# Factors producing and factors not producing time errors: An experiment with loudness comparisons

ÅKE HELLSTRÖM
*University of Stockholm, Stockholm, Sweden*

Pairs of 1-sec, 1,000-Hz tones, with interstimulus intervals of 1.5 sec, were judged by 60 subjects in categories of "louder," "softer," and "equal." The judgments referred to the first tone in the pair for half of the subjects and to the second tone for the other half. Perceived loudness differences were scaled by a Thurstonian method. The SPL of the standard tone alternated between 50 and 70 dB in one experimental series and between 30 and 50 dB in the other. Time errors (TEs) were consistently positive (first tone overestimated relative to second) at the lower SPL and negative at the higher SPL. This "classical" effect of stimulus level on TE was thus shown to depend upon the relative, rather than the absolute, level of stimulation. The judgment mode was of very little consequence, which strongly contradicts TE theories that emphasize response-bias effects. The quantitative results are interpreted in terms of a general successive-comparison model employing the concepts of adaptation and differential weighting of sensation magnitudes.

For well over a century, psychophysical experiments involving comparisons between successive stimuli, separated by a time interval, have been a source of speculation and dispute. The reason is that such comparisons are almost always subject to systematic asymmetries, so that two physically equal stimuli are not reported as such and the point of subjective equality (PSE, the physical magnitude $\phi_v$ of a variable stimulus that is judged equal to the standard stimulus) is displaced away from $\phi_s$, the physical magnitude of the standard. This "error" was first described by Fechner (1860) and was named the *time error* [TE; sometimes the term *time order error* (TOE) is used]. By Fechner's definition, the TE is positive or negative if the first stimulus is over- or underestimated, respectively, relative to the second. If, as has traditionally been the case, the standard is presented before the variable, the TE measure in physical units is equivalent to that of any constant error (CE):

$$CE = PSE - \phi_s. \tag{1}$$

It is, however, also possible to measure the TE in subjective units, as discussed in the next section.

Numerous investigations have provided evidence that the TE is dependent upon several stimulus factors, particularly the length of the interstimulus interval (ISI) and the level of stimulus magnitude. Most often, TEs have been found to be negative for high levels of stimulus intensity and positive (or less negative) for low levels (e.g., Bartlett, 1939; Lauenstein, 1933; Needham, 1935; Woodrow, 1933). Several theoretical explanations have been suggested for the TE and its dependence on various conditions (for reviews see, e.g., Guilford, 1954, chap. 12; Mikkonen, 1969; Needham, 1934b; Peak, 1940; Woodworth & Schlosberg, 1954, chap. 8; Hellström, Note 1). However, the matter has never been settled, and for a long time the most popular attitude towards the problem has been to take it for granted that TEs are simply response-bias[1] effects, such as response preferences or shifted response criteria (e.g., Engen, 1971, pp. 31-32; Galanter, 1962; Luce & Galanter, 1963, pp. 224-232; Restle, 1961, pp. 157ff, 187), or artifacts of the scaling method (Stevens, 1957b, 1975, chap. 5). John's (1975) invocation of implicit verbal responses to the absolute level of stimulation, hypothesized to influence the comparative judgment, is a somewhat more sophisticated variation on the response-bias theme.

Hellström (1977) presented experimental results indicating that, for comparisons of tone durations, TEs are virtually independent of the mode of judging and responding, and are therefore most probably of perceptual origin. Jamieson and Petrusic (1975) arrived at the same conclusions on the basis of their own duration comparison data. Woodruff, Jennings,

and Rico (1975) found judgment mode not to be a factor of importance in influencing the TE for lifted weights.

The main aim of the present study was to collect supplementary evidence from the continuum of tone loudness. In the experiment, the order of presentation of standard and variable and the judgment mode (first or second tone judged in relation to the other) were varied independently, different groups of subjects being used for the different conditions. The set of available responses was the same for all groups. If TEs arise mainly from a simple response bias, like the built-in greater availability of the response "louder" as compared to "softer," which was proposed by Restle (1961), their sign and magnitude should depend only on which tone is explicity judged, and also be independent of the loudness level of the pair; if they are mediated by implicit verbal responses ("loud," "soft") to the absolute level of stimulation, as proposed by John (1975), they should depend on both level and judgment mode; that is, the level effect should go in the usual direction when the second tone is judged relative to the first, but the effect should be reversed when the first tone is judged in relation to the second. If, however, the TEs are of a genuinely perceptual nature, the level effect should show up in the usual manner independently of the judgment mode. A further aim of the experiment was to investigate the effect of the order in which stimuli are presented, that is, the differential effects of varying the first tone [order: variable-standard (VS)] as opposed to the second [order: standard-variable (SV)]. Also, in order to investigate whether the effect is dependent on the absolute or relative level of stimulation, one group of subjects judged tones that were 20 dB lower in amplitude than those judged by the other groups (presentation order in this extra group SV only).

A fairly low number of judgments were made by each subject. This was judged necessary because practice effects might otherwise have seriously affected the results (Needham, 1934a; cf. Hellström, Note 1, Note 2). Consequently, many of the analyses had to be made on the pooled group data.

In the next section various quantitative TE measures, to be used in the analysis of the experimental results, are discussed. In the concluding section, an attempt is made to incorporate the TE effects in a comprehensive quantitative model for the comparison of successive stimuli.

## ON THE MEASUREMENT OF TIME ERRORS

As indicated above, one conventional way of measuring the TE is to compute CE, that is, the physical difference between PSE and $\phi_s$ (Equation 1). However, this measure is rather arbitrary. An estimate of the magnitude of the effect in *subjective* units should carry more theoretical meaning.

Thurstone's (1927a, 1927b) law of comparative judgment can be adapted to the three-category paired-comparison task (see Hellström, 1977; Sjöberg, 1967; cf. Glenn & David, 1960; Greenberg, 1955; Olson & Ogilvie, 1972) for the measurement of subjective differences within pairs of stimuli. The basic ideas and assumptions of this scaling method are similar to those underlying Torgerson's (1958, chap. 10) law of categorical judgment. For any pair of stimuli ($S_1$, $S_2$) there exists a distribution of momentary subjective differences $d = r_1 - r_2$, where $r_1$ and $r_2$ are the momentary subjective magnitudes of $S_1$ and $S_2$, respectively. In the case of stimuli that are presented successively, the possible judgments (which can of course be expressed in different ways, see below) are symbolized as "1" (first stimulus greater), "2" (second stimulus greater), and " = " (stimuli equal). To decide between these options, the subject uses two criteria for nonequality, with momentary positions $t_1$ and $t_2$ on the subjective difference continuum, so that he will respond "1" whenever $d > t_1$, "2" whenever $d < t_2$, and " = " whenever $t_1 > d > t_2$.

The criterion positions $t_1$ and $t_2$ are further assumed to be normally distributed with expected values symmetric about a response-bias parameter b, that is $E(t_1) = T + b$ and $E(t_2) = -T + b$, and with identical variances $\sigma_{t_1}^2 = \sigma_{t_2}^2 = \sigma_t^2$. The momentary difference, d, is likewise assumed to be normally distributed with expected value $E(d) = D$. The expression $(d - t_1) = (r_1 - r_2 - t_1)$ is denoted by $c_1$, and $(d - t_2) = (r_1 - r_2 - t_2)$ by $c_2$. The following expression for the variance of $c_1$ then follows:

$$\sigma_{c_1}^2 = \sigma_{r_1}^2 + \sigma_{r_2}^2 - 2\,Cov(r_1, r_2) + \sigma_t^2$$
$$+ 2[Cov(r_2, t_1) - Cov(r_1, t_1)]. \quad (2)$$

The expression for $\sigma_{c_2}^2$ is similar. On the assumption that the last term (which should be close to zero) is equal in these two expressions, we have $\sigma_{c_1}^2 = \sigma_{c_2}^2 = \sigma_c^2$. The corresponding standard deviation $\sigma_c$ can be termed *comparatal dispersion* (Gulliksen, 1958).

