

Comparison is not just subtraction: Effects of time- and space-order on subjective stimulus difference

ÅKE HELLSTRÖM

Stockholm University, Stockholm, Sweden

In five experiments, participants made comparative judgments of paired successive or simultaneous stimuli. Time- or space-order errors were obtained, which varied with the interstimulus interval (ISI) or stimulus duration, as well as with the stimulus level. The results, in terms of scaled subjective differences, are well described by Hellström's (1979) sensation-weighting model. With successive presentation, in comparisons of line length and tone loudness, the first stimulus had the greater weight in determining the subjective difference for short ISIs, the second for longer ISIs. In comparisons of duration (auditory and visual), the second stimulus had the greater weight. For simultaneously presented line lengths, the left stimulus had the greater weight.

Ordinal comparisons between paired stimuli, presented successively or simultaneously, are the raw material from which the discriminability of the stimuli is determined. Fechner (1860) discovered that, in determinations of the just noticeable difference (JND), the temporal or spatial presentation order of the stimuli was important: Systematic time-order errors (TOEs) or space-order errors (SOEs) kept entering the results. In the method of constant stimuli, the standard stimulus (*St*) is held constant while the comparison stimulus (*Co*) is varied in fixed steps. Also determined are the point of subjective equality (PSE) and the constant error (CE; $CE = PSE - St$). The TOE and SOE are special cases of the CE: A negative TOE (or SOE) means that the first (or left) stimulus is underestimated relative to the second.

Whereas there has been little research on the SOE, the causes of the TOE have been much debated (see Hellström, 1985). Many theorists have highlighted the often found general negative TOE, explaining it as some kind of response bias (e.g., Allan, 1977; Link, 1992; Restle, 1961; Stevens, 1957). Others, considering that the TOE is often negative with longer interstimulus intervals (ISIs) and positive (or less negative) with shorter ISIs, have conceived of some kind of retention loss (e.g., Boas, 1882; Köhler, 1923). A third group of explanations has emerged from findings—particularly in roving-standard experiments—of TOEs that are more negative the higher the stimulus magnitude level (e.g., Needham, 1935; Woodrow, 1933). These explanations vary the response bias theme, assuming an influence of the absolute level

on the relative judgment (e.g., “the tones are loud/soft” → “the tone to be judged (usually the second one) is the louder/softer” [John, 1975]) or an inference from the second stimulus to the worse-remembered first one (e.g., “the second tone is loud/soft” → “the first tone was [probably] softer/louder” [Masin, 1993]).

Much of the TOE debate seems to be due to an oversimplified view of the comparison process. For instance, one assumption underlying the concepts of the CE and of the JND is that in determining them, it should not matter which of the stimuli, the first or the second, is held constant and which is varied. This is a clear case of falling into the “physicalistic trap” (Eisler, 1982): In the use of a balance, laying 1 g in one scale has the same effect as removing 1 g from the other. However, in subjective comparisons of two successively lifted weights, this is generally not the case: The effect of a change by 1 g depends on which of the stimuli is changed (cf. Hellström, 2000; Michels & Helson, 1954). (It may help to think of a balance with levers differing in length.) Furthermore, as was mentioned above, particularly in roving-standard experiments, the direction and the size of the TOE depend on the stimulus magnitude level, so that one single TOE measure, such as the mean CE, gives incomplete information.

Hellström (1977, 1978) found that the response mode (e.g., which of the stimuli is judged in relation to the other) was of no great importance and concluded that the TOE must be a perceptual phenomenon, and not due to response processes (cf. Jamieson & Petrusic, 1975a, 1975b, 1976, 1978; Schab & Crowder, 1988; for a fuller discussion, see Hellström, 1985). For comparisons of tonal loudness, Hellström (1979) then found, using a factorial design, that variation of the first and of the second stimulus, respectively, had different degrees of impact on the response. This pattern depended on ISI and stimulus duration; using a short duration increased the effect of

