Perception & Psychophysics
1974, Vol. 15, No. 2, 291-300

Variability of magnitude estimates: A timing theory analysis*

DAVID M. GREEN†
University of California, San Diego, La Jolla, California 92037

and

R. DUNCAN LUCE
School of Social Sciences, University of California, Irvine, Irvine, California 92664

Three procedures for magnitude estimation were investigated, and a sufficient number of responses were obtained to make reasonable estimates of both the mean and variance of the responses. The conventional magnitude estimate procedure, without a standard signal, appeared to produce the most sensible data. The best method of establishing the central tendency of the data appears to be the plot of the mean ratio of successive responses against the intensity ratio of the corresponding signal intensities. When this is done, the average response ratio increases roughly as a power function of the signal ratios. The coefficient of variation, σ/m, varies from about 0.1 for small signal ratios and increases to 0.3 at about 20 dB and greater signal separations. The distribution of response ratios appears to be reasonably well approximated by a beta distribution. The change in σ/m with signal ratio is suggestive of an attention mechanism in which the sample size depends on the location of the attention band. The ratio estimation procedure suffers badly from discrete number tendencies.

1. INTRODUCTION

Anyone who has collected magnitude estimation data knows that the responses are highly variable. This variability is both between observers, with exponents of the power function fitted to loudness data ranging from at least .15 to .60 (using an intensity measure), and within observers, with the coefficient of variation (σ/m) being of the order of .2-.3 (Luce & Mo, 1965; Schneider & Lane, 1963; Stevens & Guirao, 1964). The question we wish to consider here is whether this variability contains any interesting information.

The variability among observers strikes us as no more exciting than variation of weight or height except in one circumstance, namely, if the exponents from several different modalities are strongly correlated. Surprisingly, that does not seem to have been extensively studied; the only reference we know of is Ekman et al (1968). In the absence of firm data on this point, the primary current significance of the between-observer variability is a cautionary one: one should not average over observers unless one is quite sure of the functional form of the data, so that the true form will not be distorted in the averaging process.

Our concern is not with this variability, but with the variability within a single observer emitting responses at different times. Do such data contain any interesting information about the internal representation of signals? Stevens repeatedly urged the view, both in public (1957, 1959, 1961a, b, 1971) and in private, that it does not, that it is—like most "noise" which muddies all physical measurements—ubiquitous and uninformative. This would be true were the variability due primarily to randomness wholly unrelated to the sensory-decision process underlying magnitude estimation, but is decidedly not true if it is intimately related to that process. Our goal here is to make a case for the informativeness of this variability, provided that one looks at it correctly, namely in terms of the ratios of magnitude estimates to successive signals. More of the analysis procedure later.

Theories for magnitude estimation in which variability plays a key role have been suggested by Durlach and Braida (1969), Braida and Durlach (1972), and Luce and Green (1972). In particular, the latter assumed, as was done earlier by McGill (1963, 1967), Siebert (1965, 1968, 1970), and others, that the signals are, to a first approximation, represented internally as independent Poisson processes on a number of parallel channels. The intensity parameters of these Poisson processes are assumed to be increasing functions of signal intensity. What differs in the Luce-Green approach is the emphasis on timing rather than counting procedures. A counting rule is one in which the pulses that occur in a fixed time are counted, whereas a timing rule is one in which the time is measured for a fixed number of pulses to occur. They presented data that are consistent with the interpretation that both modes of behavior are available to human observers; however, they argued that the timing mode is probably the one normally employed except in psychophysical experiments with very brief

*Supported in part by grants from the National Science Foundation to the University of California, at Irvine and at San Diego. Requests for reprints should be sent to R. Duncan Luce, School of Social Sciences, University of California, Irvine, Irvine, California 92664.

†Now at the Department of Psychology and Social Relations, William James Hall, Harvard University, Cambridge, Massachusetts.

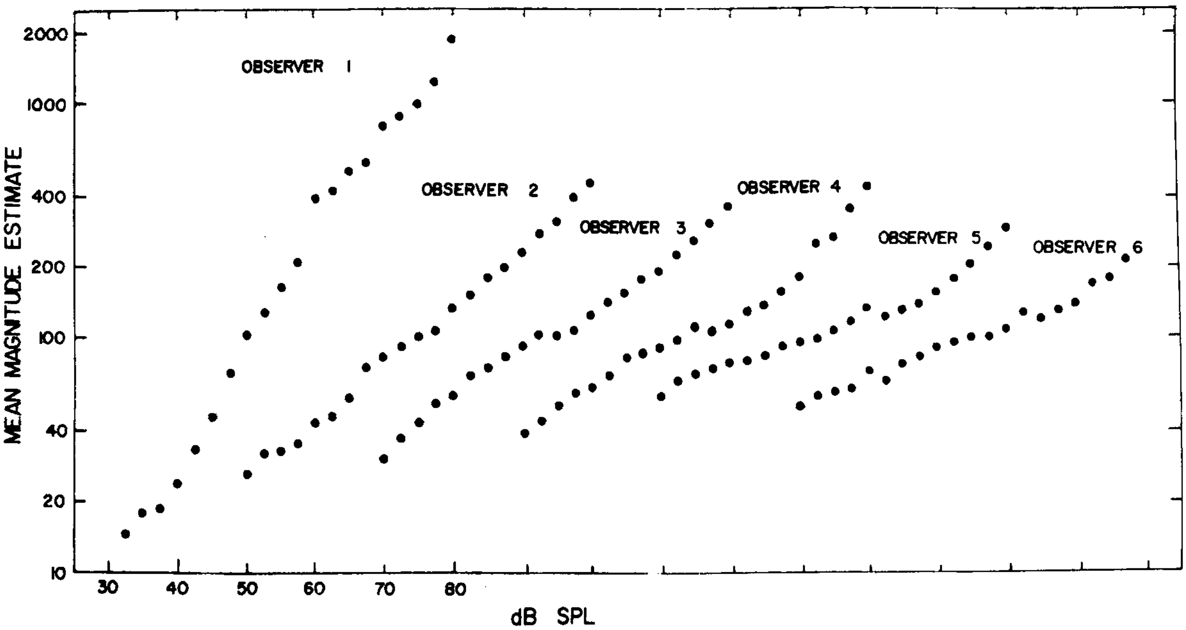

Fig. 1. Mean magnitude functions for six observers of Experiment 1. The smallest intensity in each case is at 32.5 dB SPL, and the standard, of 100, is at 50 dB SPL.

signals or extreme time pressure (Green & Luce, 1973).