Denoting by $Z_1$ and $Z_2$ the standard normal deviates which correspond to the probability $P_1$ for a judgment to fall in category "1," and the probability $P_2$ for it to fall in category "2" [i.e., $Z_1 = \Phi^{-1}(P_1)$ and $Z_2 = \Phi^{-1}(P_1 + P_2)$], we may note that

$$Z_1 = (D - T - b)/\sigma_c \quad (3)$$

and that

$$Z_2 = (D + T - b)/\sigma_c. \quad (4)$$

Thus,

$$Z_2 + Z_1 = 2(D - b)/\sigma_c \qquad (5)$$

and

$$Z_2 - Z_1 = 2T/\sigma_c. \qquad (6)$$

Setting $b = 0$ (see below), Equations 5 and 6 yield expressions for $D/T$, $\sigma_c/T$, and $D/\sigma_c$. Consistent estimates of these population measures can be obtained by substituting, for $Z_1$ and $Z_2$, the standard normal deviates $z_1$ and $z_2$, which correspond to the sample proportions $p_1$ and $p_2$ of judgments in categories "1" and ("1" or " = "), respectively:

*Mean subjective difference (D) in units of mean difference criterion (T):*

$$(D/T)^* = (z_2 + z_1)/(z_2 - z_1); \qquad (7)$$

*Comparatal dispersion ($\sigma_c$) in units of mean difference criterion (T):*

$$(\sigma_c/T)^* = 2/(z_2 - z_1); \qquad (8)$$

*Mean subjective difference (D) in units of comparatal dispersion ($\sigma_c$):*

$$(D/\sigma_c)^* = (z_2 + z_1)/2. \qquad (9)$$

As discussed in Hellström (1977), the expected half-width, T, of the "equal" category can be assumed to be constant over different conditions, such as, for example, levels of stimulation, for the same subjects in the same experiment (as long as motivation and expectancy factors do not change). On the other hand, the comparatal dispersions for different stimulus levels (in the present experiment, 20 dB apart) should be markedly different according to the massive body of evidence indicating that the discriminal dispersion, $\sigma_r$, increases with the level of stimulation, which has been formulated as Ekman's law (Ekman, 1956, 1959; cf. Eisler, 1965). Thus, the conventional assumption of constant comparatal dispersions (e.g., Glenn & David, 1960; Greenberg, 1965; Olson & Ogilvie, 1972) would almost certainly fail in cases where the stimulus level is varied. However, $(D/\sigma_c)^*$ is a more robust measure than $(D/T)^*$ and because of this technical advantage it will be used here as a complement to the latter measure.

In an experiment with the traditional design of the method of constant stimulus differences, the TE can now be estimated in subjective units as the value of the subjective difference which obtains when $\phi_1 =$

$\phi_2$, as estimated from the plot of $(D/T)^*$ or $(D/\sigma_c)^*$ against the physical magnitudes (or standardized subjective magnitude values such as sones) of the variable stimulus. Likewise, the PSE value is the x-intercept in the plot of the subjective differences against the physical magnitudes of the variable stimuli, that is, the $\phi_v$ value which yields $(D/T)^*$ [or, $(D/\sigma_c)^*$] $= 0$.

When one standard stimulus is compared with several variable stimuli which are spaced evenly and symmetrically on the subjective continuum, so that the total distribution of d for all stimulus pairs can be considered nearly rectangular, "shortcut" estimates of the series mean, $\bar{D}$, of the subjective difference, and thus of the TE, can be derived by substituting for the z values in Equations 7-9 the corresponding relative or absolute frequencies of judgments, as follows (see Hellström, 1977):

*TE in units of mean difference criterion (T):*

$$(\bar{D}/T)^* = (p_1 - p_2)/p_= = (n_1 - n_2)/n_=, \qquad (10)$$

where $p_1$, $p_2$, and $p_=$; $n_1$, $n_2$, and $n_=$ are the total proportions and numbers, respectively, of judgments in categories "1," "2," and " = " for all stimulus pairs;

*TE in units of range ($R_d$) of total distribution of d:*

$$(\bar{D}/R_d)^* = (p_1 - p_2)/2$$
$$= (n_1 - n_2)/(n_1 + n_2 + n_=); \qquad (11)$$

*Range ($R_d$) of subjective differences in units of mean difference criterion (T):*

$$(R_d/T)^* = 2/p_= = 2(n_1 + n_2 + n_=)/n_=. \qquad (12)$$

## METHOD

### Subjects

Sixty undergraduate students of psychology (30 men and 30 women, median age 25 years) took part in the experiment to fulfill a course requirement. The subjects were free from known hearing defects, and were not aware of the purpose of the experiment.

### Apparatus

A Philips PP 6050 oscillator generated a 1,000-Hz sinusoidal signal. The signal was fed into a two-channel gating unit connected to a four-channel timing device which controlled the stimulus durations and the length of the ISI. The rise time of the gates was approximately 20 msec and the fade-out time approximately 4 msec. This arrangement practically eliminated disturbing clicks. The gated signals from the two channels were fed—via two Marconi TF 2162 .1-dB step attenuators, a set of high-quality transformers (for impedance matching), and a 300-3,000-Hz band-pass filter (to eliminate possible transients and noise)—into a

**Table 1**
**Results from 50/70 Groups: Total Number of Judgments in Each Category; Computed**
**Shortcut Measures for Each Presentation Order, Standard, and Judgment Mode**

| | Presentation Order SV | | | | | | Presentation Order VS | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 50-dB SPL Standard | | | 70-dB SPL Standard | | | 50-dB SPL Standard | | | 70-dB SPL Standard | | |
| | Judgment Mode | | | Judgment Mode | | | Judgment Mode | | | Judgment Mode | | |
| | 1LS | 2LS | Total | 1LS | 2LS | Total | 1LS | 2LS | Total | 1LS | 2LS | Total |
| "First louder" | 255 | 294 | 549 | 211 | 199 | 410 | 238 | 219 | 457 | 151 | 137 | 288 |
| "Equal" | 270 | 201 | 471 | 138 | 105 | 243 | 326 | 332 | 658 | 214 | 171 | 385 |
| "Second louder" | 68 | 96 | 164 | 242 | 288 | 530 | 32 | 38 | 70 | 232 | 284 | 516 |
| "Uncertain" | 7 | 9 | 16 | 9 | 8 | 17 | 4 | 11 | 15 | 3 | 8 | 11 |
| $(\bar{D}/T)^*$ | +.69 | +.99 | +.82 | −.22 | −.85 | −.49 | +.63 | +.55 | +.59 | −.38 | −.86 | −.59 |
| $(\bar{D}/R_d)^*$ | +.16 | +.17 | +.16 | −.03 | −.08 | −.05 | +.17 | +.15 | +.16 | −.07 | −.12 | −.10 |
| $(R_d/T)^*$ | 4.39 | 5.88 | 5.03 | 8.57 | 11.28 | 9.74 | 3.66 | 3.55 | 3.60 | 5.58 | 6.92 | 6.18 |

*Note−1LS = "First tone louder or softer?"; 2LS = "Second tone louder or softer?"*

calibrated Telephonics TDH-39 headphone set with Grason-Stadler rubber cushions. The signal fed into the headphones was monitored by a Brüel & Kjaer 2603A microphone amplifier serving as a vacuum tube voltmeter.