This research was supported by the Swedish Research Council. The author thanks Jonathan Lindström and Joseph Tekle for carrying out the experiments and William Petrusic and an anonymous reviewer for valuable comments on the manuscript. Address correspondence to Å. Hellström, Department of Psychology, Stockholm University, SE-106 91 Stockholm, Sweden (e-mail: hellst@psychology.su.se).

the ISI. Detailed analysis of the data showed that the TOE could be explained as a side effect of *sensation weighting* (SW): The subjective difference, d , between the two compared stimuli represents the difference, not between their sensation magnitudes per se, but between two weighted compounds, one for each stimulus, where the sensation magnitudes of the stimulus and of the current adaptation level (AL; Helson, 1964) enter with weights s and $(1 - s)$, respectively (cf. Michels & Helson, 1954). (In the present study, as in Hellström, 2000, the more neutral term *reference level* [ReL] is used instead of AL):

$$d = k \{ [s_1 \psi_1 + (1 - s_1) \psi_{r1}] - [s_2 \psi_2 + (1 - s_2) \psi_{r2}] \} + b, \quad (1)$$

where d is the scaled subjective difference between the compared stimuli and k is a scale constant, which reflects the ability and willingness to report a difference between the two stimuli. ψ_1 and ψ_2 are the sensation magnitudes of the stimuli, s_1 and s_2 are weighting coefficients, and ψ_{r1} and ψ_{r2} are the subjective magnitudes that correspond to the current ReLs (those may be different for the two stimuli; cf. Hellström, 1979). b is a constant term, which is here added to the model to capture a possible contribution to d that is independent of the weighting mechanism—for instance, a judgmental bias.

A common finding with ISIs of a couple of seconds has been that the TOE is generally on the negative side and becomes more negative the higher the magnitude level of the compared stimuli in the series, being positive only for stimuli of low magnitude. This can be explained by defining the TOE as the value of d in a pair of identical stimuli (see Hellström, 1985) and by considering, in Equation 1, such pairs by setting $\psi_1 = \psi_2 = \psi$ [thereby assuming that the stimuli obey the same psychophysical function, $\psi = f(\phi)$, where ϕ is the physical stimulus magnitude]. Simplifying by assuming $\psi_{r1} = \psi_{r2} = \psi_r$ and $b = 0$, we obtain

$$(\text{TOE} =) d = k (s_1 - s_2) (\psi - \psi_r). \quad (2)$$

According to Equation 2, then, the inverse relation between the TOE and the level of ψ is a consequence of the fact that $s_1 < s_2$; the negative mean TOE may be due to this, in combination with the ReL being below the average series stimulus in magnitude (i.e., $\psi - \psi_r > 0$; cf. Hellström, 2000); another possibility, in reference to Equation 1, is that $\psi_{r1} < \psi_{r2}$ (for $s_1 < 1$ and $s_2 < 1$).

In Hellström (1979), the following preliminary SW model was first fitted to the empirical scale values of stimulus difference from a loudness comparison experiment with 16 temporal conditions:

$$d = W_1 \psi_1 - W_2 \psi_2 + U, \quad (3)$$

where k is a scale constant, $W_1 = k s_1$, and $W_2 = k s_2$. Relating the terms to those of Equation 1 yields

$$U = W_2 \psi_{r2} - W_1 \psi_{r1} + k (\psi_{r1} - \psi_{r2}) + b. \quad (4)$$

From the empirical relationship across conditions between the fitted values of W_1 , W_2 , and U , Hellström (1979)

obtained preliminary evidence for Equation 1 (without the b term), which was then fitted. The s values and their relation underwent dramatic changes with the ISI and tone duration. The ISI exerted its greatest effect for the shortest duration (100 msec). For this duration, the s values generally decreased with the ISI, the effect on s_2 being the greatest, so that for the longer ISIs the usual relation $s_1 < s_2$ was found, but for the shorter, $s_1 > s_2$. In accordance with Equation 2, this corresponded to a negative and a positive relation, respectively, between stimulus magnitude level and TOE.

