Operating characteristics and a priori probability of the signal[1]

ARTHUR I. SCHULMAN,[2] *UNIVERSITY OF VIRGINIA*
AND GORDON Z. GREENBERG,[3] *UNIVERSITY OF ILLINOIS*

Two experiments in auditory signal detection produced changes in the operating characteristic as the a priori probability of signal occurrence, p(SN), was varied. The signal was a sinusoid of 1,000 Hz presented for 250 msec against a continuous background of noise. In Experiment 1 three values of p(SN)—0.25, 0.50, and 0.75—were paired with each of three signal intensities. In Experiment 2 the signal intensity was fixed and p(SN) was assigned values of 0.10, 0.25, 0.50, 0.75, and 0.90. On normal-normal coordinates, operating characteristics were fitted to the points obtained from the 4-point rating scale used by the listeners. Such operating characteristics may be specified by two parameters: d_s, an index of detectability related to d', and m, its slope. While d_s was found to be independent of p(SN), m was found to be a joint function of p(SN) and signal intensity. These results are discussed against the background of the theory of signal detectability.

Experiments in signal detection commonly require an O to decide whether or not a specified signal is present in a noisy observation interval. If the O systematically adopts different criteria for acceptance or if he employs confidence ratings, his sensitivity may be described by a single operating characteristic, a function that relates the probability of a detection, $p(y \mid SN)$, to the probability of a false alarm, $p(y \mid N)$. Such an operating characteristic, when plotted on normal-normal coordinate paper, can be described as a linear function throughout a wide range of adopted criteria (Egan et al, 1960, 1961; Egan, Greenberg, & Schulman, 1961). This function is uniquely specified by two parameters: d_s, a measure of the distance of the operating characteristic from the line of chance performance,[4] and m, its slope. The great virtue of the operating characteristic is, of course, that it makes explicit the distinction between the O's decision criteria and his sensitivity; the derived statistic d_s, unlike threshold measures, is an index of detectability largely uncontaminated by those nonsensory factors that may produce large fluctuations in criterion.

For a mathematically ideal O, the operating characteristic is completely and uniquely determined when the physical properties of noise and signal are specified. Such variables as a priori probability of signal occurrence [p(SN)] and the consequences of correct and incorrect decisions merely determine the points on this operating characteristic at which the ideal O operates. Whether these variables alter detectability for the real O is an empirical question; the present research is concerned with the effect of (known) a priori probability of signal occurrence on the operating characteristics of human Os.

Variations in p(SN) are often attended by shifts in response bias (see, for example, Tanner, Swets, & Green, 1956, and Swets, Tanner, & Birdsall, 1961). By manipulating p(SN), it is thus possible to obtain distinct operating points with the use of a binary-decision (usually a yes-no) procedure. But only if evidence, and not merely theory, says that p(SN) does not affect detectability is it justifiable to summarize these obtained points by a single fitted operating characteristic. Were sensitivity to depend upon p(SN)—i.e., if each point obtained with a particular value of p(SN) in fact fell on a different operating characteristic—there would be no justification in describing these points by a common function.

Most published studies that have manipulated p(SN) (e.g., Linker, Moore, & Galanter, 1964; Blosser, 1965; Galanter & Holman, 1967; Markowitz & Swets, 1967; Sorkin, Pastore, & Gilliom, 1968) either have not been concerned with operating characteristics or else have attempted to describe with a single operating characteristic points obtained under different values of p(SN). Since the present experiments were carried out, at least two studies have been reported in which multiple points for an operating characteristic were obtained at each of a number of values of p(SN). One of these (Nachmias, 1968) found no apparent effect of p(SN) on operating characteristics used to describe the detectability of a visual signal. The other (Tanner, Haller, & Atkinson, 1967), a study of auditory signal recognition, is harder to assess. The points obtained for each value of p(SN) are rather close together so that an estimate of an operating characteristic's slope or curvature is difficult. At first glance, there seems to be no systematic effect of p(SN) but a closer look at their data suggests that, as p(SN) increases, so does the slope of the operating characteristic. The two experiments reported here were carried out so that the effects of signal probability on the parameters of the operating characteristic might be better assessed.