The experiment was carried out in an ordinary laboratory room, reasonably well shielded from outside noise. The noise level in the room was about 35 dB (A). The listening mode was binaural. The subject was seated at a table where he could not see either the apparatus or the experimenter.

**Procedure**

**Experimental design.** Three sound pressure levels (SPLs) of the standard tones were used: 30, 50, and 70 dB re 20 μPa. Two standard tones, 20 dB apart, alternated in each of three experimental groups of 20 subjects (10 men and 10 women). In two groups (the 50/70 groups), the standard levels were 50 and 70 dB, in the third group (the 30/50 group), they were 30 and 50 dB. Each standard tone was combined with a series of five variable tones differing by −4, −2, 0, +2, and +4 dB from the standard. In the 30/50 group, the order of presentation was standard-variable (SV); in one of the 50/70 groups, the order was likewise SV, in the other, variable-standard (VS). Each group was split into two subgroups of 10 subjects (five men and five women); those in one subgroup were instructed to judge whether the *first* tone was the louder or the softer in the pair (judgment mode 1LS), those in the other, whether the *second* tone was the louder or the softer (judgment mode 2LS). The standard was set in alternate trials at the group's higher and lower SPL, respectively. The experiment was run with one subject at a time and in one session for each subject. Twelve blocks of 10 pairs were presented, 5 pairs at each level, the variable stimuli being alternated in a random fashion. New random orders were used for each subject and each block. The experimental pairs were preceded by two

practice pairs of identical tones (i.e., $\phi_v = \phi_s$), one pair at each standard level. For half the subjects in each subgroup, the first practice pair was at the louder of the group's two standard levels, and for the other half, it was at the softer level. The frequency of all tones was 1,000 Hz and the duration 1.0 sec. The ISI was 1.5 sec and the intertrial interval approximately 8 sec. The session took about 25 min (instruction not included).

**Registration of judgments.** The subjects wrote down their judgments on small ticket-blocks with numbered pages, one judgment on each page. (This arrangement was designed to minimize the interdependence of successive judgments as well as any tendency by the subject to equalize the proportions of the various judgments.) Judgment categories available to the subjects were " + " for *louder*, " − " for *softer*, " = " for *equal*, and "?" for *completely uncertain*. As explained above, the judgments referred to the first and second tone in judgments modes 1LS and 2LS, respectively. The subjects were informed that most of the time there would be an objective difference between the two tones in a pair, but they were encouraged to disregard this fact and to respond in a naive manner. The "?" judgments were excluded in all computations of scale values.

## RESULTS AND DISCUSSION

For each presentation order, standard, and judgment mode in the 50/70 groups, the distribution of judgments, collapsed across all variable stimuli, is given in Table 1 together with computed shortcut values of $(\bar{D}/T)^*$, $(\bar{D}/R_d)^*$, and $(R_d/T)^*$ (Equations 10-12). Table 2 gives the corresponding results from

**Table 2**
**Results from 30/50 Group: Total Number of Judgments in Each Category;**
**Computed Shortcut Measures for Each Standard and Judgment Mode**

| | 30-dB SPL Standard | | | 50-dB SPL Standard | | |
|---|---|---|---|---|---|---|
| | Judgment Mode | | | Judgment Mode | | |
| | 1LS | 2LS | Total | 1LS | 2LS | Total |
| "First louder" | 299 | 331 | 630 | 142 | 166 | 308 |
| "Equal" | 228 | 179 | 407 | 144 | 127 | 271 |
| "Second louder" | 61 | 74 | 135 | 310 | 301 | 611 |
| "Uncertain" | 12 | 16 | 28 | 4 | 6 | 10 |
| $(\bar{D}/T)^*$ | +1.04 | +1.44 | +1.22 | −1.17 | −1.06 | −1.12 |
| $(\bar{D}/R_d)^*$ | + .20 | + .22 | + .21 | − .14 | − .11 | − .13 |
| $(R_d/T)^*$ | 5.16 | 6.53 | 5.76 | 8.28 | 9.35 | 8.78 |

*Note−1LS = "First tone louder or softer?"; 2LS = "Second tone louder or softer?"*

the 30/50 group (presentation order SV only). It is easily seen that the TE in both the 50/70 groups and the 30/50 group is positive for the softer series and negative for the louder series of the respective group, irrespective of the judgment mode. This is the effect of stimulus level on TE which is ordinarily obtained. The fact that the TE is positive for the 50-dB series in the 50/70 groups but negative for the same series in the 30/50 group shows that it is the relative, rather than absolute, level of stimulation which determines the direction of the TE. Shortcut values of $(\bar{D}/R_d)^*$ (Equation 11) in each series were computed for the individual subjects. The Mann-Whitney U test was then performed separately for each group, series, and presentation order. In no case did the effect of the judgment mode (1LS vs. 2LS) approach significance.

In the derivation of the D measures, the response-bias parameter b was set to zero. TE theories which rest on the notion of "simple" response bias predict that the subject will use the response "louder" more often than "softer" (or the reverse) in pairs of tones which are experienced as equally loud. This implies a b value which is negative (or positive) in judgment mode 1LS and positive (or negative) in 2LS, and of about the same absolute magnitude in both modes. These theories thus predict biased D estimates, yielding TEs of similar size but opposite sign in the two judgment modes. An effect of the stimulus level on the TE is not compatible with this kind of theory. John's (1975) implicit-response theory predicts that b varies with stimulus level so as to produce a level effect on the TE that goes in opposite directions in judgment modes 1LS and 2LS. The present results are obviously irreconcilable with this theory as well as with those mentioned before. Still another possible interpretation of "response bias" could be that b has a fixed negative (or positive) value, with little or no variation between judgment modes. The subject would then tend to judge the first (or second) member of a pair of equally loud tones as being the louder, regardless of the judgment mode he is instructed to use. This is strongly contradicted by the present results. The only possible explanation for these results along response-bias lines is that the subject tends to judge the first or second member, respectively, of a pair of loud or soft tones as being the louder, regardless of the judgment mode. This requires a b value whose absolute value is near constant, while its sign varies both with stimulus level and with judgment mode so as to transform experienced within-pair differences of zero (for pairs of equally loud tones) into biased D values which generate the obtained pattern of negative and positive TEs for pairs of loud and soft tones, respectively. While—here as always—some kind of response-bias notion, conceived as a sufficiently complicated set of decision rules, can be invoked and defended, an explana-

tion which is clearly more parsimonious, and to me seems much more plausible, is that the obtained TEs result primarily from the action of mechanisms which affect the *experienced* within-pair differences, the true value of b being close to zero throughout stimulus levels and judgment modes. This notion will be pursued in the following, and the mechanisms—though possibly involving memory distortion—will, for want of a better term, be called *perceptual*.

In Figures 1-4, the mean subjective difference, D, as estimated in units of T (Equation 7) and of $\sigma_c$ (Equation 9), is plotted against the SPL of the variable tone for each group, series, presentation order, and judgment mode. As can be seen, there are no essential differences between the results for the two judgment modes. For both the 50/70 groups and the 30/50 group, the plots for (D/T)* are steeper for the louder series of the respective group. This result is in line with the widely accepted notion that the psychophysical loudness function obeys Stevens' (1955, 1957a, 1957b) power law rather than Fechner's (1860) logarithm law. While the latter states that the subjective magnitude is linearly related to a logarithmic stimulus measure like dB SPL, the power law implies that this relationship is an exponential function, and thus positively accelerated.

