STATUS REPORT: NASA RESEARCH GRANT SC-NGR-23-005-159

OPTIMAL LEARNING IN DETECTION SITUATIONS

I. INTRODUCTION

The present report, although not intended for publication, will nevertheless present theoretical results which have been obtained regarding the study of sensitization learning of human observers in acoustical tasks. These results, in particular, show that it is possible to obtain performance measures from human subjects which enable the experimenter to keep track, on a trial by trial basis, of the level of learning achieved by the subject. Before turning to the details of these results, however, we shall discuss the theory of statistical adaptation devices. Most of the basic results presented here regarding these devices is known in a small but rapidly growing literature (consisting mostly of technical reports) written by engineers and communication scientists. Their primary goal appears to be, ultimately, the construction of adaptive receivers for specialized detection tasks, such as reception of radar signals or human voice patterns. Since our problem as psychologists is to provide an adequate descriptive theory for human auditory learning, our organization of the material is somewhat novel, and, we hope, a contribution to our objective.

II. PRELIMINARIES

The nature of most psychoacoustical tasks permits the consideration of the problem of adaption from a slightly less general vantage point than might be required for other tasks. Coincidentally, it is for tasks of a similar nature that the theory of adaptive detection devices is most highly developed for reasons of mathematical tractability.

A block diagram of a typical psychoacoustical detection experiment appears in Figure 1. The experimenter, on trial j, chooses (possibly with the help of a random device) one signal alternative S_i from a finite set $\{S_i\}$, $i=0,1,2,\ldots,m$,

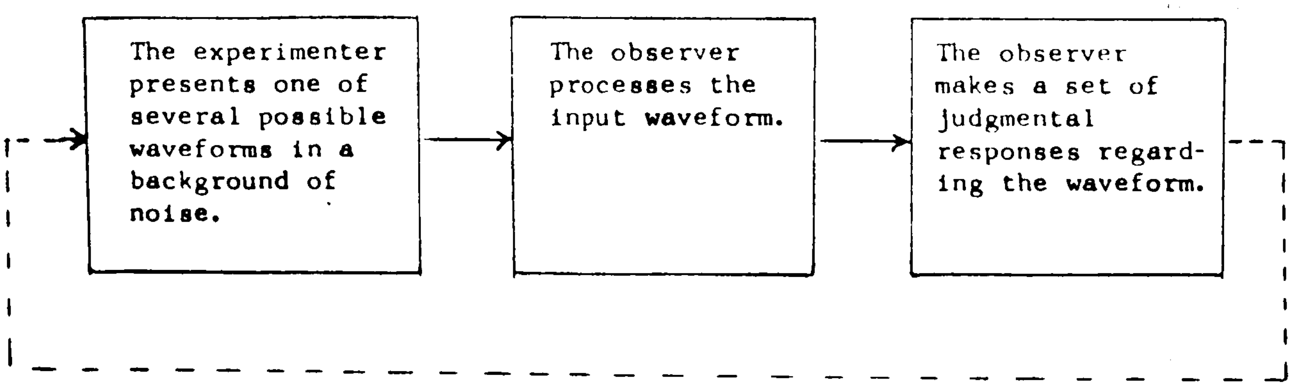

Figure 1. The typical detection experiment.

of alternatives. The (voltage or acoustical) waveform realization $s_i(t)$, $0 \leq t \leq T$, of alternative S_i superimposed on a sample of noise $n(t)$ (usually chosen from an infinite set of alternatives) is then presented to the observer for processing. After processing the input (the kinds of processing depend upon the observer's capability for various kinds of processing and his objective in processing, i.e., his "goal function"), the observer makes a set R_j of judgmental responses. Usually R_j is a subset of a well-defined predetermined set of possible responses. The experimenter may also give the observer feedback following his responses. The dotted arrow in Figure 1 indiactes that the selection of a signal alternative S_i on trial j may be at least partially observer-controlled in the sense that it depends on the sequence of previous responses $R_1, R_2, \ldots, R_{j-1}$.

We suppose that the random function $n=n(t)$ is finitely representable in the interval $0 \leq t \leq T$ and has a continuous distribution density $f(n)$. Each signal waveform $s_i(t)$, $i=1,2,\ldots,m$, also has a density (or probability mass function) denoted $g_i(s)$. The event S_0 is taken to represent "no signal" so that $s_o=s_o(t)=0(t)=0$ for all t. Thus, $g_0(s) = \delta(S - 0)$ and $Pr(s=0 \mid S_o)=1$, where δ is the impulse function*.