If we assume that the timing mode is used in magnitude estimation experiments (especially those with signals of long duration), then a natural internal measure of the intensity of the signal is the reciprocal of the average time required to get κ interpulse intervals on each of the J active channels. So each magnitude estimate is based on a total sample size of $k = \kappa J$. For a Poisson process, the time between pulses is exponentially distributed with intensity parameter $\mu(s)$, where s denotes the signal; we make the idealization that it is the same parameter on all channels. Thus, the average time for k intervals is distributed according to gamma with intensity $\mu(s) \cdot k$ and order k. This suggests analyzing the distribution of the *reciprocals* of magnitude estimates, since that quantity should be gamma distributed. This empirical distribution would provide us information both about how $\mu(s)$ grows with signal intensity, I(s), and about the sample size, k, used by the observer (Luce & Green, 1972, p. 28).

The present empirical work began with the preceding theoretical observation. The failure of the initial model, together with recent work on sequential dependencies (see Section 3), led to a more subtle analysis of the decision process underlying magnitude estimates and to two more experiments, reported in Sections 4 and 5. The Discussion section outlines one possible interpretation of the results.

2. EXPERIMENTAL PROCEDURE

In all three experiments, the random sequence of signals was programmed and the data recorded by a PDP-15 computer. Signals were 1,000-Hz tones of 1 sec duration. They were presented, in quiet, binaurally via SW-2 Superex headphones to undergraduate observers who were run individually in sound-attenuating rooms. The observers responded, with either whole or decimal numbers, via a Video System Teletype. The procedure was self-paced, with each trial being initiated by the observer's response, at an average rate of 10 trials/min. Runs consisted of 80 trials, between which there was a short break, for a total of 2 h during a period of 2-4 weeks. The exact number of trials per signal per observer varied from 80 to 300. Six observers were run in each experiment, the same ones in Experiments 2 and 3, with Observers 1, 3, and 6 having Experiment 2 first and Observers 2, 4, and 5 having Experiment 3 first. They were paid $2.13 per hour.

Experiment 1. Magnitude Estimation With a Standard (MES)

The 20 signals varied in 2.5-dB steps from 32.5 to 80 dB SPL. The standard was 55 dB and was given a numerical value of 100. The observer could call for the standard at any time and did, on the average, about 1 in 7 trials. Mistakes in responding could be corrected by pressing an appropriate key. The instructions were relatively standard (see Appendix I).

Experiment 2. Magnitude Estimation Without a Standard (ME)

As a result of the well-known perturbing effect of the standard and as a result of a possible ambiguity in analyzing the basis of the observer's responses (see Section 3), we elected to rerun Experiment 1 without using any standard. At the same time, we shifted the signal range up 10 dB (42.5 to 90 dB), and we modified the response procedure as follows in an attempt to further minimize erroneous responses. Each response was in a standard format of up to five decimal digits before the decimal point, followed by up to four digits. After each response, it was displayed and the observer was required either to type Y to verify his response and to proceed or to correct it. The exact instructions are in Appendix II.

Experiment 3. Ratio Estimation (RE)

Two signals, 400 msec apart, were presented on each trial, and the observers were instructed to judge their subjective ratios. The weaker signal was always to be called 1 and the stronger, an appropriate, larger number. So responses were either of the form 1:n when the first signal was the weaker and the second n times as loud or n:1 for the opposite order. Responses were verified as in Experiment 2. Rather than use all of the pairs generated from a random sequence of 20 different signals, we selected the following subset of 34 pairs, which were run in each order and equally often. We describe them by the (several) values of the smaller signal and the signal difference in decibels.

smaller signal (dB)	difference (dB)
40	50
40, 65	25
40, 50, 60, 70, 80	10
40, 50, 62.5, 75, 85	5
40, 50, 62.5, 77.5	2.5

The instructions are in Appendix III.

3. RESULTS AND DISCUSSION OF MES EXPERIMENT

Mean magnitude estimates for the six observers of Experiment 1 are shown in Fig. 1. As is by now familiar from a number of studies (Braida & Durlach, 1972; Luce & Mo, 1965; Pradhan & Hoffman, 1963), functions obtained with this many observations per point are roughly power functions, but with deviations of as much as 5 dB from the best fitting power functions. (This is equally true when only a few observations are obtained, but the variability of one or two responses per signal is so great that one cannot reject the hypothesis of a power function; see Ekman et al, 1967, and Stevens and Guirao, 1964.) The "slopes" vary from .15 to .55, which seems typical. A perturbation is noticeable in each case at the standard (a magnitude of 100), and this was one reason why we were led to run Experiment 2.

A primary reason for running Experiment 1 was to test the hypothesis that the reciprocal of magnitude estimates should be distributed as gamma. The order k of the gamma can be estimated from the average (over signals) of the coefficient of variation, specifically,

$$k = \frac{1}{(\sigma/m)^2},$$

and it is interpreted as the size of the sample of interpulse times on which the magnitude estimate is supposedly based. The values obtained are shown in Table 1. A correlation with slopes appears to exist. Plots of the cumulative normalizd distribution and the fitted gamma for Observers 1 and 6, the smallest and largest k, are shown in Figs. 2 and 3. The former is a relatively satisfactory fit, whereas the latter, which is more typical,