PROCEDURE

Both experiments involved the detection of a sinusoid (1,000 Hz, 0.25 sec) embedded in continuously present bandlimited white Gaussian noise. The first study employed three values of p(SN) at each of three signal strengths, whereas the second employed five values of p(SN) at one signal strength. A trial was defined by a sequence of two colored lights. The observation interval, 0.45 sec long, began 0.5 sec after the onset of an amber warning light. A green light marked the observation interval in whose center a signal might occur.

The amber and green lights went off simultaneously at the end of the observation interval, and the listener had 4 sec in which to press one of four numbered keys that represented a scale of confidence ratings. Reasonable use of a four-point scale yields three points to which an operating characteristic may be fitted; the simple computational details and rationale may be found in a previous article (Egan et al, 1960). Listeners were advised to use the rating scale in a symmetrical manner, employing the acceptance end of the scale about as often as they knew signal-plus-noise trials would occur. Their performance was monitored on counters, and they were instructed regularly either to alter or to maintain their distribution of responses over the four rating categories. Instructions were based on more or less arbitrary criteria (e.g., a desired false alarm probability between 0.075 and 0.20 for the lowest operating point on the operating characteristic) and were read to the listeners at predetermined occasions within a session.

The signal voltage was turned on without regard for phase and without the use of special devices, so that the (negligible) transients were determined by the response of the earphone (Permoflux Corporation, PDR-10). The white noise was generated by a 6D4-tube (noise generator, Model 455-B, Grason-Stadler Company). The signal and the noise were mixed electrically and then presented over a binaural headset with the

two earphones wired in parallel and in phase. Four such headsets were wired in parallel so that the instantaneous, as well as the average, signal-to-noise ratio was correlated across a group of listeners. Six listeners were employed and were tested in groups of four and two. Before each session of testing began, a demonstration trial was given with the signal present and easily detectable in the noise. The listeners therefore knew the signal they were to detect as well as its location within the observation interval. However, they were never informed if a signal had in fact been present on a given test trial and so remained ignorant of the correctness of individual decisions.

The strength of the signal relative to the noise will be specified in terms of E/N_0, where E is the signal energy, or the time integral of power, and N_0 is the noise power per unit bandwidth. The overall sound pressure level of the noise was held constant for all tests at about 65 dB re 0.0002 microbar, and E/N_0 was changed by adjusting the signal voltage.

Experiment 1

In this experiment, p(SN) was assigned the values of 0.25, 0.50, and 0.75 for each of three E/N_0s: 6.3 (8 dB), 12.6 (11 dB), and 20.0 (13 dB). Before the experiment was begun, listeners were given 4 days of practice with p(SN) = 0.50 and E/N_0 set to each of the above values. Data were then collected in 24 daily sessions, each of which was devoted entirely to one of the nine experimental conditions. A counterbalanced design was used, and the listener always knew the conditions under which he was being tested: i.e., he knew the a priori probability in effect at any time as well as the relative strength of the signal. Each condition for which p(SN) = 0.50 was presented twice, while each of the remaining six conditions was presented three times. More trials were given at the extreme a priori probabilities in order to increase the reliability of the obtained proportions.

A session consisted of 10 periods of 80 trials each and lasted about 2 h. Short rest intervals were inserted between periods. The first period of each session served as practice and was not included in the data analysis. Ignoring these practice periods, 2,160 trials were presented for each condition where p(SN) was 0.25 or 0.75, and 1,440 trials were presented for the three conditions where p(SN) = 0.50.

Experiment 2

This experiment followed immediately upon the conclusion of Experiment 1. With E/N_0 fixed at 12.6, two sessions of practice preceded 16 sessions of tests at a priori signal probabilities of 0.10, 0.25, 0.50, 0.75, and 0.90. The three intermediate values of p(SN), of course, provided a partial replication of the conditions of Experiment 1. Two different values of p(SN) were represented in each of 12 of the 16 sessions. Each of the extreme a priori probabilities was represented alone in two sessions. Each value of p(SN) appeared in six different sessions, with the

Table 1
Values of d_S Obtained for Each Listener in Experiment 2. The Value of E/N_0 was 12.6.

| Listener | p(SN) | | | | |
	0.10	0.25	0.50	0.75	0.90
L1	0.92	0.96	1.12	0.94	0.85
L2	1.16	0.92	0.98	0.75	0.74
L3	1.70	1.92	1.90	2.05	1.78
L4	1.40	1.35	1.52	1.34	1.41
L5	1.30	1.19	1.23	1.17	1.28
L6	0.65	0.54	0.54	0.62	0.58
Mean	1.19	1.15	1.22	1.14	1.11

exception of p(SN) = 0.50, which appeared in four. A counterbalanced design was again employed. As in Experiment 1, the value of p(SN) in force at any time was known to the listener.

