tau\,dA$ and the scattering area is $\frac{1}{2}c\tau\, m_v dA$. The total acoustic power intercepted by the element is

$$\tfrac{1}{2}c\tau I m_v dA = \tfrac{1}{2}c\tau I m_v r^2 d\Omega$$

where $d\Omega$ is the element of solid angle corresponding to dA. If this power is scattered uniformly in all directions (the assumption on which the measurements of m_v are based), the index intensity of the returning scattered wave is

$$dI_1' = \frac{\frac{1}{2}c\tau I m_v r^2 d\Omega}{4\pi r_1^2} \tag{1044}$$

In returning to the transducer the wave spreads out as $1/r^2$. Its effect on the transducer is proportional to the relative receiving response $\eta'(\theta,\phi)$. Therefore the received reverberation intensity is

$$dI_{RV} = \frac{r_1^2 \eta'(\theta,\phi) dI_1'}{r^2} \tag{1045}$$

Combining (1042), 1044), and (1045) and integrating over the sphere, we obtain the total reverberation intensity

$$I_{RV} = \frac{\frac{1}{2}c\tau r_1^2 I_1}{r^2} \frac{1}{4\pi} \iint m_v \eta(\theta,\phi)\eta'(\theta,\phi)\, d\Omega \tag{1046}$$

If the distribution of scatterers is uniform, so that m_v is constant, the reverberation intensity becomes

$$I_{RV} = \frac{\frac{1}{2}c\tau r_1^2 I_1 m_v}{r^2} \frac{1}{4\pi} \iint \eta(\theta,\phi)\eta'(\theta,\phi)\, d\Omega$$

Comparison with (786) shows that the integral in the above expression is the reverberation factor, d_R. Therefore

$$I_{RV} = \frac{\frac{1}{2}c\tau r_1^2 I_1 m_v d_R}{r^2} \tag{1046a}$$

The volume reverberation level L_{RV} is the decibel equivalent of (1046) or (1046a),

$$L_{RV} = 10 \log I_{RV} \tag{1047}$$

As regards units of measurement of the various quantities, it will be noted that r_1 is the standard reference distance of one yard. If r is measured in yards, then $r_1 = 1$ yd; if r is measured in kiloyards, then $r_1 = 10^{-3}$ kyd. Furthermore, since m_v is measured in yd^{-1}, $c\tau$ must be in yd. This can be done either by expressing τ in seconds and c in yd/sec, or by expressing τ in milliseconds and c in yd/ms (or kyd/sec).

To apply the above theory to practical echo-ranging operations, we must consider the physical situations involved. The simplest case to consider is that of a highly directive transducer having a narrow beam. In this case the reverberation comes from a small volume of the ocean, and the assumption of a constant m_v is reasonably valid. Furthermore, since the reverberating volume is in the immediate vicinity of the target, we may assume that the propagation loss for the reverberation is about the same as that for the target echo. In adapting the reverberation intensity (1046a) to the real world, it will be recalled that we originally had two r^2 factors in the denominator, these being associated with the spreading loss, and one r^2 in the numerator, associated with the area of the spherical shell. In changing from ideal inverse square spreading to actual propagation loss, we replace $20 \log (r/r_1)$ with N_W. Hence the reverberation intensity is

$$I_{RV} = \frac{\frac{1}{2}c\tau r^2 I_1 m_v d_R}{(10^{0.2N_W}) r_1^2} \tag{1048}$$

and the reverberation level, when r is in kyd, is

$$L_{RV} = S_0 - 2N_W - D_R + 10 \log(\tfrac{1}{2}c\tau) + 10 \log m_v + 20 \log r + 60 \tag{1049}$$

where D_R is the reverberation index (787)

If the sonar has a wide beam, the assumption that m_v is constant is of doubtful validity. The bulk of the volume scattering in many areas of the oceans is due to one or more layers of plankton which migrate to different levels at different times of the day. Additional scattering may arise from bubbles near the surface. At depths below the scattering layers the volume scattering coefficient drops to very low values. If the sonar beam is wide enough in the vertical extent to encompass large variations in m_v, it is necessary to compute an average value $\bar{m}_v$ based on the integral appearing in (1046).

and taking into account the refraction of the sound rays. To estimate the spreading loss in this case, it will be noted that a larger quantity of rays is involved, covering a greater spread of depths at the target range. Ideally, the spreading loss which should be used is a weighted average over all the scatterers. For most practical computations the inverse square law including attenuation is probably a closer approximation to reality in this case than the propagation loss to the target. If we make this assumption, the reverberation level is

$$L_{RV} = S_o - D_R + 10 \log(\tfrac{1}{2}c\tau) + 10 \log \bar{m}_v - 20 \log r - 2\alpha r - 60 \qquad (1049a)$$

Volume scattering is also expressed in an alternate form in terms of the volume scattering strength, S_v. The volume scattering strength is the target strength of the scatterers per unit volume and is defined as follows

$$S_v = 10 \log \frac{m_v}{4\pi r_1^2} \qquad (1050)$$

The similarity of volume scattering strength to the target strength of a discrete object, as given by (1008), is obvious if we consider the scattering produced by a small but finite volume of water ΔV. According to the definition of m_v, the total scattering area of the volume ΔV is $m_v \Delta V$, which corresponds to the scattering area σ of the discrete target in (1008). Hence, so far as reverberation is concerned, the volume ΔV is equivalent to a discrete target whose strength is

$$T_{\Delta V} = S_v + 10 \log \Delta V \qquad (1051)$$

The concept of volume scattering strength may be applied to the case of a narrow-beam sonar to yield an expression which gives a simple intuitive picture of the scattering process. Let Ω denote the effective solid angle of the beam, such that

$$d_R = \frac{\Omega}{4\pi} \qquad (1052)$$

Equation (1048) may then be rewritten as follows:

$$I_{RV} = \frac{(\frac{1}{2}c\tau r^2 \Omega)\ I_1 m_v}{4\pi r_1^2\ 10^{0.2N_w}} \qquad (1048a)$$

and the reverberation level, when r is expressed in yards, $c\tau$ in yards, and Ω in steradians, becomes

$$L_{RV} = S_o - 2N_w + S_v + 10 \log (\tfrac{1}{2}c\tau r^2 \Omega) \qquad (1049b)$$

Here we see that $\frac{1}{2}c\tau r^2\Omega$ is the effective scattering volume ΔV which produces the reverberation, and that the last three terms represent the target strength of the scatterers. Comparison with (1005) shows that the reverberation level has been expressed in the same form as the echo level.

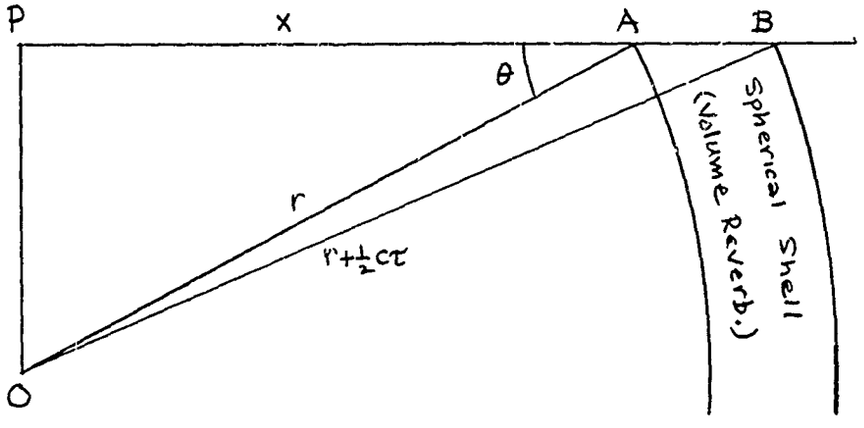
The volume scattering coefficient varies widely from place to place and from time to time. The effect of the biological scattering layers has been mentioned above. Measurements in the Pacific Ocean have shown a strong seasonal effect which has also been correlated with biological activity. The following formula was derived by Dr. H. O. Benecke from some early data taken during World War II:

$$10 \log m_v = -73 + 10 \log f \qquad (1053)$$

where f is in kc.