In a typical "Yes-No" experiment, the signal alternative set is $\{S_0, S_1\}$, and the possible responses on any trial are R_0 = "No—signal was not present" and R_1 = "Yes — signal was present." The input waveform may be described by a random function x defined by

(1)
$$x = \begin{cases} n & \text{if } S_0 \text{ is selected} \\ n + s & \text{if } S_1 \text{ is selected.} \end{cases}$$

The distribution density h of x given S_0 is $h(x|S_0) = f(x)$, while given S_1, the density of x is the convolution

$$h(x|S_1) = \int_{\Sigma} f(x-s)g(s)ds$$

where $g(s) = g_1(s)$ and Σ is the space of possible signal functions s. If, on any observation trial, the probability of selecting S_1 is p and the probability of S_0 is $q=1-p$, then the marginal density of x is the mixed density

(2)
$$\begin{aligned} h(x) &= ph(x|S_1) + qh(x|S_0) \\ &= p\int_{\Sigma} f(x-s)g(s)ds + qf(x). \end{aligned}$$

Birdsall (1963) has shown, under quite general conditions, that if a correct response (R_0 given S_0 or R_1 given S_1) is "preferable" to an incorrect response (R_0 given S_1 or R_1 given S_0), then the observer should base his decision on the value of the likelihood ratio $\ell(x) = h(x|S_1)/h(x|S_0)$ of the input

---

* $$\delta(x - x_o) = \begin{cases} \infty & \text{if } X = X_o \\ 0 & \text{if } X = 0 \end{cases} \quad \text{and} \quad \int_{X_o-\epsilon}^{X_o+\epsilon} \delta(X - X_o) = 1 \text{ for all } \epsilon > 0.$$

waveform sample x. That is, given a constant β which depends upon the observer's goal function, his decision rule should be

$$\text{If } \ell(x) \geq \beta \text{ ; make response } R_1;$$

(B) $$\text{If } \ell(x) < \beta \text{ ; make response } R_0.$$

As will be shown later, there are problems associated with the empirical implementation of this decision rule in a learning task for the observer (human or otherwise).

Using the previous definitions of the densities involved, we may write the likelihood ratio in the form

(3) $$\ell(x) = \frac{h(x \mid S_1)}{h(x \mid S_0)} = \frac{\int_\Sigma f(x-s)g(s)ds}{f(x)}.$$

The denominator is independent of the variable of integration so that

(3') $$\ell(x) = \int_\Sigma \ell(x \mid s)g(s)ds$$

where $\ell(x \mid s)$, the conditional likelihood ratio given s, has been introduced by letting

(4) $$\ell(x \mid s) = \frac{f(x - s)}{f(x)} .$$

In the case where there is only one waveform $s(t) = \bar{s}(t)$ = a (possibly different) constant for each value of t, $0 \leq t \leq T$, the distribution density g is the impulse function, i.e., $g(s) = \delta(s(t) - \bar{s}(t))$ which has unit mass concentrated at the single function $\bar{s} = \bar{s}(t)$. From (3') and (4) the likelihood ratio becomes

(5) $$\begin{aligned} \ell(x) &= \int_\Sigma (x \mid s)\ \delta(s - \bar{s})\ ds \\ &= \ell(x \mid s) \\ &= \frac{f(x - s)}{f(x)} . \end{aligned}$$

In (5) $\bar{s}$ represents the constant function $\bar{s}(t)$, whereas x is a random function with values x(t) in the observation interval, $0 \leq t \leq T$.

A device may compute $\ell(x)$ from (3') when the signal is known statistically (SKS) or from (5) when the signal is known exactly (SKE). The performance will be optimal on a wide selection of goal functions using decision rule B. Such a (non-empirically realizable) device is called the "ideal observer" (Tanner and Birdsall, 1958). Any device which computes a function $\ell^*(x)$, strictly monotone with $\ell(x)$, can also perform optimally by using decision rule B. Therefore, we shall also call a device which computes $\ell^*(x)$ an ideal observer. The one-dimensional set of numbers constituting the range of an $\ell^*(x)$ will be called an <u>ideal decision axis.</u>

Any device which uses a decision rule of the same form as B will also have a one-dimensional decision axis defined by the range of the decision function $\ell'(x)$ computed by the device. Notice that a decision axis will still be defined if ℓ' is a random function, and β', the cutoff value on the ℓ' decision axis, is a random variable. Thus the range of any (sub-optimum and possibly random) decision function in the sense of decision rule B will be called an <u>observer's decision axis.</u> Later we shall be concerned with the statistical relation between an ideal and the observer's decision axes for particular sub-optimum devices.