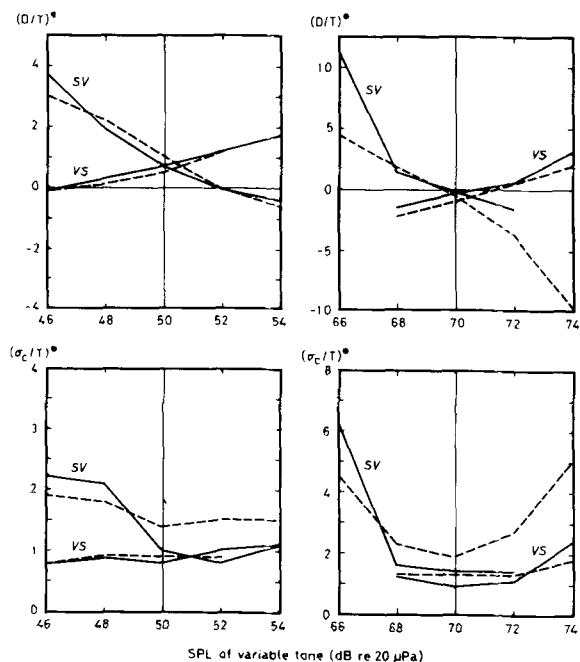In addition to the subjective difference measures, Figures 1 and 3 show the estimated comparatal dis-



Figure 1. 50/70 groups: Judgment modes 1LS (continuous lines) and 2LS (dashed lines). Mean subjective loudness difference (D) and comparatal dispersion ($\sigma_c$), as estimated in units of mean criterion difference (T), vs. SPL of variable tone for standard tone 50 dB SPL (left panel) and 70 dB SPL (right panel) in each presentation order (SV, VS).
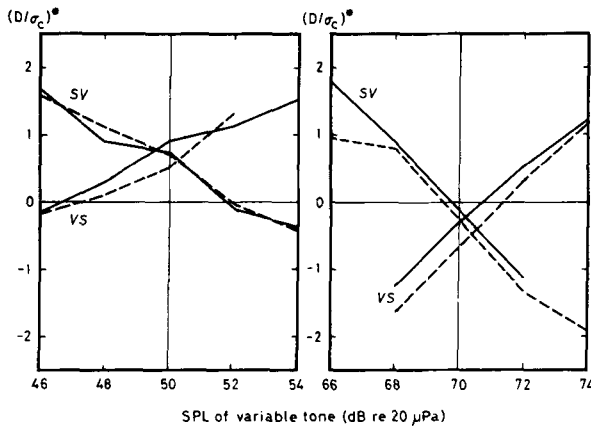
**Figure 2. 50/70 groups: Judgment modes 1LS (continuous lines) and 2LS (dashed lines). Mean subjective loudness difference (D) as estimated in units of comparatal dispersion ($\sigma_c$), vs. SPL of variable tone for standard tone 50 dB SPL (left panel) and 70 dB SPL (right panel) in each presentation order (SV, VS).**
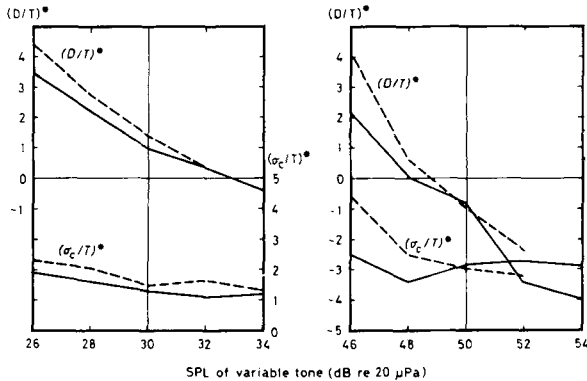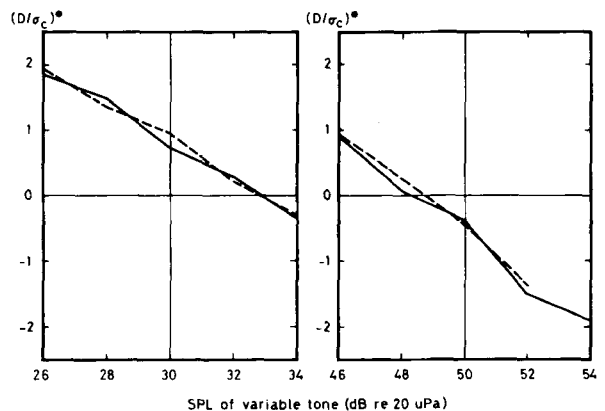


**Figure 3. 30/50 group: Judgment modes 1LS (continuous lines) and 2LS (dashed lines). Mean subjective loudness difference (D) and comparatal dispersion ($\sigma_c$) as estimated in units of mean criterion difference (T), vs. SPL of variable tone for standard tone 30 dB SPL (left panel) and 50 dB SPL (right panel) in presentation order SV.**

persions, $(\sigma_c/T)^*$ (Equation 8), for the pooled data from both judgment modes. In each group, these dispersions are larger for the louder than for the softer series, as may be expected from Ekman's law (see above). However, the average comparatal dispersion for the 30-dB series in the 30/50 group is about the same as for the 50-dB series in the 50/70 group, and for the 50-dB series in the 30/50 group it is about the same as for the 70-dB series in the 50/70 group with the same presentation order (SV). Furthermore, the plot of $(D/T)^*$ against $\phi_V$ in dB SPL is much steeper for the 50-dB series in the 30/50 group than for the same series in the 50/70 groups. These findings suggest that the unit of measurement, T, that is, the mean size of the subjective difference criterion, can be regarded as invariant between the

louder and softer series of each group, but is larger for the 50/70 groups than for the 30/50 group. This suggests that the size of T is related to the average subjective within-pair, and possibly also between-pair, difference. Both should be much larger for the 50/70 groups than for the 30/50 group because a given difference in dB SPL should be subjectively larger the higher the SPL (see above).

For both the 50-dB and the 70-dB series in the 50/70 groups, the plots in Figure 1 of $(D/T)^*$ against $\phi_V$ in dB SPL are steeper for the SV presentation order (descending lines) than for the VS order (ascending lines). The estimated comparatal dispersions are also seen to be related to the presentation order, those for the SV order being consistently larger than those for the VS order. A crucial question is whether these findings imply that the unit of measurement, T, is different for the different orders (see below).

The increase in the estimated comparatal dispersions for large subjective differences, which is easily noted for the 70-dB series, is in line with the results of Sjöberg (1963, 1965, 1968a, 1968b, 1969). As discussed in Hellström (1977), it may depend on a decrease in the correlation between the momentary subjective magnitudes of the two stimuli and hence on a decrease in the first covariance term in Equation 2. However, at least part of the effect may have a different explanation. Figure 5 shows, for the pooled data from both judgment modes in the 70-dB series (presentation order SV), the standard normal deviates $z_1$ and $z_2$, as defined above, plotted against $\phi_V$ in dB SPL. The plots are S-shaped, and the $z_1$ and $z_2$ plots approach each other in the ends. This suggests that the "innermost" tails of subjective difference distributions with extreme means are too long compared with the normal distribution, that is, that these



**Figure 4. 30/50 group: Judgment modes 1LS (continuous lines) and 2LS (dashed lines). Mean subjective loudness difference (D), as estimated in units of comparatal dispersion ($\sigma_c$), vs. SPL of variable tone for standard tone 30 dB SPL (left panel) and 50 dB SPL (right panel) in presentation order SV.**

distributions are skewed with their longer tails pointing towards zero, resulting in reduced values of the extreme normal deviates. The net result of this is that the estimates of (D/T) are exaggerated for large subjective differences. What is computed is the mean (and hence, median) and standard deviation of that normal distribution which would yield the empirical proportions of judgments in the various categories. If the actual distribution is skewed, with its longer tail pointing inward, the estimated mean should at least approximate the true median, whereas the abso-, lute magnitude of the mean is overestimated as well as the dispersion.