Table 1

O	1	2	3	4	5	6
k	3	13	20	10	30	50

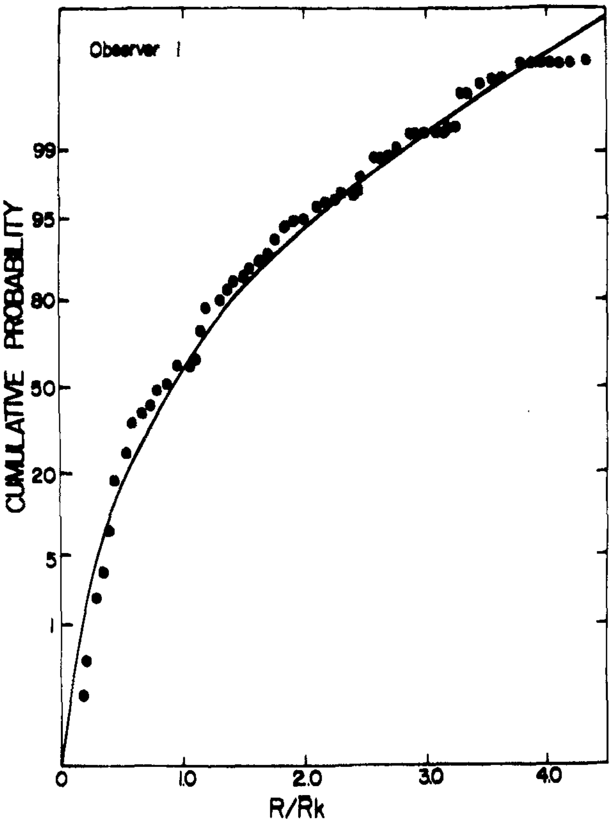

Fig. 2. Distribution for Observer 1 of magnitude estimates, normalized by dividing by the mean, $\bar{R}$, corresponding to the signal level and by the estimated sample size, k. The theoretical curve is that of the gamma distribution with intensity 1 and order k.

is very poor at the extreme values, with the tails of the observed distributions considerably higher than that predicted by a gamma. For Observer 6, this discrepancy is apparent only below the 1% or 2% and above the 98% values. For other observers, however, the discrepancies were noticeable below 10% or 20% and above 80% or 90%.

4. MOTIVATION FOR THE ME EXPERIMENT

After these data were collected, we became aware of a series of papers (Cross, 1973; Garner, 1953; Holland & Lockhead, 1968; Ward, 1972, 1973; Ward & Lockhead, 1970, 1971) showing in both absolute identification (AI) and ME procedures the existence of pronounced sequential effects. The largest sequential effect appears to arise from the preceding signal, although much attenuated effects may arise from signals occurring further back in time. One possible interpretation of these results is that the observer uses the signal and response on the preceding trial as the basis for responding on the present trial.

We make this hypothesis explicit. Using bold-faced symbols to denote random variables, the signal presented on Trial n is $\mathbf{S}_n$ and the response is $\mathbf{R}_n$. We assume that

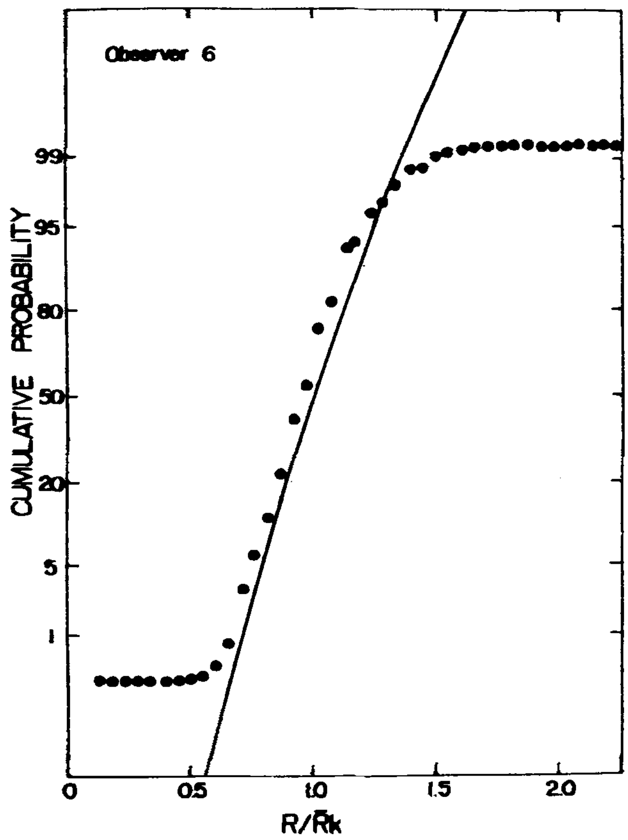

Fig. 3. This is the same as Fig. 2 for Observer 6.

whenever Signal s is presented, it is represented internally as a random variable X(s). Note that this depends only on the signal presented, not on the trial number. The hypothesis, then, is

$$R_n = R_{n-1}X(S_n)/X(S_{n-1}). \tag{1}$$

If so, then the data we should study are not the magnitude estimates themselves, but rather the ratio of successive responses to the pair of signals that generated them, as was done by Ward (1973). A detailed study of this hypothesis, including a generalization needed to yield the sequential effects observed, is given in Luce and Green (1974). This generalization may also explain the sequential effects reported by Ward and Lockhead (1971) when there is no signal at all. The theoretical distribution for the ratio of Eq. 1 on the assumption X is determined from a timing model is not a gamma, but rather that of the ratio of two gamma distributed random variables.[1] As is easily shown, if the total time T(s) for k(s) interpulse intervals is distributed as a gamma with intensity $\mu(s)$ and order k(s) when Signal s is presented and if we set X(s) = k(s)/T(s), then for those pairs of successive trials on which Signal s is preceded by Signal s′ Eq. 1 yields

$$P\left(\frac{R_n}{R_{n-1}} = x \mid S_n = s \;\&\; S_{n-1} = s'\right) = P\left(\frac{T(s')}{T(s)}\,\frac{k(s)}{k(s')} = x\right). \tag{2a}$$