2. Surface and Bottom Reverberation.

Since the method of analysis is the same for both surface and bottom reverberation, it will suffice to derive the equations applicable to the surface case. Following the method employed for volume reverberation, we shall compute the reverberation received at an instant of time $t+\tau$ seconds after the initiation of the pulse. This reverberation is returned from the scatterers whose range from the transducer is between r and $r+\frac{1}{2}c\tau$, where r is defined by (1041). As may be seen from the sketch below, the portion of the surface which contributes to the scattering is an annular ring whose inner radius is PA and whose outer radius is PB. We shall assume that the pulse length $\frac{1}{2}c\tau$ is small compared with



the range r, so that all the rays between CA and OB strike the surface at substantially the same angle θ.

Temporarily assuming inverse square spreading without attenuation, we see that the intensity of the outgoing wave as it strikes the surface is given by (1042). We shall define the scattering properties of the surface in terms of the surface scattering coefficient, m_s, which is defined as the scattering area per square yard of the surface. The surface scattering coefficient is a dimensionless number which is a measure of the fraction of the surface area which produces the scattering. The total acoustic power intercepted by an element of area dA is $Im_s dA$. If x is the horizontal distance PA, the element of area is

$$dA = x\, dx\, d\phi$$

where ϕ is the azimuth angle measured in the plane of the surface. It may be seen from simple geometry that

$$x\, dx = r\, dr$$

so that

$$dA = r\, dr\, d\phi$$

In computing the index intensity dI_1' of the scattered radiation from dA, we note that virtually all of the energy is scattered back into the water and is therefore confined to a hemisphere. Hence

$$dI_1' = \frac{I\, m_s\, r\, dr\, d\phi}{2\pi r_1^2} \tag{1054}$$

Finally, the effective intensity sensed by the transducer is

$$dI_{RS} = \frac{r_1^2\, \eta'(\theta,\phi)\, dI_1'}{r^2} \tag{1055}$$

Combining (1042), (1054), and (1055), we obtain

$$dI_{RS} = \frac{I_1\, r_1^2\, m_s\, \eta(\theta,\phi)\, \eta'(\theta,\phi)}{2\pi r^3}\, dr\, d\phi$$

In carrying out the integration over the ring we integrate the radial coordinate from r to $r+\frac{1}{2}c\tau$. In accordance with the assumption that $\frac{1}{2}c\tau$ is small compared with r, we may treat both r and θ as constants and simply multiply by $\frac{1}{2}c\tau$. The resultant intensity is therefore

$$I_{RS} = \frac{\frac{1}{2}c\tau r_1^2 I_1 m_s}{2\pi r^3} \int_0^{2\pi} \eta(\theta,\phi)\eta'(\theta,\phi)\, d\phi \qquad (1056)$$

Let us now remove the restriction imposed by the original assumption of inverse square spreading by replacing $20 \log (r/r_1)$ with N_w, so that

$$\frac{r_1^2}{r^3} = \frac{r}{r_1^2}\left(\frac{r_1}{r}\right)^4 \longrightarrow \frac{r}{r_1^2\, 10^{0.2N_w}}$$

Also, the integral in (1056) defines an effective beamwidth $\Delta\phi$,

$$\Delta\phi = \int_0^{2\pi} \eta(\theta,\phi)\eta'(\theta,\phi)\, d\phi \qquad (1057)$$

The generalized form of (1056) is

$$I_{RS} = \frac{\frac{1}{2}c\tau r m_s I_1 \Delta\phi}{2\pi r_1^2\, 10^{0.2N_w}}. \qquad (1056a)$$

In a manner analogous to the treatment of volume reverberation, this result may be expressed in terms of the surface scattering strength, which is the target strength of the surface per unit area, defined as follows:

$$S_s = 10 \log \frac{m_s}{2\pi r_1^2} \qquad (1058)$$

From the definition of m_s, the total scattering area of any small but finite area ΔA is $m_s \Delta A$. So far as reverberation is concerned, this area is equivalent to a discrete target whose strength is

$$T_{\Delta A} = S_s + 10 \log \Delta A \qquad (1059)$$

Inserting (1058) into (1056a), we find for the surface reverberation level

$$L_{RS} = 10 \log I_{RS}$$
$$= S_o - 2N_w + S_s + 10 \log (\tfrac{1}{2}c\tau r\Delta\phi) \qquad (1060)$$

where r is in yards, $c\tau$ is in yards, and $\Delta\phi$ is in radians. It will be noted that $\frac{1}{2}c\tau r\Delta\phi$ is the effective area of the surface insonified by the sonar beam and corresponds to ΔA in (1059).

To apply these results to bottom reverberation we must insert values of the bottom scattering strength S_b in place of S_s. The corresponding formula for the bottom reverberation level may be written as

$$L_{RB} = S_o - 2N_w + S_b + 10 \log (\tfrac{1}{2}c\tau r\Delta\phi) \qquad (1061)$$

Equations (1060) and (1061) express the surface and bottom reverberation levels in a form similar to that of (1005) for the echo level.

It should be noted that in cases where the pulse is of long duration the approximations made above are not valid and a more detailed analysis is needed.

Considerable work has been done in recent years, both theoretical and experimental, in determining surface and bottom scattering strengths. Recent surface measurements by Chapman and Harris* have led to useful formulas for the surface scattering strength as a function of wind speed and grazing angle in the frequency range from 400 to 6400 cps. At wind speeds below about 15 knots they found that the scattering strength increases linearly with the logarithm of the grazing angle at angles above approximately 15 degrees. At smaller angles the scattering strength was found to be essentially constant, with an abrupt change between the two regions. The constant portion at the lower angles is believed to be due to volume scattering of biological origin. At wind speeds of 20 knots or more the linear dependence was observed down to the lowest grazing angles measured, indicating that the surface scattering is predominant. The following empirical equation was derived from the measurements:

$$S_s = 3.3\beta \log(\theta/30) - 42.4 \log\beta + 2.6 \qquad (1062)$$

where β is the slope of the curve of S_s vs. $\log\theta$, expressed in decibels per grazing angle doubled and is empirically related to the wind speed v (knots) and the frequency f (cps) as follows:

$$\beta = 158\,(vf^{1/3})^{-0.58} \qquad (1063)$$

A large amount of data on bottom scattering characteristics has been accumulated over the years, but the wide variability of the sea bottom from one area to another makes it impossible to express the scattering strength adequately in any one mathematical formula. Two different assumptions are widely used regarding the angular dependence of the scattered intensity: (1) that each element of the bottom acts as an omnidirectional source of scattered sound, and (2) that the intensity of the sound scattered by each element of the bottom is proportional to the sine of the angle the scattered wave makes with the plane of the bottom (Lambert's law). In this discussion we shall restrict our attention to the

- - - - - - - - -

*R. P. Chapman and J. H. Harris, Jour. Acoust. Soc. Am., Vol. 34, pp. 1592-1597, Oct. 1962.

reverberation which is returned in the direction of the source, although there are applications in which other scattering angles are of interest. If the incident wave has an intensity I and strikes the bottom at a grazing angle θ, the acoustic power falling on a small area ΔA is $I\Delta A \sin\theta$. According to the first assumption the index intensity of the scattered wave is

$$I_1' = \frac{I \mu_1 \Delta A \sin\theta}{2\pi r_1^2}$$

where μ_1 is a constant. The corresponding scattering strength, from (1004) and (1059), is

$$S_b = k_1 + 10 \log \sin\theta \qquad (1064)$$

where

$$k_1 = 10 \log \frac{\mu_1}{2\pi r_1^2}$$

According to the second assumption, the intensity of the scattered wave is proportional to $\sin^2\theta$ and the scattering strength is

$$S_b = k_2 + 20 \log \sin\theta \qquad (1065)$$

Some of the experimental data appear to favor (1064), and some to favor (1065). Representative values of k_2 are in the vicinity of

$$k_2 = -28 \text{ db}$$

Recent measurements by Urick (NOL) and Saling (Daystrom)* and by Burstein and Keane (NADC)**, using explosive sound sources, show fair agreement with the sine-squared law at grazing angles less than about 60 degrees, but indicate a much higher scattering strength at angles nearer the vertical. These results suggest that a large part of the incident energy is scattered at angles in the vicinity of the direction of specular reflection. The data are more nearly represented by an empirical equation of the form**

$$S_b = k_3 + k_4 \log \tan\theta \qquad (1066)$$

Unfortunately, the values obtained for k_3 and k_4 in three different ocean areas vary widely.

