STOCHASTIC MODELS OF THE SCATTERING OF SOUND BY BUBBLES IN THE UPPER OCEAN*

BY

PETER C. C. WANG AND HERMAN MEDWIN

Naval Postgraduate School, Monterey, Calif.

Abstract. Stochastic models are developed to relate the statistics of sound speed fluctuations and bubble density variations as a function of sound frequency in the upper ocean. These predictions from the stochastic model have been compared with ocean experimental data of sound speed modulation in the frequency range 15 to 150 kHz, and show satisfactory agreement. Future experiments and further modification of this model are discussed.

I. Introduction. It is well known that during sound propagation in the ocean, phase and amplitude fluctuations take place under the influence of temporal changes of motion, temperature, depth and salinity. Recently, Medwin [3, 5, 8] and his students have shown experimentally that the sound dispersion and phase fluctuations for frequencies up to approximately 100 kHz depend, to a considerable extent, on the bubbles present in the upper ocean.

Earlier an experiment investigating the scattering of sound by bubbles [4] generated by an artificial wind in sea water showed that the size distribution of bubbles remains invariant to changes in wind velocity at least for large wind velocities (greater than 8–10 m/sec). However, bubble density, regardless of its size distribution, does not remain constant for various wind velocities. The bubble density near the surface is dependent largely upon the wind velocity, wave height [8] and the depth.

Medwin [5] has investigated the dependence of bubble densities on depth in iso-thermal coastal waters, at sea states one and two, over twenty-four-hour periods. It is the purpose of this paper to develop stochastic models to explain some of the experimental findings; through the model, we will be able to obtain additional information regarding the dispersion behavior at other frequencies where experiments fail to obtain data.

The initial task here is to make a study of the dependence of speed fluctuations on sound frequencies and temperature fluctuations at various given bubble densities. Denote the speed of sound as $C(T, N, f)$, where T is the temperature, N is the bubble density and f is the sound frequency and where depth and salinity are assumed to be constant. We propose to compare the following pairs of random variables:

$$C(T_i, N_i \mid f_1), C(T_i, N_i \mid f_2), \qquad i, j = 1, 2, \cdots, n$$

where f_1, f_2 denote two selected frequencies from 15–150 KHz. Corrections to the values of $C(T_i, N_i \mid f)$ must be made due to changes of depth and salinity.

For fixed frequency, we have determined the probability density function (p.d.f.) of N from the knowledge of the p.d.f. of C and p.d.f. of T.

The variance of the speed fluctuations $\sigma_c^2(f)$ as a function of f and f_0 (the bubble resonant frequency) are analysed in Sec. II and the theoretical graph of this function is provided in Fig. 1. The asymptotic variance of the speed fluctuations ($f \leq f_0$) and some critical values of $\sigma_c^2(f)$ such as $\sigma_c^2(f_0/(1 \pm d)^{1/2})$ (where d is the bubble damping constant; see Figure 5) have also been evaluated and are presented in detail in Sec. II. Some statistical procedures and numerical examples are discussed in Appendix A. An asymptotic regression model is presented in Appendix B. The Kolmogorov–Smirnov two-sample test is considered in Appendix C.

This report includes part of our joint work developing statistical models in order to explain the role of bubble density in the dispersion of sound and in the sound speed fluctuations in the upper ocean.

II. Stochastic models on dispersions of sound speed fluctuations due to bubbles.

A. Summary. Our effort is directed to developing the relation between the distributions of the sound speed C determined from the experimental equation (i.e., $C = xf/(n + \varphi/2\pi)$ where φ is the variable sound phase measured in radians, f is the sound frequency and where x (the sound path length) and n (the number of wave lengths) are known quantities [3]) and the sound speed C_T determined from the Wilson formula [3]. These two expressions for sound speed in the bubbly water in the upper ocean are assumed to be related by the following functional form [9]:
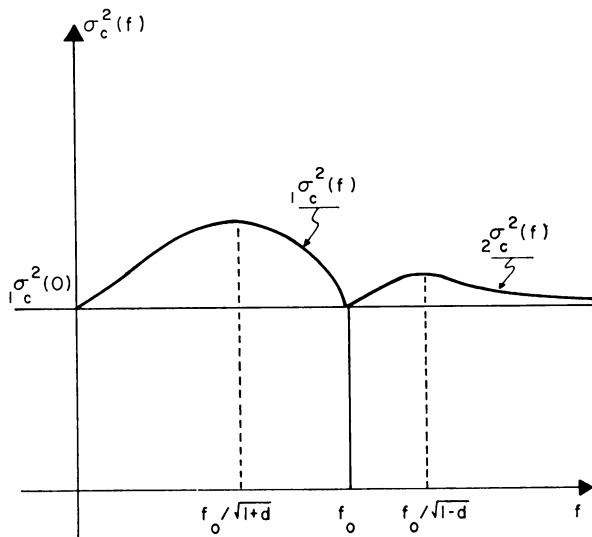
$$C = C_T[1 + a(f)V]^{-1/2} \qquad (1)$$

FIGURE 1

where

$$a(f) = \frac{3[(\omega_0/\omega)^2 - 1]}{R^2\omega_0^2\{[(\omega_0/\omega)^2 - 1]^2 + d^2(\omega_0^2/\omega^2)\}}$$ (1')

and $V \cong N$ which is the bubble density (see [9, p. 84 for details of formula (1)), d is the damping constant (see Fig. 5) and R is the radius of bubbles.