Observe that

$$P\left(\frac{T(s')}{T(s)} = y\right) = \int_0^\infty P[T(s) = z]\, z\, P[T(s') = zy]\, dz$$

$$= \frac{1}{B[k(s),k(s')]}\,\frac{y^{k(s')-1}\,[\mu(s')/\mu(s)]^{k(s')}}{[1 + y\mu(s')/\mu(s)]^{k(s)+k(s')}}, \tag{2b}$$

where

$$B(k,k') = \frac{(k-1)!\,(k'-1)!}{(k+k'-1)!}.$$

Equation 2b is the beta distribution of the second kind or equivalently the F distribution with 2k(s), 2k(s′) degrees of function. It is readily shown that

$$m = E\left(\frac{R_n}{R_{n-1}} \mid S_n = s \;\&\; S_{n-1} = s'\right) = \frac{\mu(s)}{\mu(s')}\left[\frac{k(s)}{k(s)-1}\right], \tag{3}$$

and

$$\left(\frac{\sigma}{m}\right)^2 = \frac{k(s) + k(s') - 1}{k(s')\,[k(s) - 2]}. \tag{4}$$

We could not test this theory with the data which employed a standard since the standard was used on about 15% of the trials. We assume that when it is called for, the standard, rather than the preceding signal, serves as the reference. Since we had not recorded when the standard was called for, it was necessary to repeat the experiment. For reasons cited earlier, we elected to drop the standard in the revised experiment. In addition, it seemed desirable to compare these results with a situation where the same observers were told to report the subjective ratio of two signals, which led to Experiment 3.

5. RESULTS OF ME EXPERIMENT

Figure 4 presents the mean magnitude estimates for this experiment. The functions again show systematic departures from power functions, and the "slopes" range from .18 to .35. We show these mainly to contrast them with the mean magnitude ratio on successive trials,

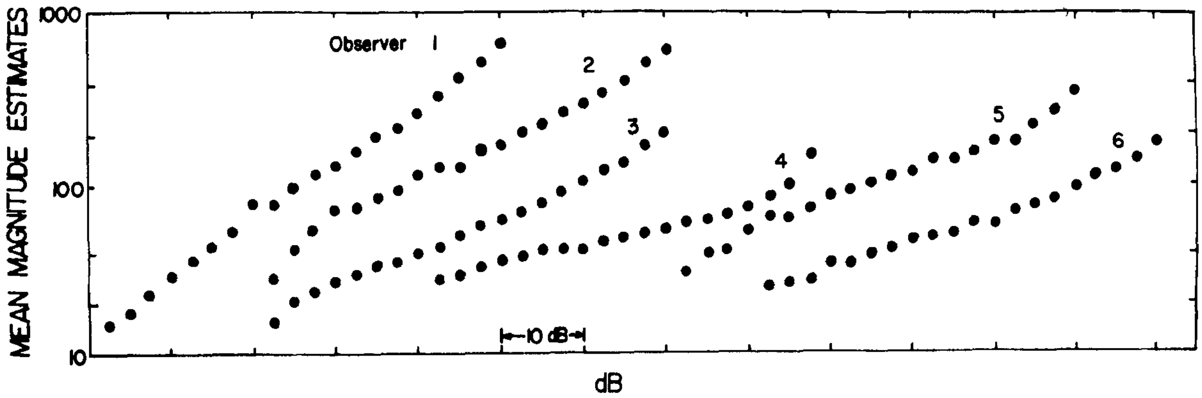

Fig. 4. Mean magnitude functions for six observers of Experiment 2. The smallest intensity in each case is at 42.5 dB SPL. There was no standard.

which are shown as circles on the top portion of each panel in Fig. 5.

Several things should be noted about these ratio functions. First, they are considerably closer to power functions than are the mean magnitude functions. Second, the slopes are little changed, now ranging from .18 to .34. Third, the averaging is not only over all response ratios to a given stimulus pair (s,s′), but over all pairs having the same intensity ratio (or difference in decibels), and also over the reciprocal ratio of responses when the signal ratio is less than 1. The empirical justification for this additional averaging is that we could not detect any systematic trend as a function of signal level or the order of the signal presentation. The theoretical motivation for it is the fact that previous work with reaction times suggested that μ is approximately a power function of signal intensity (see Luce & Green, 1972, p. 45), which in Eq. 3 suggests that the mean response ratio should be approximately a power function of the signal ratio, provided that k does not depend strongly on intensity. We argue below that it does not in a direct way. And fourth, these functions are based only on data which do not deviate too much from the central tendency. Specifically, means were calculated from all of the data, and all responses that differed by a factor of more than 5 from the mean were ignored. This makes little difference in the mean, but it matters a great deal for higher moments. One can see from Eq. 2 that, if the theory is correct, high tails of the form $y^{-k(s')-1}$ are to be expected, and with such high tails, the variability in estimates of moments is considerable. We have elected to introduce some bias in the estimates in order to increase their stability.

To estimate k uncontaminated by μ, the theory suggests (Eq. 4) examining the coefficient of variation, σ/m. For each signal pair having the same intensity ratio, we estimated σ/m (eliminating extreme responses) and again could see no evidence of any systematic trend, which suggests that it is appropriate to combine them. These data are shown as the circles in middle portion of each panel. The pattern is completely consistent: the ratio increases from its minimum value of about .1 when the two successive signals are identical to a plateau of about .3 or .4, running from about 20 dB to at least 40 dB. In the case of Observers 1, 4, and 6, there is some suggestion of a drop again at the longest separations. The average is about .2-.3, which, as we noted in Section 1, is the range of values previously reported.

Although something systematic clearly is changing with signal ratio, apparently it is not simply a change in k with intensity. For simplicity, let us consider the hypothesis that for any pair (s′, s), we have

$$k(s) = k(s') = k[I(s)/I(s')], \tag{5}$$

i.e., the sample size does not depend on signal level, but it does depend on the ratio. (See Section 7 for further discussion of this relationship.) On this assumption, we used Eq. 4 to get an initial estimate of $k[I(s)/I(s')]$, and then we sought the value of k which minimized chi squared in fitting the beta distribution to the data. All of the data were first sorted into 13 categories by the differences in successive signal level; they are (1) 0 dB, (2) 2.5 dB, (3) 5 dB, (4) 7.5 dB, (5) 10 dB, (6) 12.5 dB, (7) 15 dB, (8) 17.5 dB, (9) 20 dB, (10) 22.5 and 25 dB, (11) 27.5 and 30 dB, (12) 32.5, 35.0, and 37.5 dB, and (13) 40, 42.5, 45.0, and 47.5 dB. These choices generated roughly an equal number of observations in each category. A value of k was estimated for each change in signal level and eight expected values were computed by using the 5%, 15%, 25%, 50%, 75%, 85%, and 95% levels of the beta distribution. The chi squared has 5 df–7 independent probabilities less 2 estimated constants. Thus, the expected chi squared is 5 and the distribution of chi-squared values should have a standard deviation of 3.1. Table 2 shows the observed distribution of chi-squared values for the six observers. The obtained values of chi squared are all larger than what would be expected under the null hypothesis–two of the observers, 4 and 6, had rather poor fits. But several factors lead us to conclude that the fits are fairly good. First, the data for positive and negative changes in signal