- - - - - - - - - - -

*R. J. Urick and D. S. Saling, Jour. Acoust Soc. Am., Vol. 34, pp. 1721-1724, Nov. 1962.

**A. W. Burstein and J. J. Keane, Jour. Acoust. Soc. Am., Vol. 36, pp. 1596-1597, Aug. 1964.

3. Recognition Differential.

The aural recognition differential against reverberation, which we shall call M_R, depends upon both the ping duration τ and the Doppler shift Δf. The following information is based on World War II data taken from Vol. 7, Div. 6 of the NDRC Reports, "Principles of Underwater Sound." In the absence of Doppler, the recognition differential decreases with increasing ping length. For pings longer than about 40 milliseconds the experimental curve may be fit reasonably well with the formula

$$3 - 10 \log \frac{\tau}{200},$$

where τ is in milliseconds. The slope of the experimental curve tends to flatten off at shorter ping lengths.

The presence of a Doppler shift of about 20 cps or more, either up or down, causes a sharp decrease in the recognition differential, illustrating the ability of the ear to distinguish the difference in tone. Dr. H. O. Benecke, when at the Naval Air Development Center, fit the experimental data with the following equation:

$$M_R = 3 - 10 \log \left[\frac{\tau}{200} + 0.39 \left(\frac{\tau \Delta f}{1000} \right)^2 \right]$$

which may be transformed as follows

$$M_R = 26 - 10 \log \tau - 10 \log \left(1 + 0.375 \times 10^{-4} \tau f^2 v_T^2\right) \qquad (1067)$$

where

τ = ping duration, milliseconds

Δf = Doppler shift, cps

f = sonar frequency, kc

v_T = target Doppler speed, knots

4. Reverberation-Limited Ranges.

The sonar equation for reverberation-limited ranges is basically the same as for noise-limited ranges. In lieu of (1032) we have the following

$$S_o + T - 2N_w = L_R + M_R + \Delta L_E \qquad (1068)$$

There are two differences between the noise-limited and reverberation-limited equations.

(1) The probability distribution of the echo excess ΔL_E is not

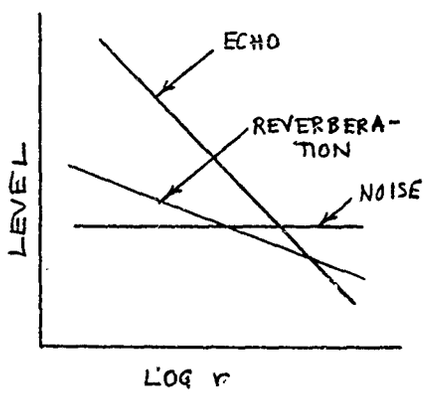
the same, due to the fact that reverberation tends to be more variable than noise. The standard deviation of the reverberation level for a single-frequency pulse appears to be about 5 db. If this is added to all the other factors, the resultant standard deviation is about 1 db higher than the value for the noise case. However, considering the uncertainties involved, this difference may be ignored and the same value of 12 db may be used.

(2) The reverberation level is a function of the range, a fact which complicates somewhat the solution of (1068). Probably the most straightforward method of solving for the range is to compute each side of the equation separately, plotting the two curves as functions of the range r. The limiting range is then the point at which the curves cross. The effect of noise (right-hand side of (1032)) may also be plotted as a horizontal line on the same graph.

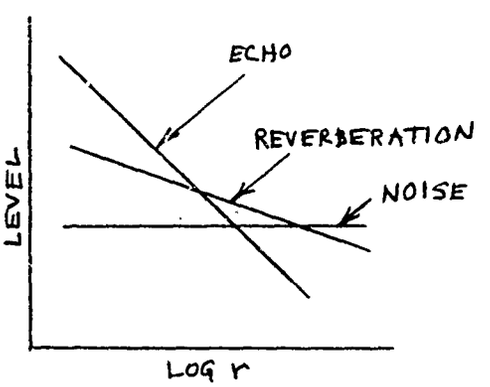
It will be noted that in the ideal case of inverse square spreading the echo, reverberation, and noise levels have the following range dependence:

Echo level:	varies as -40 log r
Surface reverberation level:	varies as -30 log r
Volume reverberation level:	varies as -20 log r
Ambient noise level:	constant

When the echo level and the reverberation and noise levels (including the respective recognition differentials) are plotted as functions of log r, two possible conditions may occur, as indicated schematically in the following sketches:



NOISE-LIMITED



REVERBERATION-LIMITED

In the first situation the echo drops below the noise at a shorter range than it drops below the reverberation; the sonar is noise-limited. In the second situation the echo drops below the reverberation level first, and the sonar is reverberation-limited.

E. Special Types of Pulses; Some Signal Processing Considerations.

The subject of signal processing was introduced in a somewhat indirect way in the discussion of signal detection. The principal purpose of the signal processing in that context was to enhance the detectability of signals relative to background noise. Signal processing is useful for other purposes as well. It can be applied to discriminate against reverberation. It can also improve the resolution of the sonar in the measurement of both range and Doppler speed. The degree of signal processing gain which can be achieved is dependent also upon the characteristics of the sonar pulse which is transmitted. In this section we shall examine a few of the types of pulses which are being used in modern sonars and shall discuss briefly some of their advantages. It must be recognized, of course, that the field of signal processing is tremendously complex and diversified; both the scope and the depth of the present discussion are extremely limited.

1. CW Pulse.

This term refers to the conventional single-frequency pulse. Consider first the processing gain against noise. We shall define the gain as the ratio of the output signal-to-noise ratio to the input signal-to-noise ratio. Let

S = average signal power,

N_o = noise power per unit bandwidth,

W_{in} = input bandwidth, that is, the bandwidth to which the incoming wave is filtered before being processed,

W_{out} = output bandwidth,

T = duration of the pulse.

A necessary requirement for good processing is that the bandwidths of all components in the system be at least as wide as the bandwidth of the signal passing through them, so that no appreciable part of the signal is lost. We shall make the assumption that this is true. We shall also assume that the spectrum of the noise is sufficiently flat that the noise power N in any band W may be expressed as

$$N = N_o W \tag{1101}$$

The input noise power is $N_o W_{in}$. Hence the input signal-to-noise ratio is

$$\left(\frac{S}{N}\right)_{in} = \frac{S}{N_o W_{in}} \tag{1102}$$

If no signal power is lost, the output signal-to-noise ratio is

$$\left(\frac{S}{N}\right)_{out} = \frac{S}{N_o W_{out}} \tag{1103}$$

The signal processing gain against the noise is the ratio of (1103) to (1102)

$$\frac{(S/N)_{out}}{(S/N)_{in}} = \frac{W_{in}}{W_{out}} \tag{1104}$$

As an example, consider the aural detection of a 10 kc pulse. If the sonar is designed to operate in a stationary position against targets with a maximum Doppler speed of 30 knots, the input bandwidth, according to (1039a) and (1040), is

$$W_{in} = 2 \times 0.7 \times 10 \times 30 = 420 \text{ cps}$$

The output bandwidth is the critical bandwidth of the ear, about 50 cps. The gain is therefore

$$\frac{420}{50} = 8.4 \qquad \text{or about 9 db.}$$

Suppose that instead of using the ear, we provide a bank of output filters and detect the output of each of these filters. The minimum bandwidth that can be used is the bandwidth of the signal. It is shown in standard books on Fourier analysis that a "single-frequency" pulse of finite duration T has a spectrum of finite width, the bulk of the energy falling within a band whose width is approximately $1/T$ and which is centered about the "single frequency". Therefore the minimum bandwidth of each of the output filters is approximately

$$W_{out} = \frac{1}{T} \tag{1105}$$

and the gain is

$$\frac{(S/N)_{out}}{(S/N)_{in}} = W_{in}\, T \tag{1106}$$

It is interesting to compare this result with equation (927), which was derived in the discussion of statistical detection theory for the case where the signal is known exactly. In (927) S/N refers to the input signal-to-noise ratio, W is the input bandwidth, and the parameter d may be considered to

be the output signal-to-noise ratio. The theoretical gain is therefore $2W_{in}T$, which is twice the value obtained for the filtered CW pulse. The essential difference here is that the CW signal is not known "exactly". All that can be detected at the output of the filter is the envelope of the pulse; all phase information is lost.