Hellström (1985) proposed that the differential SW is an adaptation to the fact that different amounts of information about the two stimuli are available at the moment of comparison; later (Hellström, 1986, 1989a, 1989b), he suggested that the purpose of the weighting is to adjust the perceived stimulus difference d so as to maximize the signal-to-noise ratio (S/N) of a change in d , caused by a change in magnitude of one of the compared stimuli.

The present study serves multiple purposes.

1. In Hellström's (1979) study of tonal loudness comparisons, one of the most interesting results was that, for short stimuli and short ISIs, the first stimulus was weighted more heavily than the second. This, contrary to all previous findings, yields a negative TOE for softer tones and a positive TOE for louder tones. In the present study, an attempt was made to reproduce this result, using a different method, and also to carry out a parallel experiment with visual stimuli (line length) under similar temporal conditions.

2. As was suggested by Hellström (1985), there are reasons to believe that the SOE, which occurs with simultaneous stimuli (presented to the left and the right), can be described by the same basic model as the TOE and has a similar explanation. In the SOE case, the differential availability of stimulus information is probably caused by lateralized processing and/or attentional scanning, giving an advantage to one or the other stimulus. For line length, one would expect an advantage for the left line, which, like other visuospatial information, is processed mainly in the right hemisphere (Beaton, 1986; Masin & Agostini, 1991). Another purpose of the present study was to test these predictions experimentally.

3. On the basis of the theoretical position outlined above, successful attempts have been made to employ computerized tests with comparisons of successive stimuli in clinical diagnostics of cerebral dysfunction. Hellström (1990), Hellström and Almkvist (1997), and Hellström, Forssell, and Fernaeus (1989) demonstrated, in patients with senile dementia of the Alzheimer type, a greatly reduced capability to discriminate tone durations—more so for ISIs of 2–4 sec than for those of 0.5–1 sec; this was suggestive of accelerated forgetting and cognitive deterioration. Almkvist, Hellström, Hårdemark, and Johansson (1996) found that one long-term effect of subarachnoidal hemorrhage was an apparent elevation of the estimated value of the reference level in patients performing a visual duration comparison task; this elevation

was related to the severity of the brain damage. Thus, one purpose of the present study was to test more rigorously the applicability of the model for auditory and visual duration and also to obtain baseline data from healthy young participants performing the same tasks.

The present article thus describes a study with five experiments requiring the participants to compare paired successive or simultaneous stimuli in a design in which the magnitudes of both stimuli were varied.

METHOD

Participants

Eighty-five first-year university students of psychology (ages, 19–50 years; median age, about 25 years) took part in the experiments in partial fulfilment of a course requirement. None reported any significant deficits of vision or hearing. The number of participants in each experiment is given in Table 2.

General Procedure

The participant was seated in a quiet laboratory room with soft background light, at a table with a Commodore Amiga 1000 micro-computer, with his/her eyes at a comfortable viewing distance (about 45 cm) from a Commodore 1081 color display screen (275 mm wide, 205 mm high) with a fine mesh filter, which efficiently reduced light reflexes. A white paper screen concealed the edges of the computer screen. Each participant took part in one session, which included from one to five experiments in irregular order.¹

The computer program was written in AmigaBASIC and employed its standard graphics and sound functions to present the stimuli via the screen and the built-in loudspeaker. The instructions for each experiment appeared on the screen and were clarified by the experimenter when necessary. The experiments were self-paced. For the visual task with simultaneous lines, an orange fixation cross (9 mm wide, 17 mm high) appeared in the center of the screen 3 sec before the presentation of the stimuli and remained for 2 sec. The participant was to watch or listen to the two stimuli (without moving the gaze from the position of the fixation cross, if one was used), to compare them, and to indicate which of them was the longer, louder, and so on by using the right hand to press a key on the numeric (right) part of the keyboard. The participant pressed "1" for *first* or *left*, "2" for *second* or *right*, and "0" for *cannot decide*, and then "Enter" or "Return." With "Backspace," the response

could be corrected before it was entered. Each experiment took about 9–18 min (see Table 2) and was followed by a break of a few minutes before the next began.