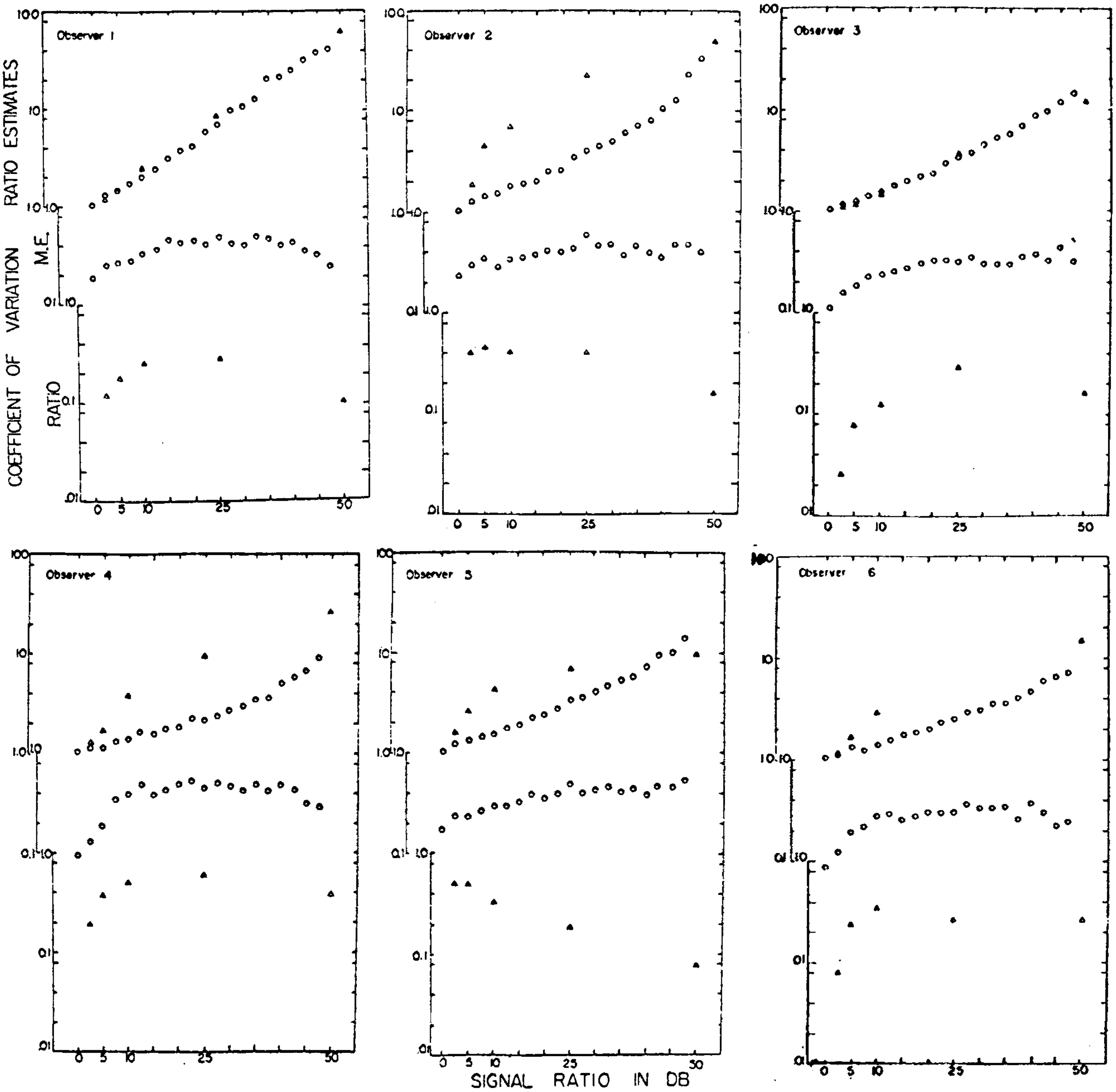

Fig. 5. Mean ratio estimates and coefficient of variation for both ME (circles) and RE (triangles) experiments.

level have been combined. Second, equal degrees of freedom have been assumed in the numerator and denominator of the beta, thus the expected distribution is nearly symmetric, especially for large values of k. We know this is false, and we will comment later on this matter. Finally, we should note that the observed distribution of responses is considerably more discrete than this continuous model would predict. Three observers (1, 3, and 6) responded with the same number (one) on more than 50% of the trials when the stimulus

Table 2

O	χ^2 Mean	χ^2 σ	χ^2 Range (Max-Min)	N
1	15.8	8.8	32.7	3134
2	15.6	10.0	27.7	3286
3	14.2	5.3	17.4	2907
4	65.3	43.2	132.6	3782
5	16.5	11.7	32.5	3038
6	46.5	35.7	125.9	2758

had the same intensity on two successive trials (0-dB difference). The other observers (2, 4, and 5) had the same tendency, but the percentage was much lower, less than 25%. For Observers 1, 3, and 6, therefore, Category 0 dB was excluded from the chi-squared analysis. This general tendency to repeat the same response was also apparent in all trials when the signal intensity changed but little. There was also the tendency to repeat common numbers, and thus certain simple ratios occur frequently in the data, although the analysis in terms of the ratio of successive responses obviously attenuates the occurrence of common ratios as compared with the frequency of the common responses themselves.