Returning to (1106), we see that the theoretical gain can be increased by increasing the length of the pulse. There are of course practical limits, such as the complexity of a very large bank of extremely narrow-band filters.

Consider now the problem of reverberation. We have seen that the intensity of the reverberation is proportional to the pulse length. This, of course, is due to the fact that the extent of the insonified volume (or area) which contributes to the reverberation is proportional to the pulse length. It will be recalled, however, that the frequency content of the reverberation is limited to the bandwidth of the transmitted pulse, plus a small amount of Doppler from the scatterers. Therefore, if the target has sufficient Doppler to remove the echo from the bandwidth of the reverberation, the two can be separated by means of the bank of output filters. Consider, for example, a 100 ms pulse of frequency 10 kc. The minimum output bandwidth is about 10 cps. According to (1039a) the minimum Doppler resolution is about

$$10 = 0.7 \times 10\, v_T$$

or

$$v_T = 1.4 \text{ knots}$$

Although this figure may be somewhat optimistic, it gives an idea of the discrimination available.

Finally, let us consider the resolution capability in range and Doppler. The importance of good range resolution is obvious when one considers the need for accurate localization of targets. Doppler resolution is also important, in that it provides information concerning one component of the target velocity. (The other component can be estimated from the bearing rate.) To simplify the discussion we shall express range in terms of time and Doppler speed in terms of the frequency shift, as expressed by (1041) and (1037) (neglecting self-Doppler). Without going through a detailed analysis, it is intuitively evident that the uncertainty in time is of the order of the pulse length T and that the uncertainty in Doppler is of the order of the bandwidth 1/T. A long

pulse has good Doppler resolution but poor range resolution. Conversely, a short pulse has good range resolution but poor Doppler resolution. It is impossible to achieve good performance in both dimensions with the same CW pulse. This limitation is one of the reasons for going to more sophisticated types of pulses.

2. Linear FM Pulse.

A linear FM (frequency-modulated) pulse is a pulse whose frequency increases or decreases linearly with time. In the following discussion we shall assume that it increases. The equation for the frequency as a function of time may be written in the form

$$f = f_0 + mt \tag{1107}$$

where f_0 is the initial frequency and m is the frequency slope in cyc/sec^2. The phase angle ϕ is the integral of the frequency. More specifically,

$$f = \frac{\omega}{2\pi} = \frac{1}{2\pi}\frac{d\phi}{dt} \tag{1108}$$

Hence

$$\phi = 2\pi(f_0 t + \tfrac{1}{2}mt^2) \tag{1109}$$

There are several ways in which a linear FM signal processing system may be implemented. We shall consider a system in which a replica of the transmitted signal is retained and compared with the returning wave in the following manner. The frequency of this reference pulse is reduced by a suitable amount. When the returning wave is received, it is beat with the reference, so that the frequency of the resultant is the difference of the two. The output is then passed through a bank of filters. For convenience of analysis, we shall ignore the frequency shift applied to the reference, so that if at any instant of time the frequency of the incoming wave exceeds that of the reference, the difference frequency will be positive. If the reverse is true, the difference frequency will be negative.

We shall divide the incoming wave into intervals T seconds long and shall process each of these intervals, one at a time. Since we do not know the range of the target, we must introduce a delay of τ seconds into the analysis. Although called a "delay", τ may have any value between -T/2 and +T/2. The effect of the target Doppler is to compress (up-Doppler) or expand (down-Doppler) the time by a factor (1+a), where

$$a = \frac{2v_T}{c} \tag{1110}$$

If we assume an ideal point target, the phase angle of the returning echo is

$$\phi_E = 2\pi\left[(1+a)f_o(t-\tau) + \tfrac{1}{2}(1+a)^2 m(t-\tau)^2\right] \tag{1111}$$

Note that the factor $(1+a)$ enters as a square in the second term because the time is squared. The frequency of the echo is, from (1108),

$$f_E = (1+a)f_o + (1+a)^2 m(t-\tau) \tag{1112}$$

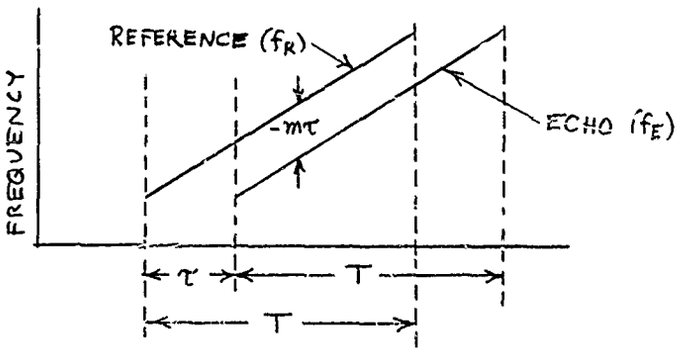
Denoting the frequency of the reference (1107) by f_R and performing the subtraction, we obtain

$$f_E - f_R = af_o + (2a+a^2)mt - (1+a)^2 m\tau \tag{1113}$$

Since a is very small, a^2 is thoroughly negligible. Also, for the purpose of the present simplified discussion we shall neglect a itself in comparison with unity, yielding the following approximation

$$f_E - f_R = af_o - m\tau + 2amt \tag{1113a}$$

Consider now the case where there is no Doppler, i.e., $a = 0$. Here there is a constant difference frequency having the value $-m\tau$. The situation is illustrated in the following sketch:

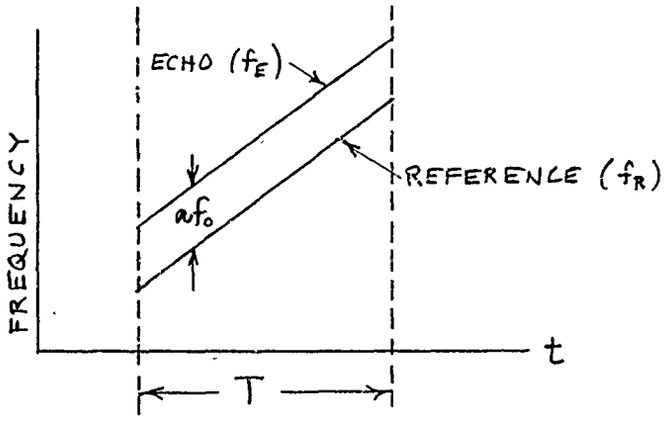


Since the frequency is constant, we may pass the output through a bank of filters whose minimum bandwidth is approximately $1/T$. Thus, if there is no Doppler, the time (and hence range) resolution of the system is approximately

$$\Delta\tau \approx \frac{1}{mT}$$

Suppose next that the range is known exactly ($\tau = 0$) but the echo has Doppler. The term $2amt$ in (1113a) is fairly small and we shall temporarily neglect it.