Estimating the subjective differences in units of $\sigma_c$, that is, computing $(D/\sigma_c)*$ (Equation 9) may be one way of coping with the skewedness problem. However, because the estimate is simply the mean of the two z values, the resulting plots display an S-shaped distortion in the "conservative" direction (see Figures 2 and 4). However, this distortion is more moderate than that affecting the $(D/T)*$ values, that is, the $(D/\sigma_c)*$ measure is more robust against nonnormality of the subjective difference distribution. Also, it is more resistant to random fluctuations of small response proportions (cf. Hellström, 1977), and may thus have some technical advantages, especially in model fitting, which is dealt with in the next section.

## TWO SENSATION WEIGHTING MODELS

The traditional concept of the TE as a "constant error, which was adhered to above" is inherently rather limited and static, as shown, for example, by the consistent effect of the stimulus level, which makes the "error" anything but constant. It can therefore be concluded that the TE, defined either as the physical difference between two subjectively equal stimuli or as the subjective difference between two physically equal stimuli, is only one manifestation of a general distortion of the subjective differences between successive stimuli which is too complicated to be reported as a single number. As we may also conclude that these differences can be estimated reasonably accurately by the measures discussed above, the estimates will be employed in this section to develop a more dynamic and comprehensive model for the comparison of successive stimuli, a model which will then include the TE effects.

In Hellström (1977), two "sensation weighting" models were developed which could well account for the TE effects in a duration comparison experiment similar in design to the present one. The basic model, *Model 1*, is:

$$d_{ik} = r_{1ik} - r_{2ik} = W_{1k}\psi_{1i} - W_{2k}\psi_{2i} + U_k, \quad (13)$$

where $d_{ik}$ is the momentary subjective difference



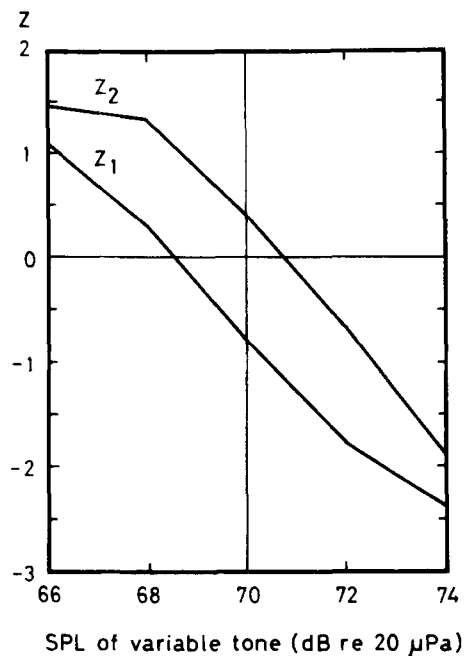SPL of variable tone (dB re 20 μPa)

Figure 5. 50/70 groups, 70-dB series, pooled data from judgment modes 1LS and 2LS: Standard normal deviates $z_1$ and $z_2$ corresponding to proportions of judgments "1" and ("1" or "=") , respectively, vs. SPL of variable tone in presentation order SV.

between the stimuli in pair i, presented in experimental condition k, $\psi_{1i}$ and $\psi_{2i}$ are the momentary magnitudes of the sensations evoked by the first and second stimulus, respectively, and $r_{1ik}$ and $r_{2ik}$ are the corresponding "weighted" subjective magnitudes. $W_{1k}$ and $W_{2k}$ are the weighting coefficients, specific to condition k, of $\psi_{1i}$ and $\psi_{2i}$, respectively, in producing $r_{1ik}$ and $r_{2ik}$, and thus in determining the value of $d_{ik}$. $U_k$ is an additive parameter required to render the mean value of $d_{ik}$ independent of the magnitudes of $W_{1k}$ and $W_{2k}$. A scale factor, common to all conditions, is contained in $W_{1k}$, $W_{2k}$, and $U_k$.

We shall assume (cf. Stevens, 1957b) that the psychophysical function for the compared stimuli is a power function,

$$E(\psi) = \alpha\phi^\beta. \quad (14)$$

$\alpha$ and $\beta$ are assumed to be independent of temporal position in the pair and constant over all experimental conditions with a given set of standards. Combining Equations 13 and 14, and setting the $\alpha$ values to 1, we arrive at *Model 1a*:

$$D_{ik}/T_k = B_{1k}\phi_{1i}{}^\beta - B_{2k}\phi_{2i}{}^\beta + C_k, \quad (15)$$

where $T_k$ is the expected value of the criterion difference for condition k, $B_{1k} = W_{1k}/T_k$, $B_{2k} = W_{2k}/T_k$, and $C_k = U_k/T_k$.

The parameters in Model 1a (the common exponent $\beta$, plus the three weighting coefficients $B_{1k}$, $B_{2k}$, and $C_k$ for each condition) were estimated by a computer program which employed the iterative function-fitting routine STEPIT (Chandler, 1969). The program minimized the sum total of squared residuals, that is, the quantity

$$S_{tot} = \sum_{k=1}^{n} \sum_{i=1}^{m_k} \left[ (D_{ik}/T_k)^* - (D_{ik}/T_k)' \right]^2 , \quad (16)$$

where $(D_{ik}/T_k)'$ is the scale value predicted by the model for pair i in condition k, n is the number of conditions, and $m_k$ is the number of pairs for which empirical scale values have been obtained in condition k. The goodness of the fit was evaluated by dividing the final value of $S_{tot}$ by the sum of the squared deviations of the empirical scale values from their total mean $M_{tot_{(D/T)^*}}$ and multiplying by 100, that is, by forming the quantity

$$V\%_{tot} = \frac{100 S_{tot}}{\sum_{k=1}^{n} \sum_{i=1}^{m_k} \left[ (D_{ik}/T_k)^* - M_{tot_{(D/T)^*}} \right]^2} . \quad (17)$$

$V\%_{tot}$ can be interpreted as the percentage of the total empirical variance that is left unexplained by the model. For the individual conditions (groups), the goodness of the fit was evaluated by computing the analogous quantity

$$V\%_k = \frac{100 \sum_{i=1}^{m_k} \left[ (D_{ik}/T_k)^* - (D_{ik}/T_k) \right]^2}{\sum_{i=1}^{m_k} \left[ (D_{ik}/T_k)^* - M_{k_{(D/T)^*}} \right]^2} , \quad (18)$$

where $M_{k_{(D/T)^*}}$ denotes the mean empirical scale value in condition k. (For the case with only one condition, $V\%_k = V\%_{tot}$.)

In the manner described above, Model 1a was fitted to the empirical scale values from the 50/70 groups (simultaneously using the values from presentation orders SV and VS) and the 30/50 group, both with judgment modes 1LS and 2LS treated as two separate conditions with an assumed common exponent $\beta$ and with the data from these two modes pooled. The $\phi$ values were expressed in Pa = N/m². The results are given in Table 3.