In this connection, we must raise the possibility that the reduction in σ/m with decreasing signal ratio is due to the well-known tendency of observers to use only a restricted set of numbers, usually simple fractions, the digits, and multiples of 5. An inspection of our raw data confirms the existence of such a round number tendency in these data, and so the possibility of an artifact is real. However, as we will see (Section 6), for at least one distribution of stimulus ratios, two observers altered their behavior in such a way that σ/m is largest for small signal ratios. This suggests that something more than a round number tendency is involved. It is not difficult to formulate plausible hypotheses about the round number tendency that support either view. Our present plan is to bypass the issue in future work by using cross-modal matches.

To appreciate how well the beta distribution fits the data, Table 3 shows the expected and obtained percentages averaged over the various categories. As the chi-squared fit indicates, the fit for Observer 4 is rather bad, but the others are reasonably satisfactory. Note, however, there is a systematic discrepancy. The percentage of large responses (top rows of table) is somewhat greater than expected; whereas, the percentage of small responses (e.g., bottom two rows) is always less than expected. Thus, for all observers, the tail on one side overshoots and on the other side undershoots the beta distribution.

6. RESULTS OF RE EXPERIMENT

The triangles in Fig. 5 are the means and coefficient of variation of estimated ratios. The means of ME and RE are superimposed, whereas the σ/m estimates are displaced. Recall that the distribution of stimuli was different for RE and ME.

Observers 1 and 3 exhibit little difference between the procedures, except Observer 3 has a smaller σ/m for RE than ME when the signal difference is small.

Observers 2 and 5 yield both different mean and different σ/m functions. The RE mean curve is concave above that from ME. The σ/m of the RE procedure is a decreasing function of signal ratio, whereas that of ME is

Table 3

	Expected	1	2	3	4	5	6
	1	1.5	1.9	2.3	3.1	1.4	1.7
Large	1.5	1.4	1.2	1.3	1.6	1.4	.9
	2.5	1.9	2.3	2.4	1.9	2.3	2.9
	5	4.6	3.9	4.2	3.1	4.2	3.5
	15	12.7	12	12.3	8.2	13.3	11.4
	25	22.9	25.7	23.8	19.9	25.8	24.2
Response	25	28.7	27.6	27.7	34.0	27.1	29.9
	15	17.5	17.1	17.3	22.1	17.0	19.7
	5	4.8	4.6	5.4	3.9	3.9	3.2
	2.5	2.3	2	1.9	1.5	1.6	1.6
Small	1.5	1.0	.8	.8	.3	1.2	.7
	1.0	.6	.8	.6	.2	.8	.3
	N =	3134	3286	2907	3782	3038	2758

Note–Entries are percentages

increasing; the difference is especially noticeable for Observer 5.

Observers 4 and 6 exhibit the same difference between the means as do Observers 2 and 5, but their σ/m functions are very similar.

This evidence strongly suggests that the deviations in the RE behavior from the ME behavior are due to two distinct mechanisms, since, given that the means differ, σ/m may or may not. The deviation in the means, when it occurs, closely resembles that obtained when observers are asked to make category judgments. If we assume that they are not distinguishing absolute level, but only signal ratios, then instead of 17 stimuli there were only 5, and it may not be surprising that observers reverted to some form of categorizing. This hypothesis is further supported by a look at the distribution of responses. The distributions are extremely discrete; for example, when the stimulus ratio was ±2.5 dB, Observer 4 used only 9 distinct numbers in 828 responses, and 677 of these were the number 1. For the ±50-dB stimulus pair, 15 different numbers (all integers) were used in 181 responses and the numbers 10, 15, and 20 were used a total of 116 times. Observer 5, who has the unusual σ/m variation with signal ratio, undoubtedly absolutely identified the large change in signal (50 dB): he called it 20 on 37% of the trials.

As can be appreciated from the brief description of the response morphology, the fit of a beta distribution to these data was extremely poor. Once again, the value of k was estimated from the coefficient of variation and a chi-squared test with 5 df was fit to the data. The resulting chi squared ranged from 4.3 to 1,795, with a median of about 200. No observer was fit well, and the expected and observed distribution always displayed the same discrepancy. There was always an excess of observations in the middle two quartiles.

Because of the discreteness of the response distribution, which is evident in both ME and RE data, we favor the ME procedure in which the ratios of successive responses are reported. The ratios muddy the discreteness to a degree.

The differences in the form of the σ/m may also be related to the distribution of the stimuli, a topic we consider more fully in the next section.

7. DISCUSSION

Had the coefficient of variation turned out to be independent of both signal level and signal ratio, we would have had little to discuss. Had it depended just on signal level, and so on signal ratio of successive signals in an indirect way, we would simply have acknowledged that sample size varies with intensity, because of some sort of sequential decision procedure or for some other reason. But the data suggest something more subtle, and interesting, than either of these possibilities. The sample size appears to depend on signal ratio in a way that may, but need not, be affected by the distribution of ratios presented: with that of a purely random schedule (ME), all six observers exhibited larger sample sizes for small ratios than for large ones, whereas, with the peculiar distribution used in RE, two observers exhibited the reverse relationship. This difference, which hints at a difference in the mode of observing in the two tasks, suggests that the mechanism is not purely sensory.

A possible explanation is that "attention" is unevenly distributed over intensities, with the main focus concentrating on a limited band of intensities. In idealized form, suppose a band of width Δ in pulse rate can be centered on each trial at any rate (equivalently, intensity). If the signal on that trial has a rate which lies within the band, then the size of the sample of interpulse intervals is K, whereas if it lies outside the band, the sample size is K_0, where $K_0 < K$. The strategic problem for the observer is where best to locate the band on each trial. One plausible idea is that it is shifted to the region of the preceding signal. If it were to center exactly on the true rate of the preceding signal, then the σ/m function for the ME experiment would increase discontinuously at $\Delta/2$. The argument is straightforward. If the signals are equally spaced over a range of rates, ρ, then with probability Δ/ρ the preceding signal had a sample size of K and with probability $1 - \Delta/\rho$ the sample size was K_0. For the following signal, if the difference is less than $\Delta/2$, the new signal has sample size K, and for larger differences, it has sample size K_0. So, for the small differences, the pair of sample sizes are

$$(K, K) \text{ with probability } \Delta/\rho$$

$$(K_0, K) \text{ with probability } 1 - \Delta/\rho$$

and for the large difference

$$(K, K_0) \text{ with probability } \Delta/\rho$$

$$(K_0, K_0) \text{ with probability } 1 - \Delta/\rho.$$

If there is some variability in the location of the interval about the preceding signal, as there must be if the estimated rate is used, the discontinuity is obviously rounded. Assuming that, the ME data suggest that Δ corresponds to approximately 20 dB.