With this approximation we see that the difference frequency is again constant and has the value af_0 (which is the Doppler shift Δf of (1037)). The situation is as shown below:



With the same filters as before, the Doppler resolution is

$$\Delta(af_0) \approx \frac{1}{T}$$

Actually, neither the range nor the Doppler is known, but only the combination $af_0 - m\tau$. If we observe a given value of difference frequency $f_E - f_R$, then we may say that we know the value of this combination to within approximately

$$\Delta(af_0 - m\tau) \approx \frac{1}{T} \qquad (1114)$$

With a linear FM pulse it is possible to determine the Doppler accurately if the range is known, and vice versa. If we plot af_0 vs. τ, there is a region of ambiguity around the line

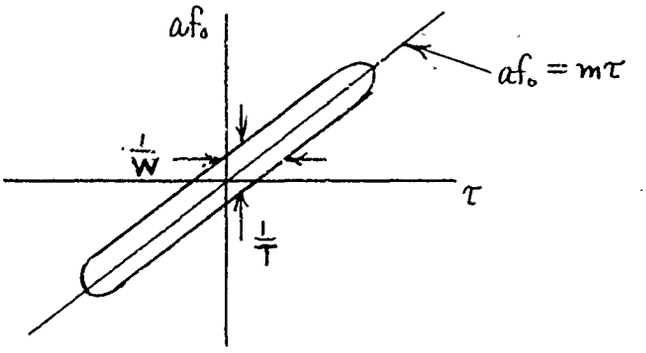
$$af_0 - m\tau = \text{constant}$$

within which the coordinates of the target cannot be resolved. The extent of this region is usually determined by computing the correlation function of the echo and reference and drawing a contour for a constant value of this function. (Some use a value of 0.5, others 0.707.) The resulting plot is called an ambiguity diagram. The following sketch illustrates a typical ambiguity diagram for a linear FM pulse. The width of the diagram in the frequency dimension (constant τ) is of the order of 1/T, as indicated by (1114). The width of the diagram in the time dimension (constant Doppler shift) is of the order of

$$\frac{1}{mT} = \frac{1}{W}$$

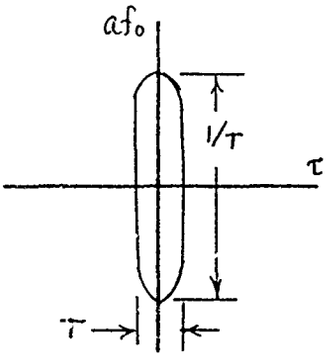
where W is the bandwidth of the signal

$$W = mT \tag{1115}$$

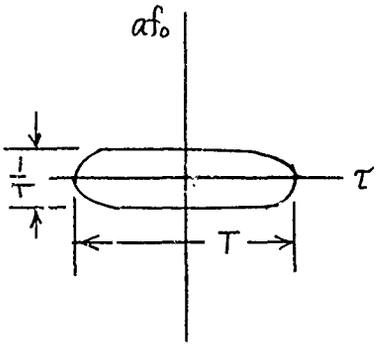


LINEAR FM PULSE

Ambiguity diagrams may also be drawn for CW pulses, as indicated below.



SHORT CW PULSE



LONG CW PULSE

We must now consider the term $2amt$ which was ignored previously. This term describes the change in slope of the frequency-time line due to the Doppler, and its presence in (1113a) shows that the difference frequency $f_E - f_R$ is not really constant. If this term is sufficiently large, $f_E - f_R$ will spread out over more than one filter and will cause a degradation in the output. A reasonable criterion for an upper limit on the tolerable error is that $f_E - f_R$ should not change more than $1/T$ during the pulse time T. This leads to the inequality

$$a < \frac{1}{2mT^2} \tag{1116}$$

or

$$\frac{2v_T}{c} < \frac{1}{2WT} \tag{1116a}$$

If Doppler speeds in excess of the maximum indicated by (1116a) are to be encountered, it is necessary to generate additional references which are artificially Dopplerized.

Another problem occurs if the target is accelerating. If the target speed changes during the time T, the resulting change in af_o will cause a change in the output difference frequency $f_E - f_R$. Let v_1 be the speed at the beginning of the echo and v_2 at the end. Then the change in $f_E - f_R$ is $\frac{2(v_2-v_1)f_o}{c}$. Using the same criterion as before, namely, that this change should not exceed $1/T$, we arrive at the inequality

$$v_2 - v_1 < \frac{c}{2f_oT}$$

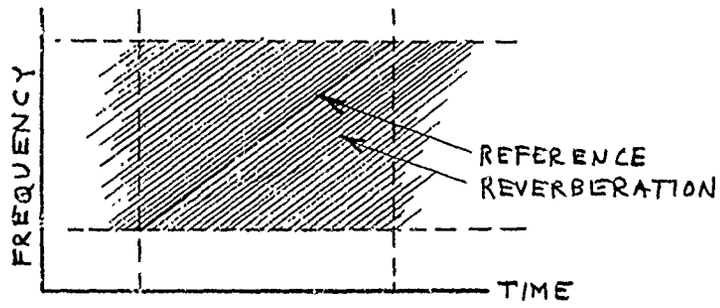
or

$$\dot{v} < \frac{c}{2f_oT^2} \tag{1117}$$

where $\dot{v}$ is the average target acceleration

$$\dot{v} = \frac{v_2 - v_1}{T} \tag{1118}$$

The processing gain of the linear FM system against noise is substantially the same as that of a CW pulse of the same duration because of the same output bandwidth. On the other hand, there is a large difference in the processing gain against reverberation. If we assume that the scatterers are uniformly distributed in range, the reverberation returned at any given instant will be spread out uniformly over the bandwidth W (1115), as indicated below:



Since the output filter has a bandwidth $1/T$, only a fraction $1/WT = 1/mT^2$ of the total reverberation appears in the output. We saw earlier that the thickness of the reverberating shell (for volume reverberation) and the width of the reverberating annulus (for surface reverberation) are equal to $\frac{1}{2}cT$. The linear FM system

effectively cuts this distance into WT pieces and rejects all but one of them. As regards reverberation, this is equivalent to shortening the pulse from T to 1/W seconds. The FM pulse can be used in cooperation with the CW pulse, the CW being most useful against high-Doppler targets and the FM against low-Doppler targets.

3. Doppler-Invariant Pulse.

We have seen that the presence of Doppler in the echo of a linear FM pulse causes a change in the slope of the frequency-time curve and introduces a residual "chirp" in the difference frequency when the echo is beat against the reference. Because of this effect, it is necessary to generate a number of artificially Dopplerized references if the system is to operate effectively over a wide range of Doppler speeds. This difficulty can be overcome by transmitting a so-called Doppler-invariant pulse, in which the reciprocal of the frequency varies linearly with time, or

$$f_R = \frac{f_o}{1 - kt} \tag{1119}$$

where k is a constant. The phase angle of the reference has the form

$$\phi_R = -\frac{2\pi f_o}{k} \log_e (1 - kt) \tag{1120}$$

The phase angle of the echo, including Doppler shift and time delay, is

$$\phi_E = -\frac{2\pi f_o}{k} \log_e \left[1 - (1+a)k(t-\tau)\right] \tag{1121}$$

Taking the derivative of ϕ_E, we obtain for the frequency of the echo

$$f_E = \frac{(1 + a)f_o}{1 - (1 + a)k(t - \tau)} \tag{1122}$$

If we neglect a in comparison with unity, as was done for the linear FM pulse in (1113a), the difference frequency is

$$f_E - f_R = \frac{(a - k\tau)f_o}{(1 - kt)\left[1 - k(t - \tau)\right]} \tag{1123}$$

Whenever the Doppler shift and time delay are related in such a way that

$$a = k\tau \tag{1124}$$

the difference frequency is zero regardless of the actual value of the Doppler shift. It is in this sense that the pulse is "Doppler-invariant". It is seen that to take advantage of the invariant nature of the pulse it is necessary to

search in time delay until (1124) is satisfied.

The Doppler-invariant pulse is sometimes called a linear-period pulse, since the period (1/f) varies linearly with time. The similarity between the two pulses is evident when (1119) is expanded in a series

$$f_R = f_o(1 + kt + k^2t^2 + \cdots) \tag{1119a}$$

If we let

$$k = \frac{m}{f_o} \tag{1125}$$

(1119a) becomes

$$f_R = f_o + mt + \frac{m^2t^2}{f_o} + \cdots \tag{1119b}$$

Comparison with (1107) shows that the two expressions differ only by terms in t^2 and above , which are relatively small (though not negligible).