The distributions of C and C_T are derived under some mild restrictions in Secs. II(B) and II(C) respectively. The distribution of V is provided in II(C). The variance of C and asymptotic variance of C are discussed in II(D).

B. *The experimental determination of the variance of sound speed C.* It is given that the speed in the real ocean can be experimentally determined by $C = fx/(n + \varphi/2\pi)$. Assume that the variances of f and φ are known (actually var $f \equiv 0$). It is the purpose of this section to demonstrate that the variance of C can be determined. We further assume that the p.d.f. of φ is a truncated Gaussian, i.e.

$$h(\varphi) = \frac{1}{a\sqrt{2\pi}\,\sigma_\varphi} \exp\left(-(\varphi - \varphi_0)^2/2\sigma_\varphi^2\right) \quad \text{if} \quad 0 < \varphi \leq 2\pi,$$ (2)

$$= 0 \qquad\qquad\qquad\qquad\qquad \text{otherwise},$$

where a is the normalizing constant. The above hypothesis is confirmed as reasonable by the research findings in Rautmann's recent work [3]. The cumulative distribution function (c.d.f.) of the random variable C is denoted by $G(c)$ and we have

$$G(c) = \Pr\{C \leq c\}$$

$$= 0 \quad \text{if} \quad c < fx/(1 + n)$$

$$= 1 - \int_0^{2\pi(fx/c-n)} h(\varphi)\,d\varphi \quad \text{if} \quad fx/(n+1) \leq c < fx/n$$ (3)

$$= 1 \quad \text{if} \quad c \geq fx/n,$$

where $h(\varphi)$ is given in formula (2).

The p.d.f. of the random variable C is $g(c) = G'(c)$. Using the fact that C is a positive random variable and

$$E(C) = \int_{fx/(n+1)}^{fx/n} [1 - G(c)]\,dc, \qquad E(C^2) = \int_{fx/(n+1)}^{fx/n} 2c[1 - G(c)]\,dc,$$

the variance of C, $\sigma_c^2(f)$, can be expressed by the following equation:

$$\sigma_c^2(f) = \int_{fx/(1+n)}^{fx/n} \int_0^{2\pi(fx/c-n)} 2Ch(\varphi)\,d\varphi\,dc - \left[\int_{fx/(n+1)}^{fx/n} \int_0^{2\pi(fx/c-n)} h(\varphi)\,d\varphi\,dc\right]^2.$$

Note that both Eqs. (3) and (4) can be evaluated either by numerical procedures or by the table on Gaussian distributions. Let us denote by $H(\varphi)$ the c.d.f. of the random variable φ; then formula (4) becomes

$$\sigma_c^2(f) = \int_{fx/(1+n)}^{fx/n} 2CH\left[2\pi\left(\frac{fx}{c} - n\right)\right]dc$$

$$- \left\{\int_{fx/(1+n)}^{fx/n} H\left[2\pi\left(\frac{fx}{c} - n\right)\right]dc\right\}^2.$$ (4')

C. Determination of the distribution of C_T from Wilson's formula and V. It is assumed for simplicity that the quantities d (depth) and S (salinity) in Wilson's formula are constants and furthermore we replace the T^2 and T^3 terms with their corresponding mean values. Wilson's formula is then reduced to the linear form

$$C_T = a + bT \tag{5}$$

where $a = 1449 + .017d + 1.39(S - 35) - .055\hat{T}^2 + .0003\hat{T}^3$, $b = 4.6 - .012(S - 35)$, and $\hat{T}^2 = 1/m \sum_{i=1}^{m} T_i^2$ and $\hat{T}^3 = 1/m \sum_{i=1}^{m} T_i^3$; m is the sample size and the T_i's are the observed values of T. We further assume that $b > 0$ since S is very close to the value of 35.

This linearized version of Wilson's formula is useful because of the distributional properties of random variable C_T immediately obtainable. We assume that the random variable T is distributed according to a Gaussian distribution; i.e., the p.d.f. of the random variable T is

$$h(t) = \frac{1}{z\sqrt{2\pi}\,\sigma_T} \exp\left[-(t - t_0)^2/2\sigma_T^2\right] \quad \text{if} \quad 0 \leq t < \infty, \tag{5'}$$

$$= 0 \quad \text{if} \quad t < 0,$$

where z is the normalizing constant. The c.d.f. of the random variable C_T is

$$
\begin{aligned}
Q(C_t) &= \Pr\{C_T \leq C_t\} \\
&= 0 \quad \text{if} \quad C_t < a, \\
&= \int_0^{(C_t - a)/b} h(t)\,dt \quad \text{if} \quad a \leq C_t < \infty.
\end{aligned}
\tag{6}
$$