We turn now to a discussion of degradations in the prior knowledge available to the observer. In order to limit the present discussion somewhat, we assume throughout the remainder of the report that the noise density f(x) is known to all observers under consideration. The distribution density for signal g(s) used in the derivation of (3') may be called the <u>environmental</u> density of the signal in order to distinguish it from some density g'(s) which characterizes an observer's <u>prior opinion</u> regarding the distribution of the signal. The likelihood ratio $\ell'(x)$ obtained with the prior density $g'(s)$, using (3'), is

$$\ell'(x) = \int_{\Sigma} \ell(x|s)g'(s)ds \tag{6}$$

where it is clear that $\ell(x|s)$ is unchanged from its definition in (4) because we have assumed that $f'(x) = f(x)$ for all observers.

The interpretation of $g'(s)$ as the prior density for signal is not the only possible interpretation. Birdsall (1960), under the conditions that the environmental density $g(s) = \delta(s-\bar{s})$, and that the mean of $g'(s)$ is $\bar{s}$, assumed that $g'(s)$ arose as a result of a faulty memory on the part of the observer. With the further assumption that the observer knows his memory distribution, Birdsall arrived at (6) for the observer's likelihood ratio. We note that if that observer did not know his memory distribution, but rather believed a sample s' from his memory to be the true signal function s, then his likelihood ratio function would be described as in (5):

$$\ell(x|s') = \frac{f(x-s')}{f(x)} \tag{7}$$

Now if s' has a distribution $g'(s')$, as might be the case for an observer with faulty memory of which he is unaware, the expected likelihood ratio computed by the observer would be

$$\mathcal{E}_{g'}[\ell(x|s')] = \int_{\Sigma} \ell(x|s')g'(s')ds'. \tag{8}$$

By using decision rule B with the random decision function $\ell(x|s')$, it is the range of the expected likelihood ratio in (8) rather than that of (7) which defines the observer's decision axis.

The preceding discussion may be summarized by giving two definitions: Regardless of the environmental signal density $g(s)$, (i), if an observer computes the likelihood ratio $\ell'(x)$ according to (6) by using a prior distribution $g'(s)$, we say that the signal is specified statistically (SSS), and (ii), if an observer computes the conditional likelihood ratio of (7) by using a prior density $g'(s) = \delta(s - s')$, then we say the signal is specified exactly (SSE). (The terminology used herein is an expansion of the terminology used by Tanner and Birdsall, 1958). The ideal observer has $g'(s) = g(s)$, so that for it, SSS is equivalent to SKS and SSE is equivalent to SKE. Descriptive models of the human observer in detection tasks have been constructed by using the assumption of SSS or SSE.

III. ON THE THEORY OF ADAPTIVE DEVICES FOR DETECTION OF UNCERTAIN WAVEFORM PATTERNS IN NOISE.

A discrepancy in performance between a sub-optimum observer and the ideal observer may depend upon the difference in the prior signal densities $g'(s)$ and $g(s)$. A Bayesian learning device with prior density $g_j(s)$ on trial j would attempt to improve its knowledge of the environmental density after each new observation trial. More precisely, given a sample waveform x_j on trial j the posterior density $g_{j+1}(s)$ is given by

$$g_{j+1}(s) = \frac{h(x_j|s)g_j(s)}{\int_{\Sigma} h(x_j|s)g_j(s)ds} \tag{9}$$

where the density $h(x|s)$ is to be identified. From (1) and the comments thereafter, we infer that $h(x|s,S_0) = f(x)$ and that $h(x|s,S_1) = f(x-s)$. Therefore, we may write

$$(10) \qquad h(x|s) = ph(x|s,S_1) + qh(x|s,S_o) = pf(x-s) + qf(x).$$

Further, we see that the unconditional density of x on trial j must be

$$(11) \qquad h_j(x) = \int_{\Sigma} h(x|s)g_j(s)ds = \int_{\Sigma} [pf(x-s) + qf(x)]g_j(s)ds = p\int_{\Sigma} f(x-s)g_j(s)ds + qf(x)$$

where the final line is obtained from the fact that f(x) is independent of s and the assumption that $\int_{\Sigma} g_j(s)ds = 1$.