4. Pseudo-Random Noise Pulse

The so-called pseudo-random noise pulse, when used in conjunction with a correlator, is very attractive as a signal processing system and is now in wide use in sonar systems. Let us temporarily ignore the "pseudo" and imagine that a sample of white Gaussian noise of bandwidth W and duration T is transmitted by a sonar projector. An exact replica of the transmitted pulse is stored in a memory circuit and is correlated with the returning wave received by the hydrophone array. Let $x_R(t)$ represent the waveform of the stored reference and $x_E\left[(1+a)(t-\tau)\right]$ represent the waveform of the echo which involves a Doppler shift a $(=2v_T/c)$ and a time delay τ. The output of an ideal correlator is

$$y(\tau,a) = \int_0^T x_R(t)\, x_E\left[(1+a)(t-\tau)\right] dt \tag{1126}$$

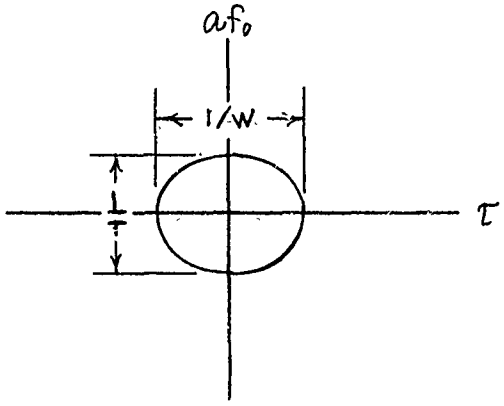
Assume for the moment that there is no Doppler shift (a = 0). The time delay must be inserted into (1126) because the exact time of arrival of the echo is not known. In order to obtain the maximum output corresponding to $\tau = 0$, a large number of correlations must be performed on the incoming wave as it is received. Since each correlation is a discrete process, τ must be varied in discrete steps, but if the interval $\Delta\tau$ between successive correlations is sufficiently small, the condition of exact time coincidence may be approximated to a sufficient degree. When this has been accomplished, the correlation function (for zero Doppler) is

$$y(0,0) = \int_0^T x_R(t)\, x_E(t)\, dt \tag{1127}$$

Since the signal $x_R(t)$ is (almost) exactly known, this is the output required by (915) for an optimum receiver, Case 1. The signal processor operating with this type of pulse is therefore, except for hardware limitations which preclude exact implementation of (1127), an optimum receiver with a signal processing gain of 2WT.

It should be pointed out that the random noise pulse referred to here applies to Case 1 (signal known exactly) rather than to Case 2 (white Gaussian noise signal), because the noise in this case is no longer a random process; it is known by virtue of the storage of an exact replica of it.

The resolution capability of a random noise pulse is found by evaluation of the integral (1126) as a function of the Doppler shift and time delay, a task which is beyond the scope of these notes. It is found that the region of ambiguity, defined as for the linear FM pulse, is enclosed within approximately the curve shown in the sketch below, in which the ordinate is the Doppler shift af_o where f_o refers to the center frequency of the signal band W. It is seen here



that the noise pulse has a high resolution in both range and Doppler.*

As regards reverberation discrimination, it is seen that the effective pulse length is reduced from T to 1/W, representing a gain equivalent to that predicted for the linear FM pulse.

There are a number of practical problems associated with the implementation of the random noise pulse system, and these have led to a host of new hardware developments--a process which is continuing at an accelerating pace. Allen and Westerfield (NEL) have published a comprehensive summary of the field.**

One of the problems has to do with the difficulty of storing an

- - - - - - - - -

*Because of the high Doppler resolution it is necessary to generate a large number of Dopplerized references.

**W. B. Allen and E. C. Westerfield, J.A.S.A., Vol.36, pp. 121-139, Jan. 1964.

undistorted replica of the transmitted pulse. A current solution to this problem is the generation of a pseudo-random noise pulse by means of a shift-register encoder (SRE). This is a device consisting of a shift register with a feedback arrangement. The shift register has been mentioned previously in connection with transducer systems (electrically steered arrays). In the SRE the output of the last stage in the shift register is added to the output of one or more of the earlier stages, using binary arithmetic (0+0 = 0; 0+1 = 1; 1+0 = 1; 1+1 = 0) and the resultant is fed back into the first stage. The shift register has 1's in all stages at the start. Upon initiation of the pulse, the feedback process continues step by step, and the digits (1 or 0) are passed along from one stage to the next down the shift register. The output from the last stage is fed through a bandpass filter and then on into the transducer. The feedback process results in a random-appearing code of 1's and 0's. The code is not really random, however, but is a completely determined, repeating sequence of finite length. When properly designed, the maximum length of the code is $2^n - 1$ bits, where n is the number of stages in the shift register.

The advantage of the pseudo-random code of the SRE is that it circumvents the necessity of storing the pulse for long periods of time, which is a difficult thing to do. After the pulse has been transmitted, the SRE is reset to its initial configuration. Then after any desired delay time, the identical pulse can be re-generated merely by reactivating the shift register.

Although the bandpass filter introduces some smoothing, the pseudo-random pulse is essentially digital and the correlation is usually done digitally. Both the incoming wave and the reference are infinitely clipped and only the signs are retained.

A second problem is that of the requirement for a very large number of correlations. Each correlation involves a piece of data T seconds long. But during this time enough correlations must be performed to provide the desired resolution in τ. If this resolution is $\Delta\tau$, the required number of correlations is $T/\Delta\tau$, a number which may run into the hundreds.

This problem has been solved by means of time compression. One of the first devices developed for this purpose is the DELTIC (DElay-Line TIme Compressor). A DELTIC is a device containing a delay line through which digits (1 or 0) are circulated. Two DELTICS are required for a correlation, one for the reference and one for the incoming wave. The latter is called the signal DELTIC.

The signal is sampled at a rate sufficiently high to retain a high percentage of the information contained in the wave. Let N denote the number of samples contained in the transmitted pulse. Then N-1 samples are circulated in the signal DELTIC, making one circuit during the interval of time between the arrival of successive samples. Because there is one too few samples in the train, one sample--the oldest--is discarded on each circulation cycle, and this is replaced with the next sample arriving from the hydrophone. The signal DELTIC is thus continuously up-dated as the returning signal arrives from the water.

In the reference DELTIC all N samples of the reference are circulated continuously. During any one circulation cycle the outputs of both DELTICS are cross-correlated. That is, during the interval of time between the arrival of successive samples of the signal, the correlation of the entire train T seconds long is performed. In this way it is possible to perform N correlations of data T seconds long in a time of T seconds. The DELTIC correlator in effect compresses time by a factor of N and provides a resolution in τ of

$$\Delta\tau = \frac{T}{N} \qquad (1128)$$

The output of the correlator is averaged by a bank of bandpass filters.

The DELTIC is actually only the beginning of a series of developments of signal processing devices in a field which is expanding rapidly.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Air Warfare Research Department U. S. Naval Air Development Center Johnsville, Warminster, Pennsylvania 18974	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP -

3. REPORT TITLE

LECTURE NOTES ON UNDERWATER ACOUSTICS

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

-

5. AUTHOR(S) (Last name, first name, initial)

BARTBERGER, CHARLES L.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
17 MAY 1965	415	10

8a. CONTRACT OR GRANT NO. -	9a. ORIGINATOR'S REPORT NUMBER(S) NADC-WR-6509
b. PROJECT NO. -	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. AVAILABILITY/LIMITATION NOTICES

QUALIFIED REQUESTERS MAY OBTAIN COPIES OF THIS REPORT DIRECT FROM DDC

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	U.S. Naval Air Development Center Johnsville, Warminster, Pa.

13. ABSTRACT

This report consists of a set of lecture notes for a course in underwater acoustics, given at the U. S. Naval Air Development Center. The course is based to a considerable extent upon J. W. Horton's book, "Fundamentals of Sonar," and covers the following topics: physical properties of acoustic waves in water, relative magnitudes and transmission loss, propagation, transducers, indicators and recorders, detection of signals, echo-ranging, and a brief discussion of signal processing for active sonars. The course is intended primarily to present basic principles, and the portions dealing with hardware and practical applications are somewhat sketchy.