In each experiment, 96 stimulus pairs were presented. These consisted of four sets of 24 pairs, each set with its separate ISI (for successive stimuli) or stimulus duration (for simultaneous stimuli). For each experiment, the same stimulus magnitude combinations were used for the pairs in each set. The pairs from all the sets were presented intermixedly in a pseudorandom order that was the same for all the participants.² The first stimulus (with successive presentation) or the left one (with simultaneous presentation) served as the base stimulus (base), and the second served as the Co, except for Experiment 5, in which both stimuli served as the base equally often. (The base/Co designation was not indicated to the participant.) The mean stimulus magnitude (M_S) was preset for each experiment, as well as the number of base levels (N_{base} ; these were equally spaced in physical units and centered around the M_S) and the spread (half-range) of the base in percentage of M_S (Spr_{base}). Also preset were the maximum absolute value (Max_D) of the difference (Co – base) in percentage of base, and the number of values (N_D) of this difference (equally spaced and centered around zero; see Table 1).

Specific Designs

The computation of the stimulus magnitudes is explained in detail for Experiment 1. In Table 1, additional details of the design for each experiment are given.

Experiment 1 (tone duration, successive). The stimuli were 600-Hz tones with a sound pressure level (SPL), as measured from the position of the participant's head, of approximately 85 dB(A). As is indicated by the figures in Table 1, the mean stimulus duration (M_S) was 1,000 msec; the duration of the base (the first tone) varied in four equal steps ($N_{\text{base}} = 4$) in the interval $1,000 \pm 60\%$ ($Spr_{\text{base}} = 60$). Thus, the base durations were 400, 800, 1,200, and 1,600 msec. The maximum difference between Co (the second tone) and the base was $\pm 30\%$ of the base ($\text{Max}_D = 30$); this difference had six equal steps ($N_D = 6$): -30% , -18% , -6% , $+6\%$, $+18\%$, and $+30\%$. Thus, for base = 400 msec, the Co durations were 280, 328, 376, 424, 472, and 520 msec; for base = 800 msec, they were 560, 656, 752, 828, 944, and 1,040 msec; for base = 1,200 msec, they were 840, 984, 1,128, 1,272, 1,416, and 1,560 msec; and for base = 1,600 msec, they were 1,120, 1,312, 1,504, 1,656, 1,888, and 2,080 msec. Each of these 24 stimulus combinations was used once for each of the four sets (with ISI = 500, 1,000, 2,000, and 4,000 msec), making a total of 96 stimulus pairs.

Table 1
Specific Conditions for Each Experiment

Experiment No.	Stimulus Type	M_S	N_{base}	Spr_{base} ($\pm\%$)	Max_D ($\pm\%$)	Duration (msec)	ISI (msec)			
							Set 1	Set 2	Set 3	Set 4
Successive Stimulus Presentation										
1	tone duration	1,000 msec	4	60	6	30	500	1,000	2,000	4,000
2	visual duration	1,000 msec	4	60	6	32	500	1,000	2,000	4,000
3	line length	96 mm	6	60	4	22	100	300	900	2,700
4	tone loudness	75 dB(A)	4	60	6	20	100	300	900	2,700
Simultaneous Stimulus Presentation										
							Duration (msec)			
							Set 1	Set 2	Set 3	Set 4
5	line length	48 mm	3	60	4	13	100	200	400	800

Note— M_S , mean stimulus magnitude; N_{base} , number of levels of base; Spr_{base} , spread (half-range) of base in percentage of M_S ; Max_D , maximum difference between comparison stimulus and base in percentage of base; N_D , number of equally spaced values of this difference; ISI, interstimulus interval.

Experiment 2 (visual duration, successive). The stimuli were orange rectangles, 33 mm wide and 26 mm high, centered on a blue background.

Experiment 3 (line length, successive). The stimuli were horizontal orange lines, 2 mm thick and, on average, 96 mm long, centered on a blue background. The stimulus duration was 100 msec.