Corresponding to (6) we define

$$(12) \qquad \ell_j(x) = \int_{\Sigma} \ell(x|s)g_j(s)ds.$$

Now, by using (10), (11), and (12), the posterior distribution density of (9) may be written as

$$(13) \qquad g_{j+1}(s) = \frac{h(x_j|s)g_j(s)}{h_j(x_j)} = \left[\frac{pf(x_j-s) + qf(x_j)}{p\int_{\Sigma} f(x_j-s)g_j(s)ds + qf(x_j)}\right] g_j(s) = \left[\frac{\ell(x_j|s) + \alpha}{\ell_j(x_j) + \alpha}\right] g_j(s),$$

where $\alpha = q/p$. Equation (13) was obtained by Fralick (1965) and generalized to the case of M signal alternatives by Hancock and Patrick (1965). The importance of (13) lies in the fact that if the signal space Σ is finite, then the posterior distribution over Σ may be updated by this iterative procedure, trial by trial, with a constant finite number of operations performed on the input waveform.

Another approach to finding a finite solution to (9) is to approximate the density $h(x|s)$ given in (10) by a density $h^*(x|s)$ which admits sufficient statistics. An approximating density $h^*(x|s)$ for which sufficient statistics exist, usually will also have a "natural" conjugate distribution density which may be used as the prior density $g'(s)$ (cf. Raiffa and Schlaifer, 1961). When the prior distribution is the natural conjugate of $h^*(x|s)$, the parameters of the prior may be combined with the sufficient statistics of $h^*(x|s)$ to yield the parameters of the posterior density $g''(s)$. In this case, $g''(s)$ belongs to the family of densities to which $g'(s)$ belongs. If the sufficient statistics of $h^*(x|s)$ are of fixed dimensionality and $g'(s)$ has a finite number of parameters, then the application of Bayes' rule in giving

$$g''(s) = \frac{h^*(x|s)g'(s)}{\int_{\Sigma} h^*(x|s)g'(s)ds} \tag{14}$$

will require a fixed finite set of operations on an input vector x for any trial j.

To illustrate these ideas, let us suppose that $f(x)$ is normal with mean vector μ_f and covariance matrix Σ_f, i.e., $f(x) = f_N(x|\mu_f, \Sigma_f)$. Then it is easy to show that $f(x-s) = f_N(x|\mu_f - s, \Sigma_f)$. Further $h(x|s)$ has mean $\mu_h = \mu_f + ps$ and covariance matrix $\Sigma_h = \Sigma_f + pqss^t$, where s^t is the transpose of the vector s. Here we could let the density $h^*(x|s) = f_N(x|\mu_h, \Sigma_h)$ be the approximation to the density $h(x|s)$. It appears difficult to find the natural conjugate density to $h^*(x|s)$ because of the occurrence of the ss^t term in Σ_h. However, if only the mean of $h^*(x|s)$ depended upon the unknown s, a natural conjugate would be the normal distribution $g'(s) = f_N(s|\mu_{g'}, \Sigma_{g'})$. If we modify $h^*(x|s)$ to use the expectation of ss^t under the prior $g'(s)$ instead of ss^t itself, then $g'(s)$ could serve as the natural conjugate of the modified $h^*(x|s)$. Implementing this idea, we find

$$\mathcal{E}_{g'}(ss^t) = \int_{\Sigma} (ss^t) f_N(s|\mu_{g'}, \Sigma_{g'})ds$$
$$= \Sigma_{g'} + \mu_{g'}\mu_{g'}^t,$$

so that $h^*(x|s)$ modified becomes $f_N(x|\mu_h, \Sigma_{h^*})$, where

$$\Sigma_{h^*} = \Sigma_f + \Sigma_{g'} + \mu_{g'}\mu_{g'}^t$$

upon substitution for ss^t by its expectation. Thus (14) becomes

$$(15) \qquad g''(s) = \frac{f_N(x|\mu_h, \Sigma_{h^*})f_N(s|\mu_{g'}, \Sigma_{g'})}{\int_{\Sigma} f_N(x|\mu_h, \Sigma_{h^*})\, f_N(s|\mu_{g'}, \Sigma_{g'})ds}$$

It is shown in Raiffa and Schlaifer (1961), with suitable normalization of the parameters of $g'(s)$, that $g''(s) = f_N(s|\mu_{g''}, \Sigma_{g''})$ is also normal with mean vector

$$(16) \qquad \mu_{g''} = (\Sigma_{g'}^{-1} + \Sigma_{h^*}^{-1})^{-1} (\Sigma_{g'}^{-1}\mu_{g'} + \Sigma_{h^*}^{-1}\mu_h)$$

and covariance matrix

$$(17) \qquad \Sigma_{g''} = (\Sigma_{g'}^{-1} + \Sigma_{h^*}^{-1})^{-1}$$

so that $g''(s)$ and $g'(s)$ are in the same family as asserted.