Eleven periods of 80 trials each constituted a daily session. The total number of trials (not including practice) at each experimental condition was as follows: p(SN) = 0.10, N = 3,520; p(SN) = 0.25, N = 1,920; p(SN) = 0.50, N = 960; p(SN) = 0.75, N = 1,920; and p(SN) = 0.90, N = 3,520.

RESULTS

From the operating characteristics obtained in Experiments 1 and 2, values of d_s and m were derived. For the range of signal strength and a priori probability used, d_s was found to be independent of p(SN). This may be seen in Fig. 1 (for Experiment 1) and in Table 1 (for Experiment 2). The eye-fitted curve in Fig. 1 shows the obtained relation between d_s and $\sqrt{2E/N_0}$. (The three clustered
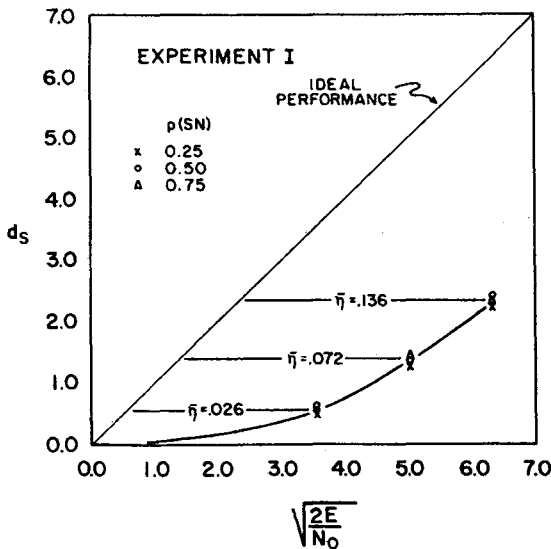

Fig. 1. These data are based upon the operating characteristics obtained in Experiment 1. The derived measure d_s is plotted against $\sqrt{2E/N_0}$ for each value of p(SN) used. The diagonal line represents the performance of an ideal detector to whom the signal is known exactly.
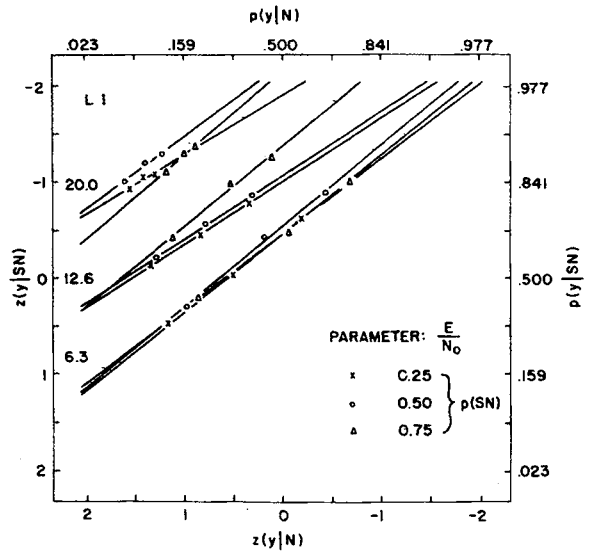