Experiment 4 (tone loudness, successive). The stimuli were 600-Hz tones with a duration of 100 msec. Their SPLs, as measured from the position of the participant's head, were approximately 55–85 dB(A). M_s was approximately 75 dB(A) (11.3 sones). The other design parameters, in terms of subjective magnitude, were the same as those for line length. The physical stimulus magnitudes were selected by transforming the desired loudness values by the inverse of an assumed psychophysical power function with an exponent of 0.5.

Experiment 5 (line length, simultaneous). The stimuli were horizontal orange lines, 2 mm thick and, on average, 48 mm long, on a blue background. They were centered 96 mm to each side of the fixation cross.

RESULTS

Basic Results

Table 2 gives, for each experiment, the number of participants and the average time (min:sec) the experiment took.

For scaling the subjective difference d between the first and the second stimulus in each pair, the measure $(D/T)^*$, which was used in Hellström (1977, 1978, 1979) and estimates the mean difference in units of the half-width of the *equal* category, requires more extensive data. Instead, to each response of “1,” “2,” or “0,” a difference value of, in order, +100, -100, or 0 was assigned. When an average is computed, this yields the percent difference ($D\%$), or the difference between the percentages of “first stimulus greater” and “second stimulus greater” (cf. Hellström, 1977, 1985; the term $D\%$ will here be used also for the individual pairs). For two stimuli differing in magnitude, $D\%$ is proportional to the measure—one half the difference between the proportions of correct responses for the two presentation orders—employed by Jamieson and Petrusic (1975a, 1975b). $D\%$ (called D/R_d by Hellström, 1977, 1978, and used there as a “shortcut measure”) was designed to scale

the subjective stimulus difference in units of the range of differences, “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 can be considered nearly rectangular” (Hellström, 1978, p. 435). As for single pairs, d is likely to have a distribution more like the Gaussian (but possibly with oversized tails; see Hellström, 1993); this should lead to a relative under-estimation of large subjective differences. The power transformation described below should help compensate for this.

Stimulus transformation. The relation between the physical stimulus magnitudes and the sensation magnitudes underlying the subjective stimulus difference was assumed to be a power function, $\psi = \phi^m$ (cf. Hellström, 1979). With this substitution, Equation 3 was fitted to the mean $D\%$ values across participants. The value (determined to two decimal places) of m that yielded the highest mean multiple correlation across stimulus sets was selected to be used in the further computations. For each experiment, this m value is given in Table 2, and the regression coefficients (scaled to estimate the s values, using a method that will be described later) are given in Figures 6–10. A linear regression equation, corresponding to Equation 3, was then fitted for each participant, experiment, and set, with $D\%$ for each pair as the dependent variable and the transformed physical magnitudes ϕ_1^m and ϕ_2^m as independent variables.

Table 2 gives the mean and standard deviation for percentage correct [counting “cannot decide” as wrong (a) and as half correct (b)] and the median for each set of the multiple correlation (R) between $D\%$ and the transformed physical stimulus magnitudes. (With 100% correct, R would be, for Experiments 1–5, in order, .87, .86, .89, .87, and .89.) For each combination of ISI and stimulus level, the individual TOE was computed as the participant's mean $D\%$ value for the pertinent stimulus pairs. This value is termed TOE% or (for Experiment 5) SOE%. Figures 1–5 show, for each experiment, the mean TOE% or SOE% plotted against the base magnitude. The corresponding mean TOE% or SOE% for each ISI or (for Experiment 5) stimulus duration is shown in Figures 6–10.

Table 2
Results for Each Experiment: Mean Time (min:sec), M and SD for Percentage Correct [Counting “Cannot Decide” as Wrong (a) and as Half Correct (b)], and Median in Each Stimulus Set for Multiple R Between Responses and Transformed Stimuli