Table 3
Results of Fitting Model 1a (Equation 15) to Mean Subjective Loudness Differences (D)
as Estimated in Units of Mean Criterion Difference (T) and Comparatal Dispersion ($\sigma_c$)

| | 50/70 Groups 50- and 70-dB SPL Standards | | | | 30/50 Group 30 - and 50-dB SPL Standards | | | |
|---|---|---|---|---|---|---|---|---|
| | Presentation Order SV and VS | | | | Presentation Order SV | | | |
| | Subjective Difference Estimate | | | | Subjective Difference Estimate | | | |
| | $(D/T)^*$ | | $(D/\sigma_c)^*$ | | $(D/T)^*$ | | $(D/\sigma_c)^*$ | |
| Percentage of total variance unexplained | 23.6 | | 3.87 | | 3.31 | | .747 | |
| $\beta$ | .460 | | .233 | | .217 | | .114 | |
| Judgment Mode | 1LS | 2LS | 1LS | 2LS | 1LS | 2LS | 1LS | 2LS |
| Percentage of variance unexplained | 43.3 | 8.55 | 3.61 | 4.28 | 2.51 | 4.68 | .970 | .429 |
| $B_1$ | 81.8 | 79.5 | 28.5 | 26.2 | 76.6 | 106.6 | 38.6 | 40.9 |
| $B_2$ | 87.3 | 93.6 | 32,6 | 31.1 | 95.4 | 124.2 | 49.1 | 51.4 |
| $C$ | 1.67 | 2.49 | 1.94 | 2.12 | 5.14 | 5.30 | 5.34 | 5.36 |
| $B_1/B_2$ | .937 | .849 | .873 | .844 | .804 | .858 | .786 | .796 |
| $a = C/(B_2 - B_1)$ | .304 | .176 | .468 | .436 | .274 | .300 | .509 | .512 |
| $SPL_a$ (dB) | 71.5 | 61.2 | 65.7 | 63.1 | 42.1 | 45.6 | 42.4 | 42.8 |
| Empirical subjective midpoint (dB SPL) | 62.6 | | 61.4 | | 41.3 | | 40.7 | |
| Judgment Mode | 1LS + 2LS | | 1LS + 2LS | | 1LS + 2LS | | 1LS + 2LS | |
| Percentage of variance unexplained | 14.35 | | 1.74 | | 4.57 | | .389 | |
| $\beta$ | .520 | | .229 | | .106 | | .109 | |
| $B_1$ | 82.3 | | 27.5 | | 82.9 | | 40.1 | |
| $B_2$ | 93.6 | | 32.1 | | 99.6 | | 50.6 | |
| $C$ | 1.87 | | 2.10 | | 9.13 | | 5.53 | |
| $B_1/B_2$ | .879 | | .856 | | .832 | | .792 | |
| $a = C/(B_2 - B_1)$ | .165 | | .455 | | .545 | | .526 | |
| $SPL_a$ (dB) | 63.9 | | 64.1 | | 44.4 | | 42.6 | |
| Empirical subjective midpoint (dB SPL) | 62.9 | | 61.4 | | 40.6 | | 40.6 | |
| Subjected midpoint computed via sone scale (dB SPL) | 63.4 | | | | 43.4 | | | |

Note—1LS = "First tone louder or softer?"; 2LS = "Second tone louder or softer?"

In Hellström (1977), it was shown that, making reasonable assumptions, if Model 1a holds for (D/T)* it should also give a satisfactory fit for $(D/\sigma_c)^*$ in cases like the present one where the within-series differences are small compared with the between-series differences. Though the values of the parameters will of course be different for the two sets of D estimates, the ratios $B_{1k}/B_{2k}$, which mirror the weighting balance, should not change appreciably. As can be seen in Table 3, the fit for $(D/\sigma_c)^*$ is far superior to that for (D/T)*. The fit for (D/T)* in judgment mode 1LS in the 50/70 group is poor but would presumably improve if one excluded the extreme stimuli, as the corresponding scale values are probably distorted. The ratios $B_{1k}/B_{2k}$, as predicted, do not change much between the corresponding fits for the two measures. The fact that these ratios are clearly and consistently below 1 agrees well with the results of Hellström (1977) for durations.

In Hellström (1977), it was found that for each of the four judgment modes employed there, the additive parameter $C_k$ was nearly proportional to the difference between $B_{2k}$ and $B_{1k}$:

$$C_k \approx a(B_{2k} - B_{1k}),    (19)$$

so that a in Equation 19 was nearly invariant over judgment modes while $B_{1k}$ and $B_{2k}$ varied somewhat (their ratio remaining nearly constant). In the present experiment, the a values were likewise quite stable over judgment modes (see Table 3). Combining Equations 15 and 19 and using Equation 14 and the definitions of $B_{1k}$, $B_{2k}$, and $C_k$, we get, after rearrangement of the terms, *Model 2*:

$$d_{ik} = W_{1k}(\psi_{1i} - a) - W_{2k}(\psi_{2i} - a).    (20)$$

The constant a may be conceived of as the subjective counterpart of the adaptation level (AL; see Helson, 1965) and can thus be converted, via the inverse of the assumed psychophysical power function (Equation 14), into a physical AL value which will be called $\phi_a$. As $\alpha$ was set to 1, we have

$$\phi_a = a^{1/\beta}.    (21)$$

For the present experiment, the $\phi_a$ values should, for convenience, be converted into values of dB SPL (denoted by $SPL_a$):

$$SPL_a = 20 \log_{10}(\phi_a/20),    (22)$$

where $\phi_a$ is given in $\mu$Pa = $\mu$N/m² (reference level for SPL is 20 $\mu$Pa).

The computed values of $SPL_a$ are given in Table 3. The table also includes two estimates of the subjec-

tive midpoint of the total set of stimuli presented to the respective group. The first estimate was calculated by taking the arithmetic mean of the fitted subjective magnitudes of the standard (weighted five times) and variable tones and converting this value back into decibels via the fitted psychophysical power function. The second estimate was made by taking, instead, the weighted arithmetic mean of the loudnesses in sones and converting this value back into decibels via the sone function. [According to I.S.O. Recommendation ISO/R 131-1959, the loudness in sones is $2^{(P-40)/10}$, where P is the loudness level in phons, which at 1,000 Hz is equal to the SPL in dB re 20 $\mu$Pa. The sone function is equivalent to a power function of the sound pressure with an exponent of 2 $\log_{10} 2$ = .6021 (cf. Stevens, 1955).] The second estimate is seen to agree better with the fitted $SPL_a$. The fitted psychophysical power functions have exponents which are lower than that of the sone function, and of course also lower than the value .67 reported by Stevens (1975, p. 15) as typical. This discrepancy is likely to depend in part on differing properties of the present scale as compared to fractionation scales, and so it may prove to be of some consequence for the quantitative interpretation of the results. However, $\beta$ is not an important parameter for the goodness of fit, and so it cannot be estimated with any great accuracy. Also, the fitted $\beta$ value is of very little importance for the relations between the other parameters. For the $(D/\sigma_c)^*$ measure, the low exponents are due to the fact that the scale unit, $\sigma_c$, increases with the subjective magnitude.

Equation 20 indicates that a, and thus $\phi_a$, can be regarded as constant, while $B_{1k}$ and $B_{2k}$ vary with the conditions. This means that the number of parameters needed to describe the data can be reduced by replacing the $C_k$s in Model 1a by a single parameter that reflects $\phi_a$. Combining Equations 14 and 20 yields *Model 2a*:

$$D_{ik} = B_{1k}(\phi_{1i}^\beta - \phi_a^\beta) - B_{2k}(\phi_{2i}^\beta - \phi_a^\beta).    (23)$$

This model was fitted to the (D/T)* values by the same computer program as was used for fitting Model 1a. As before, the sum total of squared residuals, $S_{tot}$ (Equation 16), was minimized. However, as Model 2a does not contain an additive parameter like $C_k$ in Model 1a, the sum total of the residuals, as well as their sums for the individual conditions, will not necessarily become zero, as was the case in Model 1a. In accordance with common practice, the goodness of fit was therefore evaluated by using as the denominator in $V\%_{tot}$ the sum total of the squared deviations from zero instead of from the total mean:

$$V\%_{tot} = \frac{100 S_{tot}}{\sum_{k=1}^{n} \sum_{i=1}^{m_k} \left[ (D_{ik}/T_k)^* \right]^2}. \qquad (24)$$

Analogously, the goodness of the fit for the individual groups was evaluated by computing

$$V\%_k = \frac{100 \sum_{i=1}^{m_k} \left[ (D_k/T_k)^* - (D_{ik}/T_k)' \right]^2}{\sum_{i=1}^{m_k} \left[ (D_{ik}/T_k)^* \right]^2}. \qquad (25)$$

In this way Model 2a was fitted to the empirical scale values from the 50/70 and 30/50 groups, Groups 1LS and 2LS being treated as two separate conditions with assumedly equal $\beta$ and $\phi_a$ values. The results are given in Table 4 (the estimated values of $\phi_a$ have been converted into dB SPL by Equation 22). Compared with Model 1a, the fit is very slightly impaired. (The *reduction* in the percentage of unexplained variance which takes place in one case is due to the different definitions of these percentages in Equations 24 and 25 as compared to Equations 17 and 18.)