Fig. 2. The nine operating characteristics obtained for one listener in Experiment 1. The linear functions were fitted by eye to the data points.
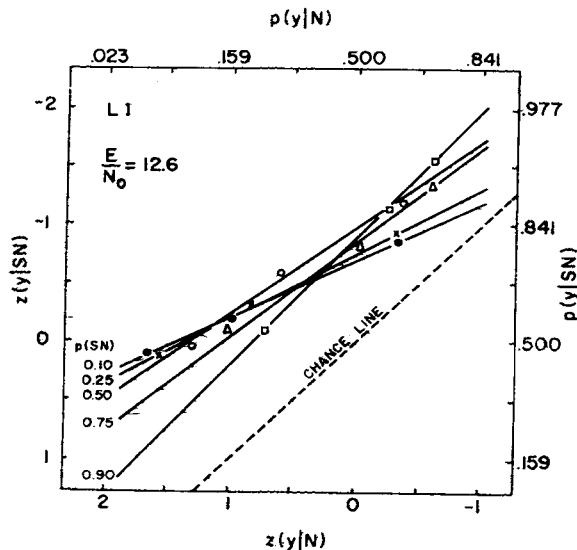
Fig. 3. The five operating characteristics obtained for one listener in Experiment 2. The linear functions were fitted by eye to the data points. The slope m of the operating characteristic increases with p(SN) while d_s remains nearly constant. This listener's performance in Experiment 1 is shown in Fig. 2.

Fig. 4. The effect of a priori probability of the signal on the slope of the operating characteristic, with E/N_0 as the parameter. Each point represents the mean slope of six operating characteristics, one for each listener.

points at each signal level differ in d_s by less than the equivalent of 1 dB, so that a single function which ignores p(SN) seems justified.) Each plotted point in Fig. 1 represents the mean of six values of d_s, one for each of the six listeners. The diagonal line represents the performance of an ideal O to whom the signal is known exactly. The mean efficiency $\bar{\eta}$ (the ratio of the signal energies required by the ideal and real listeners to achieve a given value of d_s) increases with $\sqrt{2E/N_0}$. (See Tanner and Birdsall, 1958, for a discussion of the notion of efficiency.) For Fig. 1, we have calculated the signal energy required by the ideal O to produce a d′ equal in numerical value to the obtained d_s of our real Os. The form of the function shown in Fig. 1 does not change when d′ is in fact substituted for d_s, as the reader may verify; listener efficiency, however, is somewhat improved, since d′ > d_s whenever m < 1, and m, as will be seen below, is usually less than unity. Although Fig. 1 shows that the lowest average value of d_s was obtained at p(SN) = 0.25 for each signal level, this small effect was not reproduced in Experiment 2, as may be verified in Table 1. The nine operating characteristics derived for one listener in Experiment 1 are displayed in Fig. 2. No systematic variation of d_s with p(SN) is evident.

Although d_s was found to be independent of p(SN), the slope m of the operating characteristic was not. A dramatic illustration of the effect of p(SN) on m is given in Fig. 3, which represents for a single listener the eye-fitted operating

characteristics obtained in Experiment 2. Here the slope m increases monotonically with p(SN), while d_s remains nearly constant. All listeners do not show such an orderly relation between p(SN) and m, but the trend seems to be general for $E/N_0 = 12.6$ (see Table 2). The relation between p(SN) and m is displayed in Fig. 4, with E/N_0 as the parameter. The plotted points represent the mean slopes of the operating characteristics of the six listeners, and data from both Experiments 1 and 2 are shown. Figure 4 indicates that m is a joint function of p(SN) and E/N_0. As p(SN) rose from 0.25 and 0.75, m increased from 0.70 to 0.86 when signal strength was high but remained constant at 0.85 for the lowest E/N_0 tested.