Experiment No.	Stimulus Type	n	Time (M)	% Correct (a)		% Correct (b)		Multiple R (Median) in Set				m Value for Stimulus Transformation
				M	SD	M	SD	1	2	3	4	
Successive Stimulus Presentation												
1	tone duration	32	14:00	71.0	11.2	77.1	5.0	.74	.75	.76	.72	0.19
2	visual duration	35	14:03	63.8	11.8	70.2	9.3	.66	.65	.63	.52	0.35
3	line length	39	10:06	79.4	9.9	84.0	6.0	.82	.83	.82	.78	0.11
4	tone loudness	41	10:12	61.8	14.0	70.4	5.7	.38	.55	.68	.76	0.22
Simultaneous Stimulus Presentation												
5	line length	31	13:03	74.5	10.7	79.8	5.6	.72	.73	.79	.81	0.20

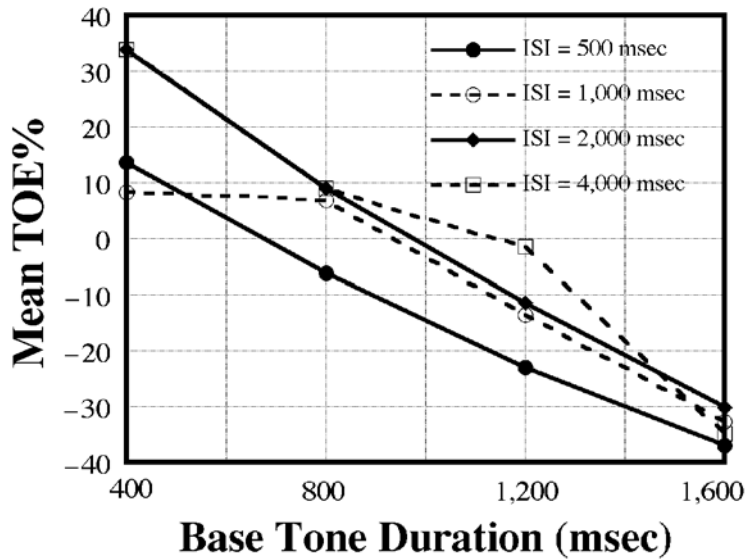


Figure 1. Experiment 1: tone duration. Mean time-order error percentage (TOE%) plotted against base duration for each interstimulus interval (ISI).

The data from the individual participants in each experiment (TOE% or SOE%; fitted W values) were analyzed by analyses of variance (ANOVAs) (SPSS MANOVA with the multivariate approach to repeated measures, using Pillai's tests) testing the effect of the ISI or stimulus duration on the weighting and on TOE% or SOE%. The results are given in Table 3. Alpha levels of .05 are used throughout this article (two-tailed, where applicable; in Tables 3 and 5, p values between .05 and .10 are reported parenthetically).

In order to evaluate the consistency of the stimulus weighting across stimulus levels, the stimulus pairs in each experiment were divided into a high and a low level, with equal numbers of pairs. This was done using the physical magnitude of the base stimulus or, for Experiment 5, the mean stimulus level in the pair. Equation 3 was fitted separately for the high and the low levels. The results were analyzed by ANOVA (same method as that described above), and the relevant results are included in Table 3.

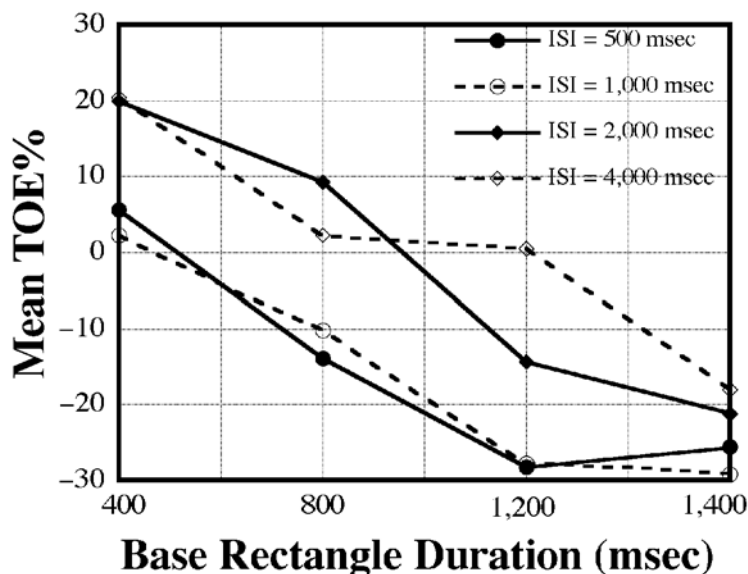


Figure 2. Experiment 2: visual duration. Mean time-order error percentage (TOE%) plotted against base duration for each interstimulus interval (ISI).