It is not claimed that the particular approximation of $h(x|s)$ given by $h^*(x|s)$ above is a good one. However, in the special case where the environmental density $g(s) = \delta(s-\bar{s})$ and the distance $\| \mu_t - \bar{s} \|$ is small, a suboptimum device using the adaptation procedure given by (15) may converge to the ideal observer with SKE.

A more detailed analysis of approximations to $h(x|s)$ and an evaluation of the performance of the resulting suboptimum Bayesian adaptation devices are contemplated for investigation as a part of our remaining research grant period.

IV. PERFORMANCE MEASURES FOR LEARNING DEVICES

We have been considering optimum and suboptimum detection devices with a fixed conception of the environmental signal density which is represented by the prior density $g'(s)$. A performance discrepancy between a suboptimum observer and the ideal observer can be measured basically in two ways. The first is in terms of the expected loss for not performing optimally; this loss is determined by the payoffs prescribed by a goal function and the operating characteristics of the device. A special case in which higher than minimum-risk is achieved by a detection device is that in which the device knows $g(s)$, so $\ell(x)$ is computed as for the ideal observer, but the cutoff value β in using decision rule B is not chosen optimally. A non-optimal β could result from either incorrect assessment of the goal function or from not knowing the environmental value p of the probability of signal occurrence. Shuford (1964) derived the optimal Bayesian learning device for estimating p when $f(x)$ is the binomial probability mass function and the observer has a Beta prior distribution on p. We have shown that when the experimenter tells the observer after each observation trial whether noise alone or signal-plus-noise occurred, the sample value x_j

observed during trial j is irrelevant to the improvement of knowledge about p. This result was implicit in Shuford's work. Fralick (1965) has considered the optimal adaptive device for estimating p when f(x) is normal and p has one of a finite number of possible values.

Although a measure of departure from minimum risk may be obtained for any suboptimum device, it may not be a particularly sensitive measure for comparative purposes (cf. Green, 1960). A departure from minimum risk may be caused either by using a suboptimal likelihood ratio $\ell'(x)$ or a suboptimal cutoff value, or both.

The second class of ways in which a performance discrepancy may be measured is cutoff-free in the sense that the index of performance does not depend upon the particular β used in decision rule B.

We define Pr(C) to be the probability that a sample ℓ_{S_1} of likelihood ratio when S_1 occurs, drawn independently of a sample ℓ_{S_0} of likelihood ratio when S_0 occurs, will be the greater of the two; i.e.,

$$\Pr(C) = \Pr\left\{\ell_{S_1} = \max(\ell_{S_1}, \ell_{S_0})\right\}.$$

This probability is a measure of sensitivity of the detection device because Pr(C) measures the effectiveness of the decision variable $\ell(x)$ in discriminating between the two hypotheses S_1 and S_0. To compute Pr(C), we let $L(x) = \left\{x \mid \ell(x) \leq \ell\right\}$ so that the density $k(\ell \mid S_1)$ of ℓ given S_1 may be found by putting

$$k(\ell \mid S_1)d\ell = dK(\ell \mid S_1) = d\int_{L(x)} h(x \mid S_1)dx. \tag{18}$$

The distribution $K(\ell \mid S_1)$ is the integral

$$(19) \qquad K(\ell \mid S_1) = \int_{-\infty}^{\ell} k(\beta \mid S_1)\, d\beta \ .$$

By using (18) and (19), Pr(C) is given by

$$(20) \qquad Pr(C) = \int_{-\infty}^{\infty} K(\ell \mid S_0) k(\ell \mid S_1) d\ell \ .$$

It is shown by Swets and Green (1966) that Pr(C) is the area under the Receiver Operating Characteristic (ROC) curve for a device which computes $\ell(x)$.