DISCUSSION

The theory of signal detectability (TSD) provides a background against which the present findings may be evaluated. According to TSD, the task of an O in a simple detection experiment is essentially one of testing statistical hypotheses, where the alternative hypotheses tested on each trial are SN (signal plus noise was presented) and N (noise alone was presented). Each noisy input must be processed and translated into a decision concerning its binary source, SN or N. An ideal O processes the input by converting it to likelihood ratio, or some monotone transformation thereof; the overlap in the two probability density functions of likelihood ratio (conditional upon SN and N) is greater for weaker signals—i.e.,
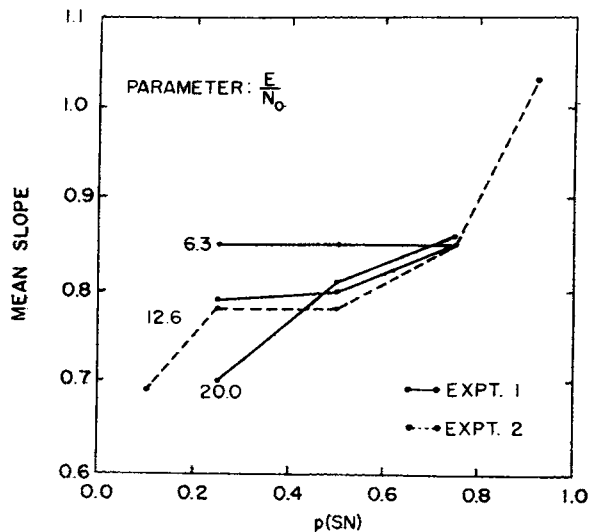
detectability is poorer—but the magnitude of likelihood ratio necessary to accept the SN-hypothesis, no matter how extensive the overlap, is determined entirely by external considerations. When a 4-point rating scale is used, as it was in the experiments reported here, the O may accept the SN-hypothesis with varying degrees of confidence: a rating of "1" (high confidence) may be assumed to correspond to a high value of likelihood ratio, a "4" to a relatively low one. If the conditional distributions of likelihood ratio are both Gaussian, the operating characteristic obtained will be linear in a normal-normal plot, and its slope will depend on the relative variance of these distributions: $m = \sigma_N/\sigma_{SN}$. If the signal to be detected is the same from trial to trial then, for an ideal O, the SN distribution is merely displaced to the right of the N distribution with no increment in its variance. If the signal is known inexactly or statistically,

Table 2
The Slope (m) of the Operating Characteristic as a Function of p(SN) for Each of the Six Listeners. The Data were Obtained in Experiment 2.

Listener	p(SN)				
	0.10	0.25	0.50	0.75	0.90
L1	0.49	0.56	0.74	0.81	1.10
L2	0.80	0.74	0.84	0.86	0.99
L3	0.73	0.91	0.90	0.67	1.12
L4	0.83	1.00	0.71	1.07	1.07
L5	0.64	0.79	0.79	0.94	1.04
L6	0.64	0.69	0.71	0.77	0.87
Mean	0.69	0.78	0.78	0.85	1.03

$\sigma_{SN} > \sigma_N$ and m will be less than unity. The measure of detectability, d', has been defined as the difference between the means of the SN and N distributions, divided by σ_N, and the relationship between d' and d_s is given in Note 4.

In the previous section it was seen that, for relatively strong signals at least, an increase in p(SN) elevates the slope of the operating characteristic but leaves d_s unchanged, so that d' varies inversely, as p(SN). This dependence of d' upon the a priori probability of the signal seems to be at odds with the assumption of TSD that p(SN) may affect the decision criteria but not the sensitivity of the O. But a set of operating characteristics that varies in m but not in d_s does not force upon us the conclusion that sensitivity is altered by the variable that gives rise to a change in slope. It is certainly plausible, for example, that normally insignificant response determinants are differentially amplified according to p(SN), and that they produce the observed changes in m as a result. But plausibility is not proof, and the source of the correlation between p(SN) and m remains unclarified. Appeals to the status of the internally stored comparison signal, to systematic shifts in attention, and to the instability of response criteria, for example, have as yet provided no satisfying explanation for this correlation. It may be significant that a set of operating characteristics, obtained under identical conditions except for p(SN), cross at the only point where the percentage of correct decisions is independent of p(SN). Rather than speculate further, however, we at this time simply observe that m systematically varies with p(SN) and that no other change in stimulus conditions is known to produce a comparable variation. (Uncertainty with respect to the time of signal onset, a variable which might appear likely to affect σ_{SN}, has never been clearly shown to do so.) Slope varies with p(SN), moreover, whether or not that probability is known to the O, as unpublished work by the senior author has disclosed. Finally, the specific effect of p(SN) on m is remarkable in that it must outweigh the "natural" and considerable variability in m found generally in fixed-interval observation experiments.