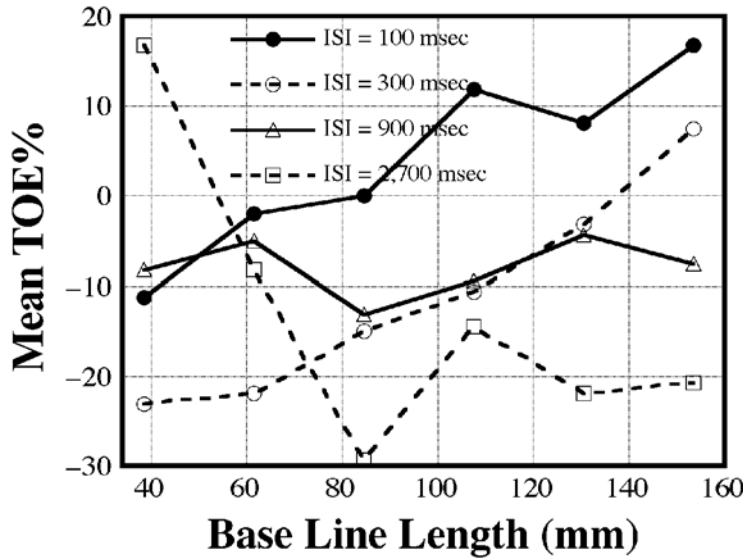


Figure 3. Experiment 3: line length, successive. Mean time-order error percentage (TOE%) plotted against base length for each interstimulus interval (ISI).

According to the SW model, as defined in Equations 1 and 3, for each experiment and temporal condition, the individual U values should be dependent on W_1 and W_2 (or W_{left} and W_{right}), as prescribed by Equation 4. This equation can be rewritten as

$$\begin{aligned}
 U = & (W_2 - W_1) (\psi_{r1} + \psi_{r2})/2 \\
 & + (W_1 + W_2) (\psi_{r2} - \psi_{r1})/2 \\
 & + k (\psi_{r1} - \psi_{r2}) + b.
 \end{aligned}
 \tag{5}$$

We may expect that ψ_{r1} and ψ_{r2} , resulting from the pooling of sensation magnitudes, should be close to each

other and vary little across individuals. If so, U should correlate highly positively with $(W_2 - W_1)$ but should correlate close to zero with $(W_1 + W_2)$. Thus, it should be possible to partition the total variability in $(W_1 + W_2)$, $(W_2 - W_1)$, and U into two relatively independent components: one large, with $(W_2 - W_1)$ and U , and one smaller, with $(W_1 + W_2)$. To evaluate the model, a principal components analysis with varimax rotation was performed for each experiment and temporal condition. The results are summarized in Table 4.

For further evaluation of the comparison model, multiple regression analyses were performed across partici-