As in the duration experiment (Hellström, 1977), the data analyses thus indicate that the effect of stimulus level on TE can be explained as the result of a differential weighting of the sensory information from the first and second stimulus—specifically, of their subjective distances from the AL, with a heavier weighting of the distance corresponding to the second tone than of that corresponding to the first. The weighting effect can also be discerned at each stimulus level separately in the 50/70 group for the (D/T)* measure, in the form of steeper plots against

$\phi_v$ (in dB SPL) in the SV than in the VS presentation order (see Figure 1). The same effect was noted for lifted weights by Michels and Helson (1954; Helson, 1964, chap. 4). It is obviously consistent with the notion of a larger weighting coefficient for the second stimulus than for the first. For the $(D/\sigma_c)^*$ measure, however, the slopes of the two plots at each level are nearly equal. The estimated comparatal dispersions, $(\sigma_c/T)^*$, are larger in the SV than in the VS presentation order, and this compensates for the weighting effect. One explanation could be that the scale unit, T, is different for the two presentation orders, so that the "weighting" effect is only apparent and disappears when the units are equal. The level effect on the TE would then have to depend on some other factor that works only *between* series, that is, with stimuli that are as far apart as 20 dB, but not with the much smaller intensity variation *within* each series. The different values of $(\sigma_c/T)^*$ for the two presentation orders may also just mean that the corresponding "genotypical" comparatal dispersions are different. As the discriminal dispersion of a stimulus can probably be expected to be independent of whether the stimulus is designated a standard or a variable, the differences in the estimated dispersions are then likely to depend on the criterion variances being greater in the SV than in the VS presentation order (cf. Wickelgren, 1968).

A pilot study was undertaken which may give at least a partial answer to the question of the nature of the level effect on the TE. A brief report of this study follows.

## AN EXPLORATORY STUDY WITH SMALL INTERSERIES DIFFERENCES

In order to study whether the level effect on the

**Table 4**
**Results of Fitting Model 2a (Equation 23) to Mean Subjective Loudness Differences (D)**
**as Estimated in Units of Mean Criterion Difference (T) and Comparatal Dispersion ($\sigma_c$)**

| | 50/70 Groups 50- and 70-dB SPL Standards | | | | 30/50 Group 30 - and 50-dB SPL Standards | | | |
|---|---|---|---|---|---|---|---|---|
| | Presentation Order SV and VS | | | | Presentation Order SV | | | |
| | Subjective Difference Estimate | | | | Subjective Difference Estimate | | | |
| | (D/T)* | | $(D/\sigma_c)^*$ | | (D/T)* | | $(D/\sigma_c)^*$ | |
| Percentage of total variance unexplained | 24.8 | | 3.95 | | 3.96 | | .731 | |
| $\beta$ | .475 | | .240 | | .232 | | .116 | |
| Judgment Mode | 1LS | 2LS | 1LS | 2LS | 1LS | 2LS | 1LS | 2LS |
| Percentage of variance unexplained | 39.5 | 9.29 | 3.27 | 4.61 | 3.34 | 4.57 | .982 | .387 |
| $B_1$ | 89.0 | 79.8 | 28.9 | 26.0 | 79.1 | 111.5 | 38.4 | 40.8 |
| $B_2$ | 96.2 | 94.1 | 33.4 | 30.7 | 97.5 | 131.1 | 48.8 | 51.3 |
| $B_1/B_2$ | .925 | .848 | .867 | .847 | .811 | .850 | .787 | .796 |
| $SPL_a$ (dB) | 63.2 | | 64.2 | | 43.8 | | 42.6 | |

*Note—1LS = "First tone louder or softer?"; 2LS = "Second tone louder or softer?"*

Table 5
Results from Study with Small Interseries Differences (Presentation Order SV, Judgment Mode "Which was louder?"): Number
of Judgments in Each Category and Computed Shortcut TE Measures for Each Subject and for the Whole Group

| | 66-dB SPL Standard | | | | 70-dB SPL Standard | | | | |
| | Number of Judgments in Category | | | | Number of Judgments in Category | | | | TE $[(\overline{D}/T)*]$ |
| Subject | 1 | = | 2 | TE $[(\overline{D}/T)*]$ | 1 | = | 2 | TE $[(\overline{D}/T)*]$ | Difference |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 24 | 30 | 6 | + .600 | 22 | 21 | 17 | + .238 | + .362 |
| 2 | 29 | 19 | 12 | + .895 | 21 | 16 | 23 | − .125 | +1.020 |
| 3 | 23 | 35 | 2 | + .600 | 20 | 27 | 13 | + .259 | + .341 |
| 4 | 32 | 15 | 13 | +1.267 | 27 | 6 | 27 | .000 | +1.267 |
| 5 | 13 | 40 | 7 | + .150 | 14 | 29 | 17 | − .103 | + .253 |
| 6 | 26 | 17 | 17 | + .529 | 19 | 12 | 29 | − .833 | +1.363 |
| Total | 147 | 156 | 57 | | 123 | 111 | 126 | | |
| Mean | | | | + .673 | | | | − .094 | + .768 |
| | | | Whole Group | | | | | | |
| $(\overline{D}/T)*$ | | | | + .577 | | | | − .027 | + .604 |
| $(\overline{D}/R_d)*$ | | | | + .125 | | | | − .004 | + .129 |
| $(R_d/T)*$ | | | | 4.615 | | | | 6.486 | |

TE can be demonstrated even when the between-series difference in stimulus level is only of the same order of magnitude as the within-series differences, six subjects performed a comparison task that was identical with that of the main study, except that the SPLs of the standard tones were 66 and 70 dB. The variable tones, as before, differed from the respective standard tones by −4, −2, 0, +2, and +4 dB. The presentation order was SV. The judgment mode used—"which tone was the louder?" (cf. Hellström, 1977, Note 2)—was considered to be more "neutral" than those used in the main study. Written responses available to the subjects were "1," "2," and "=." No "?" responses were thus allowed. Shortcut TE values, $(\overline{D}/T)*$ (Equation 10), were calculated for the individual subjects. The results are given in Table 5. As can be seen, all six differences between the TE values for the 66-dB series and the 70-dB series were positive. A randomization test yields p = .015625 for the hypothesis that the population value of the difference in TE between the series is zero. It can thus be concluded that large between-series differences in stimulus level are not essential to the level effect on the TE.