Note that this implies that the response distributions are based on mixes of either (K, K) and (K_0, K) sample sizes or of (K, K_0) and (K_0, K_0) sample sizes, and so our fitting them with a (k, k) sample size model is surely wrong in detail. Indeed, our estimates must satisfy the inequality

$$K_0 < \min k < \max k < K.$$

It is evident, by an analogous argument, that if the observer avoids attending to the region of the previous signal, then the value of σ/m will be large for small differences and small for large ones. As differences of 10 dB and larger were almost as common in the RE design as the smaller ones, 5 and 2.5, and perhaps more noticeable, it is plausible that some observers may have adopted this strategy. Indeed, even from the ME data there is some indication that several observers may have shifted to the opposite end of the range following either a soft or a loud signal, as evidenced by the decrease in σ/m for large ratios in Observers 1, 4, and 6.

It is interseting that this hypothesis accounts qualitatively for a well-known, but perlexing, fact about AI results. It is this: Consider an AI design with $n \geqslant 8$ signals, equally spaced in decibels over the range R. With n fixed, vary R. Were we to assume a fixed sample size associated with each signal, then increasing R, which necessarily increases the separation betweeen adjacent signal intensities would decrease the overlap of the sampling distributions of the signals, thereby making them easier to identify. In fact, the predicted improvement occurs only for smaller ranges, after which the actual improvement grows considerably more slowly than predicted by such an argument (Braida & Durlach, 1972; Garner 1952, 1953; Hake & Garner, 1951; Pollack, 1956). If, however, there is an attention band of about 20 dB, improved performance is predicted only as long as all signals remain within the band, after which there is an increasing probability that a presented signal will fall outside the band, in which case its representation is based on a considerably smaller sample size. Thus, increasing the range beyond 20 dB brings into play opposing forces: an increase in the separation of the means of the sampling distribution counteracted by an increase in the variance due to reduced sample sizes. The relative balance of forces depends, of course, on the exact numbers involved, and we have not worked out any detailed models for either AI or ME experiments.

The postulate that observers tend to center their attention on the preceding signal, when all signal intensities are equally likely, predicts the known tendency for observers to be considerably more accurate when a signal is repeated (Swets & Green, 1961; Swets, Shipley, McKey, & Green, 1959).

Although this attention hypothesis seems to make sense of both the ME and AI data, as a theory it involves enormous freedom, since we can, post hoc, select a strategy of attenting for each set of data. This is not satisfactory unless we can devise methods either to determine the strategy in use or to manipulate it experimentally. The use of strong sequential dependencies in the schedule of signal presentations or differential payoffs for small σ/m for certain signal ratios are all suggested by this hypothesis, but they have not yet been tried. So, at present, the attention hypothesis is largely speculative, but a possible source of direction for future development.

APPENDIX I. INSTRUCTIONS FOR MES

This is an experiment to determine your ability to judge the relative loudness of tones. You will hear a series of tones of different intensities; your task is to tell me how loud they are by assigning numbers to them. The ratios between the numbers you assign to the different tones correspond to the ratio between the loudnesses of the tones. In other words, try to make the numbers proportional to the loudness as you hear them. We will call the standard stimulus 100 units. If the second stimulus sounds 7 times as loud, call it 700. If it sound ½ as loud, call it 50, and so forth. Don't worry about the first few judgments too much; we will give you several hours of practice before we begin collecting data. Try to arrive at some systematic procedure so that after listening to the tone you will be able to assign the numbers quickly and consistently. Since there are a large number of different stimuli, it is unlikely the same stimulus will repeat within any brief interval of time. Thus, make each judgment by comparing it with the standard and ignore as much as possible the preceding stimulus. Any time you wish to hear the standard, you need only type the letter S and the standard will be presented. You may also listen to the loudest and the softest stimuli, at the very beginning of a session (see procedure instruction sheet).

APPENDIX II. INSTRUCTIONS FOR ME

This is an experiment to determine your ability to judge the relative loudness of tones. You will be presented a series of tones of different intensities in an irregular order; your task is to tell me how loud they are by assigning numbers to them. The ratios between the numbers you assign to the different tones should correspond to the ratio between the loudness of the tones. For example, if one tone sounds 20 times louder than another, then, if you had assigned 8 to the weaker tone, assign 160 to the louder one. In other words, try to make the number proportional to the loudness as you hear it. You may assign any positive number you wish to the tones. There are a large number of intensities, so you will use many different numbers. However, try to be consistent in the sense that if two intensities are near each other then the numbers you assign should be near each other; but they should exhibit the proper ratio. The computer program does not accept fractions, so it may be helpful and easier for you to assign a fairly large number like 100 to a middle loudness. The program will accept decimal numbers, so if you want ¼ enter .25 as your response. Don't worry about the first few judgments; we will give you several hours' practice before we begin collecting data.

APPENDIX III. INSTRUCTION FOR RE

This is an experiment to determine your ability to judge the relative loudness of tones. You will be presented a pair of tones of different intensities. They will be presented in an irregular order; your task is to tell me how they differ in loudness by assigning a ratio to them. The ratios between the numbers you assign to the different tones should correspond to the ratio between the loudness of the tones. Make the softer tone equal to one. If the louder tone is 20 times louder and it was the first tone, type "20 ESC 1." If the louder tone is the second tone of the pair, type "1 ESC 20" [on the VST it looks like 1:20]. In other words, try to make the number proportional to the loudness as you hear it. You may assign any positive number you wish to the tones. There are a large number of intensities, so you will use many different numbers. However, try to be consistent in the sense that if two intensities are near each other, then the numbers you assign should be near each other; but they should exhibit the proper ratio. Don't worry about the first few judgments; we will give you several hours of practice before we begin collecting data.