This implies that the p.d.f. of the random variable C_T is

$$q(C_t) = \frac{1}{b} h\left(\frac{C_t - a}{b}\right) \quad \text{if} \quad a \leq C_t < \infty, \tag{7}$$

$$= 0 \qquad\qquad \text{otherwise,}$$

The randomness of C_T can be obtained more accurately if T^2 and T^3 terms in Wilson's formula are included in the calculation, and the procedures used above can be extended immediately to include terms such as T^2 and T^3. The variance of $\sigma_{C_T}^2$ has the form $\sigma_{C_T}^2 = b^2 \sigma_T^2$ and σ_T^2 can be obtained from Eq. (5'). We assume for simplicity that the random variables C and C_T are stochastically independent random variables because C does not depend upon T and C_T does not assume bubble density. Hence the joint p.d.f. of random variables C and C_T can be written as $g(C, C_t) = g(c) \cdot q(C_t)$ if $a \leq C_t < \infty$ and $fx/(n + 1) \leq C \leq fx/n$. We can rearrange (1) and have

$$V = \frac{1}{a(f)} [C_T^2/C^2 - 1]. $$

Using the distributions of C and C_T derived in II(B), II(C) together with the condition that the random variables C and C_T are independent, the distribution function $H(v)$ of the random variable V can be derived in each of the following three cases depending on whether $a(f) > 0$, $a(f) < 0$ or $a(f) = 0$. These cases correspond, respectively, to sound frequencies ω below, above, and equal to the bubble resonant frequency ω_0.

Case 1. $a(f) > 0$ for $\omega < \omega_0$. The c.d.f. of the random variable V is

$$H_1(v) = 0 \quad \text{if} \quad v < 0$$

$$= \int_{fx/n/(1+a(f)v)^{1/2}}^{fx/n} \int_{fx/n}^{C(1+a(f)v)^{1/2}} g(C, C_t) \, dC_t \, dC$$

$$\text{if} \quad 0 \le v < \frac{1}{a(f)} [(n + 1/n)^2 - 1] \qquad (8)$$

$$\int_{fx/(n+1)}^{fx/n} \int_{fx/n}^{C(a(f)v+1)^{1/2}} g(C, C_t) \, dC_t \, dC \quad \text{if} \quad v \ge \frac{1}{a(f)} [(n + 1/n)^2 - 1].$$

Case 2. $a(f) < 0$ if $\omega > \omega_0$. The c.d.f. of the random variable V in this case can be expressed as

$$H_2(v) = 0 \qquad\qquad\qquad\qquad \text{if} \quad v < 0$$

$$= \int_{fx/(n+1)}^{fx/n} \int_{C[1+a(f)v]^{1/2}}^{\infty} g(C, C_t) \, dC_t \, dC \quad \text{if} \quad 0 < v < \frac{-1}{a(f)}$$

$$= 1 \qquad\qquad\qquad\qquad \text{if} \quad v > \frac{-1}{a(f)}.$$

Case 3. $a(f) = 0$ for $\omega = \omega_0$. The c.d.f. on V, denoted as $H_3(v)$, is degenerate along the line $C_T = C$ for all values of v, i.e.,

$$H_3(v) = 0 \quad \text{if} \quad v < \infty, \qquad\qquad (10)$$

$$= 1 \quad \text{if} \quad v = \infty.$$

The distribution function of V is governed by essentially two different functional forms. For $\omega < \omega_0$ (for frequencies less than bubble resonance frequency) the distribution of V is given in formula (8). For $\omega > \omega_0$, the distribution of V is given in formula (9).

  D. The behavior of the variance of sound-speed fluctuation and the asymptotic variance of sound speed at high frequencies. The behavior of the variance of C on the low-frequency portion (relative to the resonant frequency) is provided by formulae (4) and (4') in II(B). We must utilize information expressed in formula (8) to obtain $\sigma_c^2(f)$ for the high-frequency portion as well. To approach this problem, we make certain assumptions which are fairly reasonable from the experimental standpoint of view, namely:

  1. The random variable V is Gaussian,
  2. V and C_T are independent random variables.

The p.d.f. $k(v)$ of random variable V is Gaussian, i.e.,

$$k(v) = \frac{1}{w\sqrt{2\pi}\,\sigma_V} \exp \left[-(v - v_0)^2 / 2\sigma_V^2 \right], \qquad 0 \le v < \infty$$

$$= 0, \quad v < 0,$$

where w is the normalizing constant. The distribution function of the random variable C_T is given in formula (6). The distribution function of the random variable C can be calculated as follows:

Case 1. $a(f) > 0$ for $\omega < \omega_0$.

$$G_1(C) = 0 \qquad\qquad\qquad\qquad\qquad \text{if} \quad C \le 0 \qquad\qquad (11)$$

$$= \int_0^\infty \int_{C[1+a(f)v]^{1/2}}^\infty q(C_t)k(v)\, dC_t\, dv \quad \text{if} \quad C > 0.$$

Case 2. $a(f) < 0$ for $\omega > \omega_0$.

$$G_2(C) = 0 \qquad\qquad\qquad\qquad\qquad \text{if} \quad C \le 0 \qquad\qquad (12)$$

$$= \int_0^{-1/a(f)} \int_{C[1+a(f)v]^{1/2}}^\infty q(C_t)k(v)\, dC_t\, dv \quad \text{if} \quad C > 0$$

$$= 1 \qquad\qquad\qquad\qquad\qquad\qquad \text{if} \quad C = \infty$$

Note that here we have the condition $0 \le v \le -1/a(f)$. For $v > -1/a(f)$, we have

$$G_2(C) = Pn\{C \le c\}$$

$$= Pn\left\{ \frac{C_T{}^2}{[1 + a(f)v]} \le c \right\}$$

$$= 1.$$

Formula (11) may be rewritten as

$$G_1(C) = \int_0^\infty k(v)Q[C(1 + a(f)v)^{1/2}]\} \, dv, \qquad\qquad (13)$$

and formula (12) may be rewritten as

$$G_2(C) = \int_0^{-1/a(f)} k(v)\{1 - Q[C(1 + a(f)v)^{1/2}] \, dv. \qquad\qquad (14)$$