DD FORM 1 JAN 64 1473

UNCLASSIFIED

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
UNDERWATER ACOUSTICS SONAR PROPAGATION, ACOUSTIC TRANSDUCERS, ACOUSTIC ECHO-RANGING, ACOUSTIC						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through

_____."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through

_____."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through

_____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS), (S), (C), or (U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.

DD FORM 1 JAN 64 1473 (BACK)

UNCLASSIFIED

THE FOLLOWING TOPICS: PHYSICAL PROPERTIES OF ACOUSTIC WAVES IN WATER, RELATIVE MAGNITUDES AND TRANSMISSION LOSS, PROPAGATION, TRANSDUCERS, INDICATORS AND RECORDERS, DETECTION OF SIGNALS, ECHO-RANGING, AND A BRIEF DISCUSSION OF SIGNAL PROCESSING FOR ACTIVE SONARS. THE COURSE IS INTENDED PRIMARILY TO PRESENT BASIC PRINCIPLES, AND THE PORTIONS DEALING WITH HARDWARE AND PRACTICAL APPLICATIONS ARE SOMEWHAT SKETCHY.

UNCLASSIFIED

THE FOLLOWING TOPICS: PHYSICAL PROPERTIES OF ACOUSTIC WAVES IN WATER, RELATIVE MAGNITUDES AND TRANSMISSION LOSS, PROPAGATION, TRANSDUCERS, INDICATORS AND RECORDERS, DETECTION OF SIGNALS, ECHO-RANGING, AND A BRIEF DISCUSSION OF SIGNAL PROCESSING FOR ACTIVE SONARS. THE COURSE IS INTENDED PRIMARILY TO PRESENT BASIC PRINCIPLES, AND THE PORTIONS DEALING WITH HARDWARE AND PRACTICAL APPLICATIONS ARE SOMEWHAT SKETCHY.

UNCLASSIFIED

THE FOLLOWING TOPICS: PHYSICAL PROPERTIES OF ACOUSTIC WAVES IN WATER, RELATIVE MAGNITUDES AND TRANSMISSION LOSS, PROPAGATION, TRANSDUCERS, INDICATORS AND RECORDERS, DETECTION OF SIGNALS, ECHO-RANGING, AND A BRIEF DISCUSSION OF SIGNAL PROCESSING FOR ACTIVE SONARS. THE COURSE IS INTENDED PRIMARILY TO PRESENT BASIC PRINCIPLES, AND THE PORTIONS DEALING WITH HARDWARE AND PRACTICAL APPLICATIONS ARE SOMEWHAT SKETCHY.

UNCLASSIFIED

THE FOLLOWING TOPICS: PHYSICAL PROPERTIES OF ACOUSTIC WAVES IN WATER, RELATIVE MAGNITUDES AND TRANSMISSION LOSS, PROPAGATION, TRANSDUCERS, INDICATORS AND RECORDERS, DETECTION OF SIGNALS, ECHO-RANGING, AND A BRIEF DISCUSSION OF SIGNAL PROCESSING FOR ACTIVE SONARS. THE COURSE IS INTENDED PRIMARILY TO PRESENT BASIC PRINCIPLES, AND THE PORTIONS DEALING WITH HARDWARE AND PRACTICAL APPLICATIONS ARE SOMEWHAT SKETCHY.

UNCLASSIFIED

UNCLASSIFIED

U. S. NAVAL AIR DEVELOPMENT CENTER, JOHNSVILLE, WARMINSTER, PENNSYLVANIA
AIR WARFARE RESEARCH DEPARTMENT

LECTURE NOTES ON UNDERWATER ACOUSTICS; BY C. L. BARTBERGER; 17 MAY 1965; 415 P. REPORT: UNCLASSIFIED SECOND PRINTING 20 SEP 1965

THIS REPORT CONSISTS OF A SET OF LECTURE NOTES FOR A COURSE IN UNDERWATER ACOUSTICS, GIVEN AT THE U. S. NAVAL AIR DEVELOPMENT CENTER. THE COURSE IS BASED TO A CONSIDERABLE EXTENT UPON J. W. HORTON'S BOOK, "FUNDAMENTALS OF SONAR," AND COVERS

1. UNDERWATER ACOUSTICS
2. SONAR
3. PROPAGATION, ACOUSTIC
4. TRANSDUCERS, ACOUSTIC
5. ECHO-RANGING, ACOUSTIC

REPORT NADC-WR-6509

UNCLASSIFIED

U. S. NAVAL AIR DEVELOPMENT CENTER, JOHNSVILLE, WARMINSTER, PENNSYLVANIA
AIR WARFARE RESEARCH DEPARTMENT

LECTURE NOTES ON UNDERWATER ACOUSTICS; BY C. L. BARTBERGER; 17 MAY 1965; 415 P. REPORT: UNCLASSIFIED SECOND PRINTING 20 SEP 1965

THIS REPORT CONSISTS OF A SET OF LECTURE NOTES FOR A COURSE IN UNDERWATER ACOUSTICS, GIVEN AT THE U. S. NAVAL AIR DEVELOPMENT CENTER. THE COURSE IS BASED TO A CONSIDERABLE EXTENT UPON J. W. HORTON'S BOOK, "FUNDAMENTALS OF SONAR," AND COVERS

1. UNDERWATER ACOUSTICS
2. SONAR
3. PROPAGATION, ACOUSTIC
4. TRANSDUCERS, ACOUSTIC
5. ECHO-RANGING, ACOUSTIC

1. REPORT NADC-WR-6509

UNCLASSIFIED

U. S. NAVAL AIR DEVELOPMENT CENTER, JOHNSVILLE, WARMINSTER, PENNSYLVANIA
AIR WARFARE RESEARCH DEPARTMENT

LECTURE NOTES ON UNDERWATER ACOUSTICS; BY C. L. BARTBERGER; 17 MAY 1965; 415 P. REPORT: UNCLASSIFIED SECOND PRINTING 20 SEP 1965

THIS REPORT CONSISTS OF A SET OF LECTURE NOTES FOR A COURSE IN UNDERWATER ACOUSTICS, GIVEN AT THE U. S. NAVAL AIR DEVELOPMENT CENTER. THE COURSE IS BASED TO A CONSIDERABLE EXTENT UPON J. W. HORTON'S BOOK, "FUNDAMENTALS OF SONAR," AND COVERS

1. UNDERWATER ACOUSTICS
2. SONAR
3. PROPAGATION, ACOUSTIC
4. TRANSDUCERS, ACOUSTIC
5. ECHO-RANGING, ACOUSTIC

1. REPORT NADC-WR-6509

UNCLASSIFIED

U. S. NAVAL AIR DEVELOPMENT CENTER, JOHNSVILLE, WARMINSTER, PENNSYLVANIA
AIR WARFARE RESEARCH DEPARTMENT

LECTURE NOTES ON UNDERWATER ACOUSTICS; BY C. L. BARTBERGER; 17 MAY 1965; 415 P. REPORT: UNCLASSIFIED SECOND PRINTING 20 SEP 1965

THIS REPORT CONSISTS OF A SET OF LECTURE NOTES FOR A COURSE IN UNDERWATER ACOUSTICS, GIVEN AT THE U. S NAVAL AIR DEVELOPMENT CENTER. THE COURSE IS BASED TO A CONSIDERABLE EXTENT UPON J. W. HORTON'S BOOK, "FUNDAMENTALS OF SONAR," AND COVERS