The present findings underline the difficulty of evaluating listener performance whenever p(SN) is a variable. For example, consider the following question: Other things being equal, is performance better at high or low values of p(SN)? If we take d_s as our index of performance, p(SN) is of no consequence but, if d' is chosen, performance declines (except for very weak signals) as p(SN) increases. Clearly, a meaningful evaluation of the listener's performance must take into account not only (1) the two independent parameters of his operating characteristic (m and either d' or d_s) but also (2) the points on his operating characteristic at which he actually operates. Even with identical operating characteristics, of course, it is possible that a listener's operating points, and consequently his effectiveness relative to some performance criterion, will vary markedly as we change p(SN). On the other hand, if a set of operating characteristics varies in m but not in d_s, it is possible that the O operates so as to obtain the same total payoff (say) for each value of m. That is to say, the O might act as if he knew that m were changing with p(SN), and that adjustments in decision criteria were therefore necessary. Our crew of listeners were periodically instructed to adjust their criteria, so that it is impossible to state with confidence if they would have given us similar operating points if left to their own devices. As a consequence, no easy comparison of performance at different values of p(SN) can be made. Listeners did tend, however, to operate higher on the operating characteristic as p(SN) increased, independently of E/N_o, thereby corroborating the finding of others (Tanner, Swets & Green, 1956; Swets, Tanner, & Birdsall, 1961) with uninstructed Os.

REFERENCES

BLOSSER, A. B. The effects of signal probability and observation period duration upon signal detection performance. Tracor Document No. 65-277-U, 1965, Tracor, Inc., 1701 Guadalupe St., Austin, Tex. 78701.

EGAN, J. P., GREENBERG, G. Z., & SCHULMAN, A. I. Interval of time uncertainty in auditory detection. Journal of the Acoustical Society of America, 1961, 33, 771-778.

EGAN, J. P., SCHULMAN, A. I., & GREENBERG, G. Z. Operating characteristics determined by binary decisions and by ratings. Journal of the Acoustical Society of America, 1960, 31, 768-773.

EGAN, J. P., SCHULMAN, A. I., & GREENBERG, G. Z. Memory for waveform and time uncertainty in auditory detection. Journal of the Acoustical Society of America, 1961, 33, 779-781.

GALANTER, E., & HOLMAN, G. L. Some invariances of the isosensitivity function and their implications for the utility function of money. Journal of Experimental Psychology, 1967, 73, 333-339.

LINKER, E., MOORE, M. E., & GALANTER, E. Taste thresholds, detection models, and disparate results. Journal of Experimental Psychology, 1964, 67, 59-66.

MARKOWITZ, J., & SWETS, J. A. Factors affecting the slope of empirical ROC curves: Comparison of binary and rating responses. Perception & Psychophysics, 1967, 2, 91-100.

NACHMIAS, J. Effects of presentation probability and number of response alternatives on simple visual detection. Perception & Psychophysics, 1968, 3, 151-155.

SORKIN, R. D., PASTORE, R. E., & GILLIOM, J. D. Signal probability and the listening band. Perception & Psychophysics, 1968, 4, 10-12.

SWETS, J. A., TANNER, W. P., JR., & BIRDSALL, T. G. Decision processes in perception. Psychological Review, 1961, 68, 301-340.

TANNER, T. A., JR., HALLER, R. W., & ATKINSON, R. C. Signal recognition as influenced by presentation schedules. Perception & Psychophysics, 1967, 2, 349-358.

TANNER, W. P., JR., & BIRDSALL, T. G. Definitions of d' and η as psychophysical measures. Journal of the Acoustical Society of America, 1958, 30, 922-928.

TANNER, W. P., JR., SWETS, J. A., & GREEN, D. M. Some general properties of the hearing mechanism. Technical Report 30, 1956, Electronic Defense Group, University of Michigan, Ann Arbor, Michigan.

NOTES

4. More precisely, d_s is defined as the difference between $z(y \mid N)$ and $z(y \mid SN)$ at the point where the operating characteristic intersects the diagonal whose equation is $z(y \mid SN) = z(y \mid N)$. In other words, d_s is computed by subtracting the ordinate value from the abscissa value at that point on the operating characteristic where the two types of error, the false alarm and the incorrect rejection, are equiprobable. The measure d' is related to d_s by the following equation:

$$d' = \frac{1 + m}{2m} d_s.$$

When m = 1, the two indices of detectability are equivalent. For studies of human detection in which operating characteristics are secured, d_s is to be preferred to d' since it is more stable, i.e., less affected by the estimated value of m than is d'.

(Accepted for publication January 2, 1970.)