The variance of C can be computed from formula (13) for the case where $\omega < \omega_0$ and from formula (14) for the case where $\omega > \omega_0$. We then have for $\omega < \omega_0$,

$$_1\sigma_c{}^2(f) = \int_0^\infty \int_0^\infty 2Ck(v)Q[C(1 + a(f)v)^{1/2}] \, dv \, dc \qquad\qquad (15)$$

$$- \left[\int_0^\infty \int_0^\infty k(v)Q[C(1 + a(f)v)^{1/2}] \, dv \, dc \right]^2.$$

Similarly, for $\omega > \omega_0$ we have

$$_2\sigma_c{}^2(f) = \int_0^\infty \int_0^{-1/a(f)} 2Ck(v)\{1 - Q[C(1 + a(f)v)^{1/2}]\} \, dv \, dc \qquad\qquad (16)$$

$$- \left[\int_0^\infty \int_0^{-1/a(f)} k(v)\{1 - Q[C(1 + a(f)v)^{1/2}]\} \, dv \, dc \right]^2.$$

It is clear that we have the identities

$$_1\sigma_c{}^2(f_0{}^-) = {}_2\sigma_c{}^2(f_0{}^+) = \sigma_c{}^2 = {}_1\sigma_c{}^2(0). \qquad\qquad (17)$$

(see Fig. 1). To differentiate Eq. (15) with respect to f, we get

$$D_f \, _1\sigma_c{}^2(f) = \int_0^\infty \int_0^\infty 2k(v)(A - c) \, D_fQ \, dv \, dc$$

where

$$A = \int_0^\infty \int_0^\infty k(v)\{1 - Q[c(1 + a(f)v)^{1/2}]\}\, dv\, dc$$

and

$$D_f Q = \frac{cv}{2} \frac{q[c(1 + a(f)v)^{1/2}]}{[1 + a(f)v]^{1/2}} D_f a(f).$$

Letting $D_f a(f) = 0$, we have found that the critical value of $_1\sigma_c(f)^2$ is at the point $f_0/\sqrt{1 + d}$. Similarly, we can find from Eq. (16) that the critical value of the function $_2\sigma_c(f)$ is at the point $f_0/\sqrt{1 - d}$. The graph of $_1\sigma_c{}^2(f)$ and $_2\sigma_c{}^2(f)$ for all frequencies $f > 0$ is either an M-shaped curve or a W-shaped curve. Based on the physics of the problem, we conclude that this graph is an M-shaped curve which satisfies the following inequality:

$$_2\sigma_c{}^2(f_0/\sqrt{1 - d}) > \sigma_c{}^2(f_0).$$

This inequality will be established statistically in Eqs. (19–21).

In the remaining portion of this section we will demonstrate that the asymptotic variance, defined below, depends on ω_0 and the lower bound of this asymptotic variance is $\sigma_c{}^2(0)$. As $f \to \infty$, we have from (1') that $a(f) \to -3/(R^2\omega_0{}^2)$.

The asymptotic variance is denoted as

$$_2\sigma_c{}^2(\infty) = \lim_{f \to \infty} {}_2\sigma_c{}^2(f)$$

$$= \int_0^\infty \int_0^{R^2\omega_0{}^2/3} 2ck(v)\left\{1 - Q\left[c\left(1 - \frac{3v}{R^2\omega_0{}^2}\right)^{1/2}\right]\right\} dv\, dc$$

$$- \left[\int_0^\infty \int_0^{R^2\omega_0{}^2/3} k(v)\left\{1 - Q\left[c\left(1 - \frac{3v}{R^2\omega_0{}^2}\right)^{1/2}\right]\right\} dv\, dc\right]^2$$

$$= \int_0^{R^2\omega_0{}^2/3} \frac{k(v)\, dv}{1 - 3v/(R^2\omega_0{}^2)} \cdot \sigma_c{}^2(f)$$

$$+ E^2(C)\left[\int_0^{R^2\omega_0{}^2/3} \frac{k(v)\, dv}{1 - 3v/(R^2\omega_0{}^2)} - \left(\int_0^{R^2\omega_0{}^2/3} \frac{k(v)\, dv}{\left(1 - \frac{3v}{R^2\omega_0{}^2}\right)^{1/2}}\right)^2\right].$$

From the Cauchy–Schwarz inequality, we have

$$\int_0^{R^2\omega_0{}^2/3} \frac{k(v)\, dv}{1 - \frac{3v}{R^2\omega_0{}^2}} \geq \left[\int_0^{R^2\omega_0{}^2/3} \frac{k(v)\, dv}{\left(1 - \frac{3v}{R^2\omega_0{}^2}\right)^{1/2}}\right]^2.$$

Hence, we have the following inequality:

$$_2\sigma_c{}^2(\infty) \geq \sigma_c{}^2(f_0) \cdot \int_0^{R^2\omega_0{}^2/3} k(v)\, \frac{dv}{1 - \frac{3v}{R^2\omega_0{}^2}} \tag{18}$$

$$= \sigma_c{}^2(f_0) + \sigma_c{}^2(f_0)[aE(V) + a^2E(V^2) + \cdots + a^nE(V^n) + \cdots]$$

where $a = 3/R^2\omega_0{}^2$ and $E(V^n) = \int_0^\infty v^n k(v)\, dv$, the nth moment of V.