When both densities $k(\ell|S_i)$, $i=0, 1$, are Gaussian (20) may be simplified to the expression

$$(21) \qquad Pr(C) = \int_{-\infty}^{d'/\sqrt{2}} f_N(\ell) d\ell = \Phi(d'/\sqrt{2}).$$

where d' is the (normalized) mean of ℓ_{S_1}. Thus, we may interpret d' as a measure of sensitivity, independent of the decision cutoff β, of an optimum detection device when SKE and the noise density is Gaussian. Since d' may be found for any device for which Pr(C) is known (or can be estimated), it is a conanical measure of sensitivity through which detection may be compared <u>regardless</u> of the distributions of x given S_0 and S_1.

It has been customary in the psychological literature to define the efficiency η of a device α with respect to the ideal observer as

$$(22) \qquad \eta = (d'_{\alpha} / d'_{opt.})^2$$

(See Tanner and Birdsall, 1958). We now give an interpretation of η :

<u>Theorem.</u> If f(x) is Gaussian and both a detection device α and the ideal observer have SSE, then the efficiency of α is the square of the correlation coefficient between its decision axis and that of the ideal observer.

The development leading to the proof of the theorem will be made somewhat more general by considering two arbitrary devices α and β, and then specializing one of them to be the ideal observer. When both signal and noise functions are Fourier series band limited in the same way over the interval $[0, T]$ the sample functions x and s may be represented as the vectors $x = (x_1, x_2, \ldots, x_u)$ and $s = (s_1, s_2, \ldots, s_u)$, respectively, where u = 2WT equally spaced components, and W is the bandwidth of the series (Peterson, Birdsall, and Fox, 1954). Since the noise is Gaussian (and let us assume white, without loss of generality) and SSE, i.e., $g_\alpha(s) = \delta(s - s_\alpha)$, $g_\beta(s) = \delta(s - s_\beta)$, we may write

(23)
$$h(x \mid S_0) = \left(\frac{1}{2\pi N}\right)^{u/2} \exp\left[-\frac{1}{2N} \sum_{i=1}^{u} x_i^2\right]$$

$$h_\gamma(x \mid S_1) = \left(\frac{1}{2\pi N}\right)^{u/2} \exp\left[-\frac{1}{2N} \sum_{i=1}^{u} (x_i - s_{\gamma_i})^2\right.$$

where N = noise power = $\frac{1}{2W} \mathcal{E}_n\left(\sum_{i=1}^{u} n_i^2\right)$ and $\gamma = \alpha, \beta$. The observers' decision axes may be taken as the logarithm of likelihood ratios $\ell_\gamma(x)$. The decision variables may be written by using (7) and (23) as

(24)
$$\ell_\gamma^* = \ln \ell_\gamma = \ln \frac{h_\gamma(x \mid S_1)}{h(x \mid S_0)} = \frac{1}{N} \sum x_i s_{\gamma_i} - \frac{1}{2N} \sum s_{\gamma_i}^2$$

$$= \frac{1}{N} \sum_{i=1}^{u} x_i s_{\gamma_i} - \frac{E_\gamma}{N_0},$$

where $E_\gamma = \int_0^T [s_\gamma(t)]^2 dt = (1/2W) \sum_{i=1}^{u} s^2_{\gamma i}$ is the effective energy of the specified signal s_γ, and $N_0 = N/W$ is the noise power per unit bandwidth.

When noise alone is present, $x_i = n_i$ and the mean of ℓ^*_γ is

$$(25) \qquad \mu_{\gamma 0} = \mathcal{E}_{S_0}(\ell^*_\gamma) = \frac{1}{N} \sum_{i=1}^{u} \mathcal{E}(n_i) s_{\gamma i} - \frac{E_\gamma}{N_0}$$

$$= -\frac{E_\gamma}{N_0}$$

and the variance of ℓ^*_γ is given by

$$(26) \qquad \sigma^2_{\gamma 0} = \mathcal{E}_{S_0}[(\ell^*_\gamma)^2] - [\mathcal{E}_{S_0}(\ell^*_\gamma)]^2$$

$$= 2E_\gamma/N_0.$$

When signal plus noise is present $x_i = n_i + \bar{s}_i$, and the mean of ℓ^*_γ is

$$(27) \qquad \mu_{\gamma 1} = \mathcal{E}_{S_1}(\ell^*_\gamma) = \frac{1}{N} \sum_{i=1}^{u} \mathcal{E}(n_i + \bar{s}_i) s_{\gamma i} - \frac{E_\gamma}{N_0}$$

$$= \frac{1}{N} \sum_{i=1}^{u} \bar{s}_i s_{\gamma i} - \frac{E_\gamma}{N_0}$$

$$= \frac{2R_\gamma}{N_0} - \frac{E_\gamma}{N_0},$$

where

$$(28) \qquad R_\gamma = \int_0^T \bar{s}(t) s_\gamma(t) dt = \frac{1}{2W} \sum_{i=1}^{u} \bar{s}_i s_{\gamma i}$$

has been introduced. The variance given S_1 may be shown by similar calculation to be the same as the variance given S_0:

(29) $$\sigma^2_{\gamma_0} = \sigma^2_{\gamma_1} = 2E_\gamma / N_0.$$

Intuitively, we expect (29) because of SKE; therefore $\bar{s}$ contributes no variability to x.

The sensitivity of a device γ when SKE is defined by

$$d'_\gamma = \frac{\mu_{\gamma_1} - \mu_{\gamma_0}}{\sigma_{\gamma_0}}$$

so that for SSE

(30) $$(d'_\gamma)^2 = \left[\frac{2R_\gamma}{N_0} - \frac{E_\gamma}{N_0} - \left(-\frac{E_\gamma}{N_0}\right)\right]^2 \Big/ \left(\frac{2E_\gamma}{N_0}\right)$$

$$= 2R^2_\gamma / E_\gamma N_0.$$

For γ the ideal observer $s_\gamma = \bar{s}$ so that (30) becomes

(31) $$(d'_{opt})^2 = 2E^2/EN_0 = 2E/N_0$$

where the <u>known</u> signal energy

(32) $$E = \int_0^T [\bar{s}(t)]^2 dt = \frac{1}{2W} \sum_{i=1}^{u} \bar{s}_i^{\,2}.$$

By using (22), (30) and (31), we find that when SKE and a device α has SSE the efficiency of α is

(33) $$\eta_\alpha = \left(\frac{d'_\alpha}{d'_{opt}}\right)^2 = \frac{2R_\alpha^2}{E_\alpha N_0} \Big/ \frac{2E}{N_0} = \frac{R_\alpha^2}{E_\alpha E}$$

To complete the proof of the theorem, we compute the covariance between ℓ_α^* and ℓ_β^* (which will be the same for S_0 and S_1, considering the reasoning following (29)). We have

(34) $$\begin{aligned} \sigma_{\alpha\beta} &= \mathcal{E}_{S_0}(\ell_\alpha^* \cdot \ell_\beta^*) - \mu_{\alpha_0}\mu_{\beta_0} \\ &= \left[\frac{1}{N^2}\sum_{i=1}^{u}\sum_{j=1}^{u} s_{\alpha_i} s_{\beta_j}\, \mathcal{E}(n_i n_j) + \frac{E_\alpha E_\beta}{N_0}\right] - \frac{E_\alpha E_\beta}{N_0^2} \\ &= \frac{1}{N}\sum_{i=1}^{u} s_{\alpha_i} s_{\beta_i} \\ &= \frac{2R_{\alpha\beta}}{N_0}. \end{aligned}$$

where

(35) $$R_{\alpha\beta} = \int_0^T s_\alpha(t) s_\beta(t)\,dt = \frac{1}{2W}\sum_{i=1}^{u} s_{\alpha_i} s_{\beta_i}.$$

Then by using (26) and (34) the square of the correlation $r_{\alpha\beta}$ between the decision axes of α and β is

(36) $$\begin{aligned} (r_{\alpha\beta})^2 &= \frac{\sigma_{\alpha\beta}^2}{\sigma_\alpha^2\,\sigma_\beta^2} = \left(\frac{2R_{\alpha\beta}}{N_0}\right)^2 \Big/ \left(\frac{2E_\alpha}{N_0}\right)\left(\frac{2E_\beta}{N_0}\right) \\ &= \frac{R_{\alpha\beta}^2}{E_\alpha E_\beta} \end{aligned}$$

When β is the ideal observer $R_{\alpha\beta} = R_\alpha$ and $E_\beta = E$, so

$$(r_{\alpha\ opt})^2 = \frac{R_\alpha^2}{E_\alpha E} = \eta_\alpha \tag{37}$$

by referring to (36) and (33).

V. SUMMARY OF COMPLETED AND PROPOSED RESEARCH OBJECTIVES.

1. Adaptation Using Bayes' Rule.

The form of $h(x|s)$ given in (13) enables an observer to adapt sequentially to a signal of a fixed, but unspecified, waveform, when approximations to $h(x|s)$ are given. As indicated above, we shall continue to investigate approximating densities to $h(x|s)$ and evaluate their applicability to the problem of human pattern discrimination.