1. UNDERWATER ACOUSTICS
2. SONAR
3. PROPAGATION, ACOUSTIC
4. TRANSDUCER, ACOUSTIC
5 ECHO-RANGING, ACOUSTIC

1. REPORT NADC-WR-6509

THE FOLLOWING TOPICS: PHYSICAL PROPERTIES OF ACOUSTIC WAVES IN WATER, RELATIVE MAGNITUDES AND TRANSMISSION LOSS, PROPAGATION, TRANSDUCERS, INDICATORS AND RECORDERS, DETECTION OF SIGNALS, ECHO-RANGING, AND A BRIEF DISCUSSION OF SIGNAL PROCESSING FOR ACTIVE SONARS. THE COURSE IS INTENDED PRIMARILY TO PRESENT BASIC PRINCIPLES, AND THE PORTIONS DEALING WITH HARDWARE AND PRACTICAL APPLICATIONS ARE SOMEWHAT SKETCHY.	UNCLASSIFIED
THE FOLLOWING TOPICS: PHYSICAL PROPERTIES OF ACOUSTIC WAVES IN WATER, RELATIVE MAGNITUDES AND TRANSMISSION LOSS, PROPAGATION, TRANSDUCERS, INDICATORS AND RECORDERS, DETECTION OF SIGNALS, ECHO-RANGING, AND A BRIEF DISCUSSION OF SIGNAL PROCESSING FOR ACTIVE SONARS. THE COURSE IS INTENDED PRIMARILY TO PRESENT BASIC PRINCIPLES, AND THE PORTIONS DEALING WITH HARDWARE AND PRACTICAL APPLICATIONS ARE SOMEWHAT SKETCHY.	UNCLASSIFIED
THE FOLLOWING TOPICS: PHYSICAL PROPERTIES OF ACOUSTIC WAVES IN WATER, RELATIVE MAGNITUDES AND TRANSMISSION LOSS, PROPAGATION, TRANSDUCERS, INDICATORS AND RECORDERS, DETECTION OF SIGNALS, ECHO-RANGING, AND A BRIEF DISCUSSION OF SIGNAL PROCESSING FOR ACTIVE SONARS. THE COURSE IS INTENDED PRIMARILY TO PRESENT BASIC PRINCIPLES, AND THE PORTIONS DEALING WITH HARDWARE AND PRACTICAL APPLICATIONS ARE SOMEWHAT SKETCHY.	UNCLASSIFIED
THE FOLLOWING TOPICS: PHYSICAL PROPERTIES OF ACOUSTIC WAVES IN WATER, RELATIVE MAGNITUDES AND TRANSMISSION LOSS, PROPAGATION, TRANSDUCERS, INDICATORS AND RECORDERS, DETECTION OF SIGNALS, ECHO-RANGING, AND A BRIEF DISCUSSION OF SIGNAL PROCESSING FOR ACTIVE SONARS. THE COURSE IS INTENDED PRIMARILY TO PRESENT BASIC PRINCIPLES, AND THE PORTIONS DEALING WITH HARDWARE AND PRACTICAL APPLICATIONS ARE SOMEWHAT SKETCHY.	UNCLASSIFIED

UNCLASSIFIED

U. S. NAVAL AIR DEVELOPMENT CENTER, JOHNSVILLE, WARMINSTER, PENNSYLVANIA
AIR WARFARE RESEARCH DEPARTMENT

LECTURE NOTES ON UNDERWATER ACOUSTICS; BY C. L. BARTBERGER; 17 MAY 1965; 415 P. REPORT: UNCLASSIFIED SECOND PRINTING 20 SEP 1965

THIS REPORT CONSISTS OF A SET OF LECTURE NOTES FOR A COURSE IN UNDERWATER ACOUSTICS, GIVEN AT THE U. S. NAVAL AIR DEVELOPMENT CENTER. THE COURSE IS BASED TO A CONSIDERABLE EXTENT UPON J. W. HORTON'S BOOK, "FUNDAMENTALS OF SONAR," AND COVERS

1. UNDERWATER ACOUSTICS
2. SONAR
3. PROPAGATION, ACOUSTIC
4. TRANSDUCERS, ACOUSTIC
5. ECHO-RANGING, ACOUSTIC

REPORT NADC-WR-6509

UNCLASSIFIED

U. S. NAVAL AIR DEVELOPMENT CENTER, JOHNSVILLE, WARMINSTER, PENNSYLVANIA
AIR WARFARE RESEARCH DEPARTMENT

LECTURE NOTES ON UNDERWATER ACOUSTICS; BY C. L. BARTBERGER; 17 MAY 1965; 415 P. REPORT: UNCLASSIFIED SECOND PRINTING 20 SEP 1965

THIS REPORT CONSISTS OF A SET OF LECTURE NOTES FOR A COURSE IN UNDERWATER ACOUSTICS, GIVEN AT THE U. S. NAVAL AIR DEVELOPMENT CENTER. THE COURSE IS BASED TO A CONSIDERABLE EXTENT UPON J. W. HORTON'S BOOK, "FUNDAMENTALS OF SONAR," AND COVERS

1. UNDERWATER ACOUSTICS
2. SONAR
3. PROPAGATION, ACOUSTIC
4. TRANSDUCERS, ACOUSTIC
5. ECHO-RANGING, ACOUSTIC

1. REPORT NADC-WR-6509

UNCLASSIFIED

U. S. NAVAL AIR DEVELOPMENT CENTER, JOHNSVILLE, WARMINSTER, PENNSYLVANIA
AIR WARFARE RESEARCH DEPARTMENT

LECTURE NOTES ON UNDERWATER ACOUSTICS; BY C. L. BARTBERGER; 17 MAY 1965; 415 P. REPORT: UNCLASSIFIED SECOND PRINTING 20 SEP 1965

THIS REPORT CONSISTS OF A SET OF LECTURE NOTES FOR A COURSE IN UNDERWATER ACOUSTICS, GIVEN AT THE U. S. NAVAL AIR DEVELOPMENT CENTER. THE COURSE IS BASED TO A CONSIDERABLE EXTENT UPON J. W. HORTON'S BOOK, "FUNDAMENTALS OF SONAR," AND COVERS

1. UNDERWATER ACOUSTICS
2. SONAR
3. PROPAGATION, ACOUSTIC
4. TRANSDUCERS, ACOUSTIC
5. ECHO-RANGING, ACOUSTIC

1. REPORT NADC-WR-6509

UNCLASSIFIED

U. S. NAVAL AIR DEVELOPMENT CENTER, JOHNSVILLE, WARMINSTER, PENNSYLVANIA
AIR WARFARE RESEARCH DEPARTMENT

LECTURE NOTES ON UNDERWATER ACOUSTICS; BY C. L. BARTBERGER; 17 MAY 1965; 415 P. REPORT: UNCLASSIFIED SECOND PRINTING 20 SEP 1965

THIS REPORT CONSISTS OF A SET OF LECTURE NOTES FOR A COURSE IN UNDERWATER ACOUSTICS, GIVEN AT THE U. S NAVAL AIR DEVELOPMENT CENTER. THE COURSE IS BASED TO A CONSIDERABLE EXTENT UPON J. W. HORTON'S BOOK, "FUNDAMENTALS OF SONAR," AND COVERS

1. UNDERWATER ACOUSTICS
2. SONAR
3. PROPAGATION, ACOUSTIC
4. TRANSDUCER, ACOUSTIC
5 ECHO-RANGING, ACOUSTIC

1. REPORT NADC-WR-6509

SUPPLEMENTARY

INFORMATION

(3) Infrared. Can detect wakes by measuring temperature difference between disturbed water and surrounding water. Has some capability against submerged submarines.

b. Acoustic. Good propagation through water. Has capability against submarines at all depths. Subject to limitations due to thermal conditions (which cause curving of sound rays).

c. Magnetic. Based on disturbance of earth's magnetic field caused by submarine. Good method for classification (distinguish submarines from extraneous objects). Limited to very short ranges.

d. Other methods. Other methods being investigated. None shows great promise.

While some non-acoustic methods are valuable as auxiliaries, primary method for submarine detection is acoustic.

3. Sonar

a. Historically, the term sonar was applied only to active systems which measured the azimuth and range of a target. In these notes the word will be used in its broadest sense, including not only both active and passive systems, but also sonobuoys and explosive echo ranging.

b. Signal Detection. Detection depends upon the relation between a signal and the interfering noise. A signal is defined as an item understood or recognized if we receive it. Whether any given set of sounds is a signal or not depends upon the circumstances. For example, the noise generated by a submarine is a signal if we are attempting to detect the submarine by passive listening; it is interfering noise if we are attempting to detect an echo returned by the submarine.

c. Passive Sonar. (Passive Listening). Signal depends upon the amount of noise generated by the target and upon the propagation loss involved in transmission of the sound from target to listening device. One-way transmission. Interference consists of ambient noise of the sea and self-noise of the listening platform.