Since all moments are nonnegative, we conclude that

$$_2\sigma_c{}^2(\infty) \geq \sigma_c{}^2(f_0). \tag{19}$$

Furthermore, letting $\omega_0 \to \infty$, we have

$$\lim_{\omega_0 \to \infty} {}_2\sigma_c^2(\infty) \geq \lim_{\omega_0 \to \infty} \left\{ \sigma_c^2(f_0) + \sigma_c^2(f_0) \left[\frac{3}{R^2 \omega_0^2} E(V) + \frac{9}{R^4 \omega_0^4} E(V^2) + \cdots \right] \right\} = \sigma_c^2(f_0),$$

(20)

which is to be verified.

The behavior of formula (20) is described in the graph in Fig. 2 for two superposed curve with bubble frequencies f_0 and f_0', and this checks out with the experimental data collected in [8] for high frequencies.

It can be shown that the variance at the first peak $f = f_0/\sqrt{1+d}$ is greater than the variance at the second peak where $f = f_0/\sqrt{1-d}$. We have

$$_2\sigma_c^2(f_0/\sqrt{1-d}), \qquad {}_1\sigma_c^2(f_0/\sqrt{1+d}).$$

(21)

The comparison of these predictions with results of an actual experiment at sea is interesting. Figs. 3 and 4 are from [8]. We tabulate below the suggested resonant frequencies from this work, the predicted maxima of the variance following Eqs. (15), (16), and the observed maxima at sea.

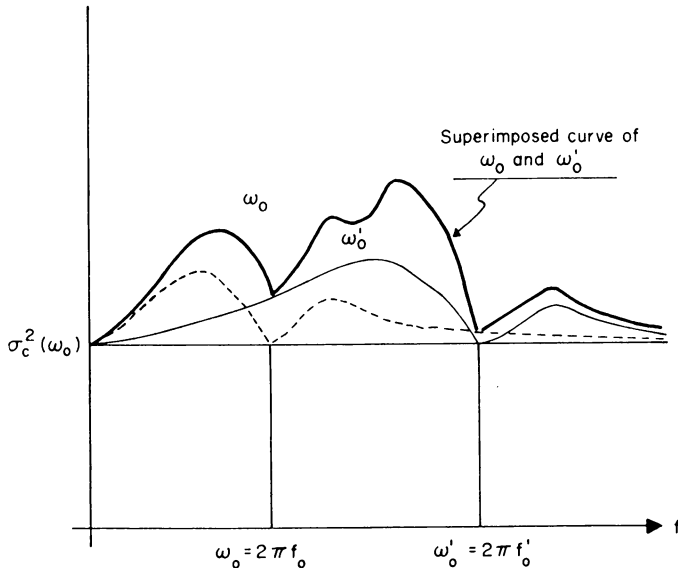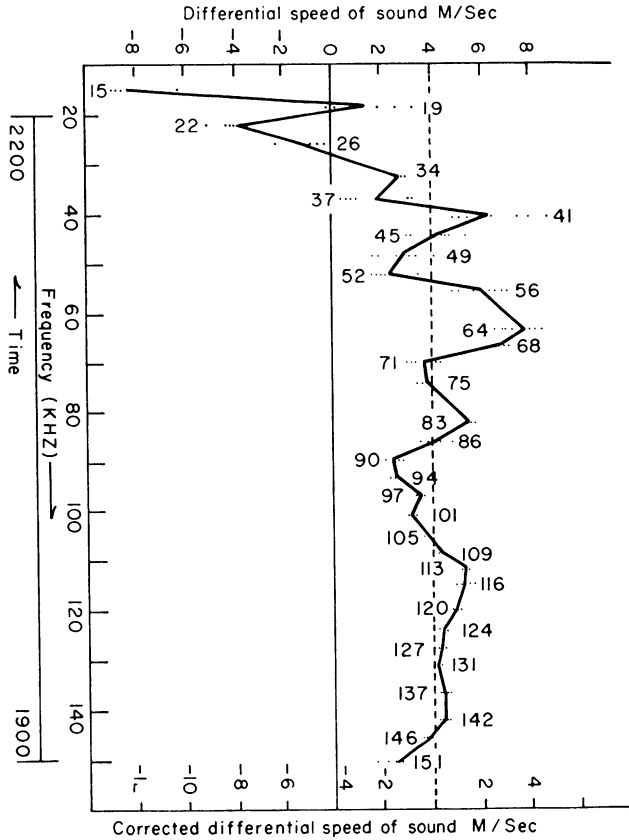| Damping constant d from Fig. 5	Resonance frequency in KHz from Fig. 3	Predicted maximum of σ_c^2 using Eqs. (15) and (16)		Nearest experimental maxima of σ_c^2 from Fig. 4
		Eq. (16)	(Eq. (15))	
.075	20	27	(29)	26*
.082	39	37	(41)	37*
.092	54	51	(57)	57–64 plateau*
.105	75	71	(79)	71*
.12	105	99	(112)	94, 109

FIGURE 2

FIGURE 3

The starred values represent what we believe to be substantial confirmation of the predictions of the stochastic model.