## CONCLUSION

The outcome of the pilot study reinforces the conclusion from the main experiment that the level effect, and the entire TE phenomenon, is due to differential efficiency of the information received from two stimuli presented in succession. To study these effects in more detail and avoid the ambiguity of interpretation which results from possibly unequal scale units with different standard-variable presentation orders, the standard-variable paradigm has to

be abandoned in favor of a design where *both* stimuli are varied at the same time. This has been done in a further loudness-comparison study (Hellström, Note 2), the results of which corroborate the adequacy of the sensation-weighting explanation. In this study, the quantitative models are also further developed.

## REFERENCE NOTES

1. Hellström, A. *On the nature of the time-error.* Reports from the Department of Psychology, University of Stockholm, 1977, Suppl. 38.
2. Hellström, A. *Differential sensation weighting as the basic cause of time-errors.* Reports from the Department of Psychology, University of Stockholm, 1977, No. 498.

## REFERENCES

BARTLETT, R. J. Measurement in psychology. *Advancement of Science,* 1939, 1, 422-441.
CHANDLER, J. P. STEPIT—Finds local minima of a smooth function of several parameters. *Behavioral Science,* 1969, 14, 81-82. (CPA 312)
EISLER, H. The connection between magnitude and category scales and direct and indirect scaling methods. *Psychometrika,* 1965, 30, 271-289.
EKMAN, G. Discriminal sensitivity on the subjective continuum. *Acta Psychologica,* 1956, 12, 15-18.
EKMAN, G. Weber's law and related functions. *Journal of Psychology,* 1959, 47, 343-352.
ENGEN, T. Psychophysics. 1. Discrimination and detection. In J. W. Kling & L. A. Riggs (Eds.), *Woodworth & Schlosberg's experimental psychology* (3rd ed.). New York: Holt, 1971.
FECHNER, G. T. *Elemente der Psychophysik* [Elements of psychophysics]. Leipzig, Germany: Breitkopf & Härtel, 1860.
GALANTER, E. Contemporary psychophysics. In *New directions in psychology.* New York: Holt, 1962.
GLENN, W. A., & DAVID, H. A. Ties in paired-comparison experiments using a modified Thurstone-Mosteller model. *Biometrics,* 1960, 16, 86-109.
GREENBERG, M. G. A modification of Thurstone's law of comparative judgment to accommodate a judgment category of

"equal" or "no difference." *Psychological Bulletin*, 1965, **64**, 108-112.

GUILFORD, J. P. *Psychometric methods* (2nd ed.). New York: McGraw-Hill, 1954.

GULLIKSEN, H. Comparatal disperson, a measure of accuracy of judgment. *Psychometrika*, 1958, **23**, 137-150.

HELLSTRÖM, Å. Time errors are perceptual. An experimental investigation of duration and a quantitative successive-comparison model. *Psychological Research*, 1977, **39**, 345-388.

HELSON, H. *Adaptation-level theory*. New York: Harper & Row, 1964.

JAMIESON, D. G., & PETRUSIC, W. M. Presentation order effects in duration discrimination. *Perception & Psychophysics*, 1975, **17**, 197-202.

JESTEADT, W., & BILGER, R. C. Intensity and frequency discrimination in one- and two-interval paradigms. *Journal of the Acoustical Society of America*, 1974, **55**, 1266-1276.

JOHN, I. D. A common mechanism mediating the time-order error and the cross-over effect in comparative judgments of loudness. *Australian Journal of Psychology*, 1975, **27**, 51-60.

LAUENSTEIN, O. Ansatz zu einer physiologischen Theorie des Vergleichs und der Zeitfehler. [Attempt at a physiological theory of comparison and of the time-error]. *Psychologische Forschung*, 1933, **17**, 130-177.

LUCE, R. D., & GALANTER, E. Discrimination. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1). New York: Wiley, 1963.

MICHELS, W. C., & HELSON, H. A quantitative theory of time-order effects. *American Journal of Psychology*, 1954, **67**, 327-334.

MIKKONEN, V. On the retention of perceptual quantities. *Commentationes Humanarum Litterarum, Societas Scientiarum Fennica*, 1969, **44**, Whole No. 3.

NEEDHAM, J. G. The time error as a function of continued experimentation. *American Journal of Psychology*, 1934, **46**, 558-567. (a)

NEEDHAM, J. G. The time error in comparison judgments. *Psychological Bulletin*, 1934, **31**, 229-243. (b)

NEEDHAM, J. G. The effect of the time-interval upon the time-error at different intensive levels. *Journal of Experimental Psychology*, 1935, **18**, 530-543.

OLSON, C. L., & OGILVIE, J. C. The method of constant stimuli with two or more categories of response. *Journal of Mathematical Psychology*, 1972, **9**, 320-338.

PEAK, H. The time order error in successive judgment and in reflexes: III. Time error theories. *Psychological Review*, 1940, **47**, 1-20.

RESTLE, F. *Psychology of judgment and choice*. New York: Wiley, 1961.

SJÖBERG, L. An empirical application of a new case of the law of comparative judgment. *Scandinavian Journal of Psychology*, 1963, **4**, 97-107.

SJÖBERG, L. A study of four methods for scaling paired comparisons data. *Scandinavian Journal of Psychology*, 1965, **6**, 173-185.

SJÖBERG, L. Successive intervals scaling of paired comparisons. *Psychometrika*, 1967, **32**, 297-308.

SJÖBERG, L. The dimensionality paradox in comparative judgment: A resolution. *Scandinavian Journal of Psychology*, 1968, **9**, 97-108. (a)

SJÖBERG, L. Unidimensional scaling of multidimensional facial expressions. *Journal of Experimental Psychology*, 1968, **78**, 429-435. (b)

SJÖBERG, L. The dimensionality of similarity judgments. *American Journal of Psychology*, 1969, **82**, 441-456.

STEVENS, S. S. The measurement of loudness. *Journal of the Acoustical Society of America*, 1955, **27**, 815-829.

STEVENS, S. S. Concerning the form of the loudness function. *Journal of the Acoustical Society of America*, 1957, **29**, 603-606. (a)

STEVENS, S. S. On the psychophysical law. *Psychological Review*, 1957, **64**, 153-181. (b)

STEVENS, S. S. *Psychophysics* (ed. by G. Stevens). New York: Wiley, 1975.

THURSTONE, L. L. A law of comparative judgment. *Psychological Review*, 1927, **34**, 273-286. (a)

THURSTONE, L. L. Psychophysical analysis. *American Journal of Psychology*, 1927, **38**, 368-389. (b)

TORGERSON, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.

WICKELGREN, W. A. Unidimensional strength theory and component analysis of noise in absolute and comparative judgments. *Journal of Mathematical Psychology*, 1968, **5**, 102-122.

WOODROW, H. Weight discrimination with a varying standard. *American Journal of Psychology*, 1933, **45**, 391-416.

WOODRUFF, B., JENNINGS, D. L., & RICO, N. L. Time-error in lifted weights as affected by presentation order and judgment mode. *Perception & Psychophysics*, 1975, **18**, 98-104.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology* (Rev. ed.). New York: Holt, 1954.

## NOTE

1. The term *response bias* is used here, as by the authors cited, in its literal sense, that is, to denote an effect which operates in the response system and affects the *judgments* but not the experiences that are to be reported. However, some authors have tended to use the term to denote any deviation of a judgment from veridicality or expectation, regardless of the nature of the effect. For instance, Jesteadt and Bilger (1974) write, in discussing the results of their pitch and loudness comparison experiments: "The pattern of these *response biases* suggests that the *remembered value* of the first-interval tone has a tendency to shift towards the center of the context" (p. 1274; italics mine). This use of terms is definitely not conducive to understanding the nature of the TE phenomenon. Indiscriminate use of the term "response bias" probably helps divert interest from the search for underlying mechanisms. In the cited case, and many others, the term "bias" would be much more adequate if it were not qualified by "response."