2. Trial-by-trial Estimates of Efficiency.

The definition of Pr(C) given in IV allows direct estimates of an observer's efficiency via equations (21) and (22). If trial-by-trial estimates of the likelihood are given by the observer, $Pr(C) = Pr(\ell_{S_1} = \max(\ell_{S_1}, \ell_{S_0}))$ can be estimated directly and used to estimate the observer's efficiency by a trend analysis. Further, when the observer can be considered as performing a linear operation on the input waveform, a rank order correlation coefficient can be computed between the observer's likelihood ratio judgments and the output of an electronic device designed to compute a close approximation of the true likelihood ratio (the device is nearly an ideal observer). As shown by the theorem in IV, this correlation may be used to estimate the observer's efficiency.

By using both of these efficiency estimation procedures, it is possible to make deductions regarding the linearity of the observer's operations on the input.

3. Supervised Learning VS Non-Supervised Learning.

The discussion in the present report has emphasized non-supervised learning of the signal waveform pattern. If supervised learning is considered, a great simplification in the form of the ideal adapting observer is obtained. For some types of experimental situations, it may be realistic to apply the theory of adaptive devices in supervised learning tasks. In tasks for human observers where it is not clear whether or not the experimenter's supervision is <u>effective</u>, it is possible to compare the two kinds of models of adaptive devices to attempt to ascertain the effectiveness of the supervision. We expect to review the existing literature concerned with experiments comparing supervised and non-supervised learning in psychoacoustical detection tasks.

4. Correlation Between Observers.

We have worked out a procedure under the aegis of this grant, but not reported here, for estimating the correlation between observers (ideal or otherwise) which uses only the contingency tables obtained from observers serving in the same experimental task. The degree of linear agreement between two observers, which does <u>not</u> depend upon their individual agreement with the ideal observer, may be found by computing a partial correlation between observers' decision axes. The partial correlation may be used to measure interobserver agreement at various stages of learning. We intend to pursue this line of reasoning with the objective of providing a way of ascertaining whether or not different observers use the same type of processing on the input to achieve similar levels of adaptation.

5. Experimental Designs.

In addition to the theoretical efforts being made under this grant, we have been exploring the feasibility of constructing an electronic device to estimate the ideal observer's likelihood ratio on a trial-by-trial basis. Such a device appears possible at moderate cost, and we are proceeding with its design. The output of this device may be compared with observers' estimates of their own decision variable to obtain measures of efficiency as outlined in paragraph 2.

Finally, we are investigating the feasibility of various particular experimental designs which incorporate the preceding ideas and may be used in the further study of psychoacoustic learning.

References

Birdsall, T. G. Detection of a signal specified exactly with a noisy stored reference signal. J. Acoust. Soc. Amer., 1960, 32, 1038-1045.

Birdsall, T. G. Random Processes, Engineering Summer Conference, The University of Michigan, Ann Arbor, Michigan, 1963.

Fralick, S. C. Learning to recognize patterns without a teacher. Stanford Electronics Laboratory. Tech. Rep. No. 6103-10. SU-SEL-65-011.

Green, D. M. Psychoacoustics and detection theory. J. Acoust. Soc. Amer., 1960, 32, 1189-1203.

Hancock, J. A., and Patrick, E. A. Iterative computation of a posteriori probability for M - ary nonsupervised adaptation. Submitted to technical journal, June 1, 1965.

Peterson, W. W., Birdsall, T. G., and Fox, W.C. The theory of signal detectability. Trans. IRE., 1954. PGIT2-4, 171-212.

Raiffa, H. and Schlaifer, R. Applied Statistical Decision Theory. Boston: Graduate School of Business Administration, Harvard University, 1961.

Shuford, E. H., Jr. Some Bayesian Learning Processes. Chapter 9 in M. W. Shelly, II, and G. L. Bryan (Eds.), Human Judgments and Optimallity. New York: John Wiley and Sons, 1964.

Swets, J. A., and Green, D. M. Signal Detection and Psychophysics. New York: John Wiley and Sons, 1966. (In Press).

Tanner, W. P., Jr., and Birdsall, T. G. Definition of d' and η as psychophysical measures. J. Acoust. Soc. Amer., 1958, 30, 922-928.