It is observed that it is for frequencies above 100 KHz that the bubble model is inadequate. In fact, the differential speeds at these high frequencies are rather small
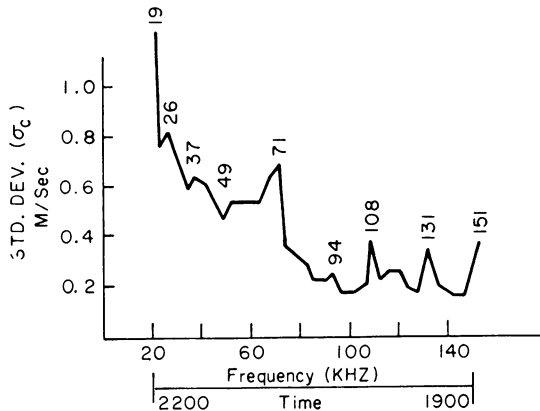

FIGURE 4

and the variance of the speed is more likely to be due to temperature fluctuations, according to other evidence presented in [8].

E. Side experiments. In the experiment investigating the scattering of sound by bubbles generated by wind in sea water under laboratory setup, several side experiments can be considered. Namely:

(i) The temperature effect on the fluctuation of sound speed can be tested when depth, salinity and bubble density are assumed to be kept at constant level.

(ii) Since bubble density and depth are inversely related, the exact graph of this relation can be determined from experiments.

(iii) We conjecture that any modifications to Wilson's formula to include the effect have to come from the depth term (i.e., $0.17d$ [3 p. 21]) if bubble density is incorporated into Wilson's formula. We propose specifically that the linear depth term is to be replaced by a nonlinear term in Wilson's formula; i.e., the linear depth term $(0.17d)$ is replaced by the following nonlinear term:

$$b_1(d + b_2)(N + b_3)$$

where b_1, b_2 and b_3 are parameters to be estimated while d is the depth and N is the bubble density. This side experiment should be carried out under conditions that the salinity of sea water used is kept at 35 and temperature of sea water used is kept constant at 50–60° F.

(iv) Since changes of depth (say due to waves overhead) cause changes in the resonant frequency of the bubbles, we propose to study the consequences this dependence in terms of the variance of the speed of sound as a function of wave height. Such effects have been postulated by Fitzgerald [8] in explanation of the large cross-correlation of sound phase with wave height near f_0.

Appendix A. Statistical procedures illustrated with numerical examples. The basic question concerning the experiment is to determine how various bubble-density fluctuations can change the sound-speed fluctuations at various temperature fluctuations and frequencies of sound input. Here we discuss statistical procedures for testing the difference of the two random variables C_{f_1} and C_{f_2}. Two statistical procedures for testing the difference of two empirical distributions are provided for illustrative purposes. Sample sizes required to attain certain confidences are also discussed. Two simple numerical examples are provided to illustrate how to apply these procedures to data collected in the experiment.

A. *Two-sample Kolmogorov–Smirnov test.* The two-sample Kolmogorov–Smirnov test can be used to test the difference of two empirical distributions given that they differ by no more than $p\%$ of the average sound speed.

Let n be the size of the sample and let $F_n(t)$ and $G_n(t)$ be the empirical distributions corresponding to the two random variables; that is:

$$F_n(c) = \frac{\text{number of } C_{f_1}\text{'s} \le c}{n} \, ;$$

$$G_n(c) = \frac{\text{number of } C_{f_2}\text{'s} \le c}{n} .$$

This test is based on the following statistic:

$$D_{nn} = \sup |F_n(c) - G_n(c)|.$$

The null hypothesis that the samples are from the same distribution is rejected if D_{nn} is sufficiently large. The exact distribution of the random variable nD_{nn} is given by Birnbaum and Hall [7]. A numerical example is provided in the following. Assume that observations $C_{f_1}(1)$, $C_{f_1}(2)$, $C_{f_1}(3)$, $C_{f_1}(4)$, $C_{f_1}(5)$ are made under system one and observations $C_{f_2}(1)$, $C_{f_2}(2)$, $C_{f_2}(3)$, $C_{f_2}(4)$, $C_{f_2}(5)$ are made under system two. These observations are assumed to be

$$(1400.5,\ 1401.2,\ 1401.4,\ 1402.3,\ 1403.0)$$

and

$$(1401.3,\ 1402.4,\ 1402.6,\ 1404.0,\ 1405.0).$$

The computation of $D_{5,5} = .6$ is shown in Table 1. In this case, $nD_{nn} = 5(.6) = 3.0$ It can be found that the shaded region has an exact area $A = .36$ under the null hypothesis. In this case, D_{nn} is not significant at any level less than 36% (say 10% level).

B. χ^2-*test of goodness of fit.* An useful alternative to the Kolmogorov–Smirnov test for testing the equality of two empirical distributions is the χ^2-test. Let C_{f_1} and C_{f_2} be stochastically independent random variables with empirical distributions $F_n(C)$ and $G_n(C)$ respectively, as defined in the last example. Suppose $[A_1, A_2, \cdots, A_K]$ is a partition of the set of possible values of C_{f_1} and C_{f_2} and let

$$P_i = P_r\{C_{f_1} \in A_i\}, \qquad i = 1, 2, \cdots, K;$$

$$q_i = P_r\{C_{f_2} \in A_i\}, \qquad j = 1, 2, \cdots, K.$$

TABLE I.