REFERENCES

Braida, L. D., & Durlach, N. I. Intensity perception. II. Resolution in one-interval paradigms. Journal of the Acoustical Society of America, 1972, 51, 483-502.

Cross, D. V. Sequential dependencies and regression in psychophysical judgments. Perception & Psychophysics, 1973, 14, 547-552.

Durlach, N. I., & Braida, L. D. Intensity perception. I. Preliminary theory of intensity resolutions. Journal of the Acoustical Society of America, 1969, 46, 372-383.

Ekman, G., Hosman, B., Lindman, R., Ljungberg, L., & Åkesson, C. A. Interindividual differences in scaling performance. Perceptual & Motor Skills, 1968, 26, 815-823.

Garner, W. R. An equal discriminability scale for loudness of judgments. Journal of Experimental Psychology, 1952, 43, 232-238.

Garner, W. R. An informational analysis of absolute judgments of loudness. Journal of Experimental Psychology, 1953, 46, 373-380.

Garner, W. R., & Hake, H. W. The amount of information in absolute judgments. Psychological Review, 1951, 58, 446-459.

Green, D. M., & Luce, R. D. Speed-accuracy tradeoff in auditory detection. In S. Kornblum (Ed.), *Attention and performance, IV.* New York: Academic Press, 1973. Pp. 547-569.

Holland, M. K., & Lockhead, G. R. Sequential effects in absolute judgments of loudness. Perception & Psychophysics, 1968, 3, 409-414.

Luce, R. D., & Green, D. M. A neural timing theory for response times and the psychophysics of intensity. Psychological Review, 1972, 79, 14-57.

Luce, R. D., & Green, D. M. The response ratio hypothesis for magnitude estimation. Journal of Mathematical Psychology, 1974, 11, in press.

Luce, R. D., & Mo, S. S. Magnitude estimation of heaviness and loudness by individual subjects: A test of a probabilistic response theory. British Journal of Mathematical & Statistical Psychology, 1965, 18, 159-174.

McGill, W. J. Stochastic latency mechanism. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology.* Vol. I. New York: Wiley, 1963. Pp. 309-360.

McGill, W. J. Neural counting mechanisms and energy detection in audition. Journal of Mathematical Psychology, 1967, 4, 351-376.

Pollack, I. Identification and discrimination of components of elementary auditory displays. Journal of the Acoustical Society of America, 1956, 28, 906-909.

Pradhan, P. L., & Hoffman, P. J. Effect of spacing and range of stimuli on magnitude estimation judgments. Journal of Experimental Psychology, 1963, 66, 533-541.
Schneider, B., & Lane, H. Ratio scales, category scales, and variability in the production of loudness and softness. Journal of the Acoustical Society of America, 1963, 35, 1953-1961.
Siebert, W. M. Some implications of the stochastic behavior of primary auditory neurons. Kybernetic, 1965, 2, 206-215.
Siebert, W. M. Stimulus transformations in the peripheral auditory system. In P. A. Kolers and M. Eden (Eds.), *Recognizing patterns*. Cambridge, Mass: M.I.T. Press, 1968. Pp. 104-133.
Siebert, W. M. Frequency discrimination in auditory system: Place or periodicity mechanisms? Proceedings of the IEEE, 1970, 58, 723-730.
Stevens, S. S. On the psychophysical law. Psychological Review, 1957, 64, 153-181.
Stevens, S. S. On the validity of the loudness scale. Journal of the Acoustical Society of America, 1959, 31, 995-1003.
Stevens, S. S. Toward a resolution of the Fechner-Thurstone legacy. Psychometrika, 1961a, 26, 35-47.
Stevens, S. S. To honor Fechner and repeal his law. Science, 1961b, 133, 80-86.
Stevens, S. S. Issues in psychophysical measurement. Psychological Review, 1971, 78, 426-450.
Stevens, J. C., & Guirao, M. Individual loudness functions. Journal of the Acoustical Society of America, 1964, 36, 2210-2213.
Swets, J. A., & Green, D. M. Sequential observations by human observers of signals in noise. In C. Cherry (Ed.), *Information theory*. London: Butterworths, 1961. Pp. 177-195. (Reprinted in J. A. Swets (Ed.), *Signal detection and recognition by human observers*. New York: Wiley, 1964. Pp. 221-242.)
Swets, J. A., Shipley, E. F., McKey, M. J., & Green, D. M. Multiple observations of signals in noise. Journal of the Acoustical Society of America, 1959, 31, 514-521. (Reprinted in J. A. Swets (Ed.), *Signal detection and recognition by human observers*. New York: Wiley, 1964. Pp. 201-220.)
Ward, L. M. Category judgments of loudness in the absence of an experimental-induced identification function: Sequential effects and power functions fit. Journal of Experimental Psychology, 1972, 94, 179-187.
Ward, L. M. Repeated magnitude estimations with a variable standard: Sequential effects and other properties. Perception & Psychophysics, 1973, 13, 193-200.
Ward, L. M., & Lockhead, G. R. Sequential effects and memory in category judgments. Journal of Experimental Psychology, 1970, 84, 27-34.
Ward, L. M., & Lockhead, G. R. Response system processes in absolute judgment. Perception & Psychophysics, 1971, 9, 73-78.

NOTE

1. Of course, other approaches to these data are possible. One is the counting model. We do not study this for two reasons. First, we purposely used signals of long duration (1 sec) so that the observers would not be driven to counting behavior from, what we believe to be, their natural tendency to use a timing rule (Green & Luce, 1973). Second, we have been uanble to derive neat formulas for the distribution of ratios under the counting hypothesis. Another avenue is the generalized detectability analysis of Braida and Durlach (1972). Basically, they approach the problem as one in the framework of signal detectability in which the variance of the signal representation is the sum of two terms, one of which is attributed to the signal and the other of which grows as the square of the range in decibels of signals used. We are not sympathetic with that hypothesis and are pursuing a different approach here and elsewhere. As will be seen below, we posit an attentional mechanism, which affects the sample size and so the variability of the representation, as an alternative mechanism. It has the advantage of being operative immediately before the observer has any experience with the range or distribution of signals being used.

(Received for publication September 17, 1973;
revision accepted October 22, 1973.)