$C_{f_1}(i)$	$C_{f_2}(j)$	F_n	G_n	$|F_n - G_n|$
1400.5		.2	0	.2
1401.2		.4	0	.4
	1401.3	.4	.2	.2
1401.4		.6	.2	.4
1402.3		.8	.2	.6
	1402.4	.8	.4	.4
	1402.6	.8	.6	.2
1403.0		1.0	.6	.4
	1404.0	1.0	.8	.2
	1405.0	1.0	1.0	0

$$nD_{nn} = 3.0$$
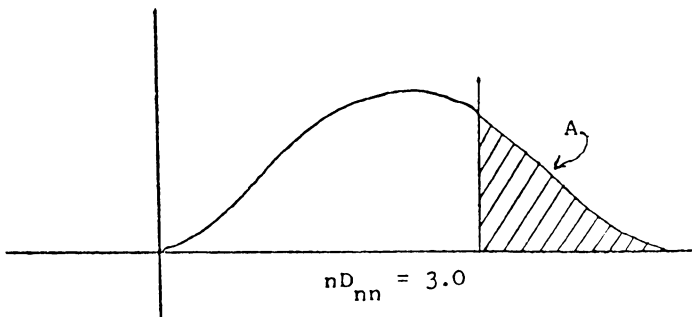
TABLE 2

Partitions	Observed Frequencies n_i			Expected Frequencies e_i			$\dfrac{(m_i - e_i)^2}{e_i}$	
Range	C_{f_1}	C_{f_2}		C_{f_1}	C_{f_2}		C_{f_1}	C_{f_2}
1401–1420	2	10	12	4.8*	7.2**	12	1.63Δ	1.09
1421–1440	6	21	27	10.8	16.2	27	2.13	1.42
1441–1460	8	15	23	9.2	13.8	23	.16	.10
1461–1480	13	7	20	8.0	12.0	20	3.12	2.08
1481–1500	11	7	18	7.2	10.8	18	2.45	1.63
	40	60	100	40	60	100	$\chi_f = 15.8$	

$$* \ 4.8 = 40\ \frac{12}{100} \qquad ** \ 7.2 = 60\ \frac{12}{100} \qquad \Delta \ 1.63 = \frac{(4.8 - 2)^2}{4.8}$$

If C_{f_1} and C_{f_2} have the same distribution, then $P_i = q_i$, $i = 1, 2, \cdots, K$. Accordingly, the hypothesis of interest is:

$$H_0 : P_i = q_i : i = 1, 2, \cdots, K.$$

This is the problem of testing the equality of two independent multinominal distributions. The partition $A_1, A_2, \cdots, A_K$ may be determined by the experimental values. The test statistic has an approximate $\chi^2(K - 1)$ distribution. Based on the above discussion, a two-sample χ^2-test of the hypothesis that two independent sets of independent observations are from the same distribution may be conducted by classifying the observations on C_{f_1} and C_{f_2} into sets of K cells, $A_1, A_2, \cdots, A_K$, and computing, as expected frequencies for each cell, the products of its marginal totals divided by the number of observations. χ^2 is computed from the resulting observed and expected frequencies. The approximate distribution of the χ^2 statistic is that of the χ^2—table with $(K - 1)$ degrees of freedom. As a general guideline, this χ^2 approximation should be used only when the sample size n is such that $n \geq \max_{1 \leq i \leq k} (5/p_i)$.

Example. Observations, already grouped, are given in Table 2, along with the table of expected frequencies and the computation of χ^2. From the χ^2 table with $v = 4$, the tail area is $A_{\chi}^2(H_0) = .003$. The null hypothesis that C_{f_1} and C_{f_2} are identically distributed is therefore rejected at the .10 level.

The sample size n required to attain α confidence that the two empirical distributions may differ by no more than 25% (arbitrarily picked) may be computed in the following manner. If $\alpha = 10\%$, then $1.22 = \sqrt{n/2}\, D_{nn}$. Since $D_{nn} = 25\%$, one may compute for the required sample size:

$$n \geq 2(16)(1.22)^2 \doteq 48.$$

In a similar manner, one obtains $n \doteq 60$ if $\alpha = 5\%$ and $n \doteq 85$ if $\alpha = 1\%$. The sample sizes corresponding to a few selected values of α are summarized in Table 3.

Remarks: The 25% difference of the two empirical distributions required is used only for illustrative purposes.

TABLE 3

α	10%	5%	1%
n	48	60	85

Appendix B. Asymptotic regression model. In this section, a curve-fitting technique is briefly discussed in the study of sound-speed fluctuations and bubble density.

Let us assume that the sound-speed fluctuation is determined solely by the bubble-density fluctuations. A nonlinear model is proposed for the above phenomenon, namely the equation:

$$C(N) = A + B\rho^N; 0 < \rho < 1,$$

where C is the speed of sound and N is the bubble density and A, B, ρ are parameters. The least-squares method for fitting the proposed regression curve above requires the estimates ρ, A and B by successive approximations. For equally spaced values of N, solutions of the appropriate normal equations can be simplified. Stevens [1] gave an ingenious method for estimating the parameter by maximum likelihood. The practical application of that method has been explored by Hiorns [2]. Most of the techniques discussed in the above references can immediately be used to our problem except certain modifications must be made in order to increase the efficiency of the estimators $\hat{\rho}$, $\hat{A}$ and $\hat{B}$, particularly for the case when $\hat{\rho}$ is small and the sample size is large.

This may not be necessary since no prior knowledge about ρ is known at this moment. A similar regression form has been used in Medwin [5] to describe the dependence of bubble density on depth and wind speed.

There are different ways of writing the regression curve that are essentially the same insofar as the parameters in one form can be expressed in terms of the parameters in the other forms. For example, the parameters in

$$C = a(1 - b\gamma^N),$$

$$C = p(1 - \exp[-q(N + s)])$$

are related as follows:

$$a = A, ab = -B, \gamma = \rho, \exp(-qs) = b, \ln \gamma = -q.$$

The proposed regression curve can be used not only where sound-speed fluctuations and bubble density are concerned but can also be applied to the following experiments:

a. Bubble density versus depth [5, Figure 5],

b. Speed of sound versus temperature with bubbles,

c. Speed of sound versus temperature without bubbles.

The procedures suggested above will be useful in reducing the gross standard error and mean errors at each temperature and in achieving better accuracy in the sound speed.

Appendix C. Kolmogorov–Smirnov two-sample test. In comparing random variables $C(T, N, f_1)$ and $C(T, N, f_2)$, the sound speed will be adjusted for the difference in temperature. Let us denote the adjusted random variables by $C_{f_2}(T, N, f_1)$ and $C_{f_1}(T, N, f_2)$. A non-parametric two-sample test should be used here to establish whether $C_{f_2}(T, N, f_1)$ and $C_{f_1}(T, N, f_2)$ are affected by the bubbles in a similar manner. Without any prior knowledge of these quantities, we propose adopting a two-sample Kolmogorov–Smirnov test in testing the differences between C_{f_1} and C_{f_2}.

A few words must be said about the sample size when one is dealing with observations of the random variables $C(T, N, f_i)$. The sample size and confidence can now be deter-
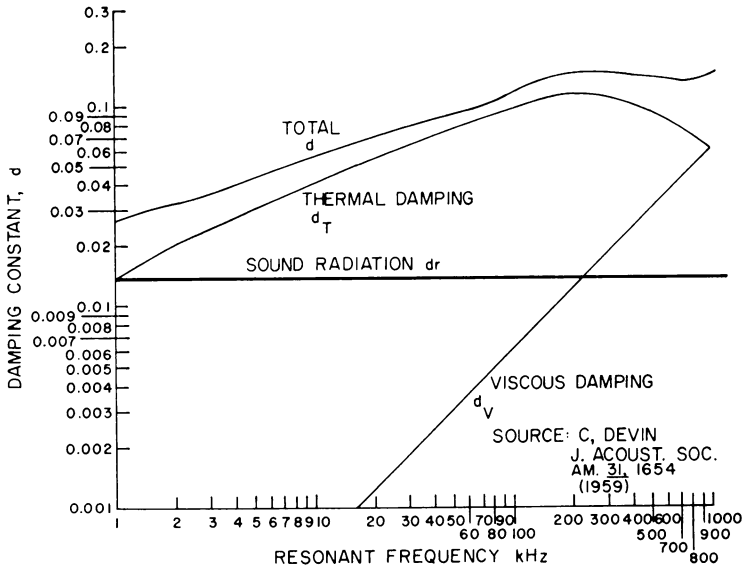
mined from the two-sample non-parametric test (see Table 3). In-situ measurement of the speed of sound in the upper ocean is not recommended for the experiment proposed above because of lack of control to the variable "bubble density". However, any in-situ measurements of the speed of sound in the upper ocean should include the measurements of bubble density at the time where the experiment is conducted, and these data can be used to provide a cross check with results in the laboratory. The above experiment can best be conducted under laboratory conditions by wind-agitated breaking waves because the bubble density can be controlled by wind speed [4]. A cross check among experiments conducted in the laboratory and in-situ measurements in the upper ocean should be made so that inferences about the speed fluctuations at an assumed level of bubble density in the upper ocean can be made. For details of the procedures, see Appendix A.

REFERENCES

[1] W. L. Stevens, *Asymptotic regression*, Biometrics **7**, 247–267 (1951)
[2] R. W. Hiorns, *The fitting of growth and allied curves of the asymptotic regression type by Steven's method*, Tracts for Computers, XXVIII, University College, London, Cambridge University Press (1965)
[3] Jurgen Rautmann, *Sound dispersion and phase fluctuations in the upper ocean*, Thesis, Naval Postgraduate School (1971)
[4] V. P. Glotov, P. A. Kolobaev and G. G. Neuimin, *Investigation of the scattering of sound by bubbles generated by an artificial wind in sea water and the statistical distribution of bubble sizes*, Soviet Physics —Acoustics **7**, 341–345 (1962)
[5] Herman Medwin, *In-situ acoustic measurements of bubble populations in coastal ocean waters*, J. Geophys. Research **75**, 599–611 (1970)
[6] Vincent A. Del Grosso, *Sound speed in pure water and sea water*, J. Acoust. Soc. Amer. **47**, 947–950 (1969)
[7] Z. W. Birnbaum and R. A. Hall, *Small sample distribution for multi-sample statistics of the Smirnov type*, Ann. Math. Stat. **31**, 710–720 (1960)

[8] James Fitzgerald, *Statistical study of sound speed in the inhomogeneous upper ocean*, Thesis in Engineering Acoustics, Naval Postgraduate School, December, 1972

[9] E. Skudrzyk Meyer, *Sound absorption and sound absorbers in water*, NAVSHIPS 900.164, **1**, 1 December 1950

[10] Peter C. C. Wang and Herman Medwin, *Statistical considerations to experiments on the scattering of sound by bubbles in the upper ocean*, Technical Report NPS-53WG72101A, Naval Postgraduate School, Monterey, California (1972)