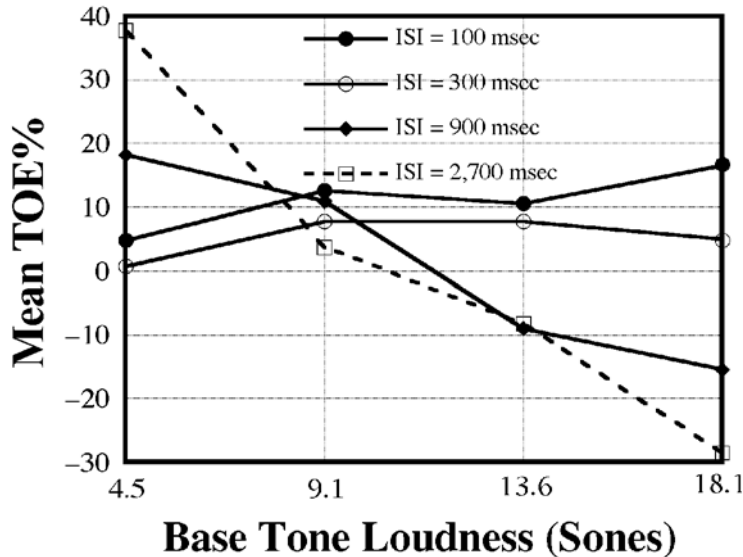


Figure 4. Experiment 4: tone loudness. Mean time-order error percentage (TOE%) plotted against base loudness for each interstimulus interval (ISI).

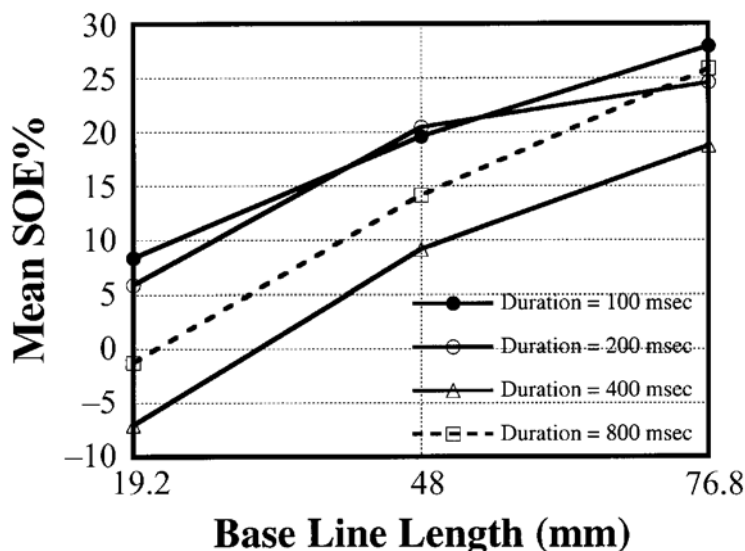


Figure 5. Experiment 5: line length, simultaneous. Mean space-order error percentage (SOE%) plotted against base length for each stimulus duration.

pants for each experiment and ISI or stimulus duration, with U as the dependent variable and the values of $(W_2 - W_1)$ and $(W_1 + W_2)$ as independent variables. Referring to Equation 5, a significant positive coefficient for $(W_2 - W_1)$ indicates that the model is basically correct, in the sense that $D\%$ depends on $(W_2 - W_1)$ and that $(\psi_{r1} + \psi_{r2})$ is >0 and reasonably stable. A significant coefficient for $(W_1 + W_2)$ suggests $\psi_{r1} \neq \psi_{r2}$. Disregarding the outcomes of the significance tests, the fitted coefficients for $(W_2 - W_1)$ and $(W_1 + W_2)$ were used to estimate ψ_{r1} and ψ_{r2} for each experiment and stimulus set according to Equation 5 and also to estimate ϕ_{r1} and ϕ_{r2} , using the relation $\psi_r = \phi_r^m$. The results are summarized in Table 5.

Estimation of Scale Constant and Bias Term

The intercept in the regression just described should estimate $k(\psi_{r1} - \psi_{r2}) + b$ for the particular stimulus set.

Across sets, the obtained value of this intercept was regressed on the corresponding difference between the estimates of ψ_{r1} and ψ_{r2} . The regression coefficient and intercept were taken as estimates of the scale constant k and the bias term b , respectively.³ The results of these analyses are shown in Table 6. The mean values of W_1 and W_2 for each experiment were divided by the corresponding k estimate to obtain approximate estimates of the mean "absolute" weights s_1 and s_2 or s_{left} and s_{right} . In Figures 6–9, the means of these estimates and of the mean TOE% across base stimulus levels are plotted against the ISI for Experiments 1–4. Figure 10 shows the corresponding results for Experiment 5, with the mean estimates of s_{left} , s_{right} , and the mean of SOE% plotted against the stimulus duration.

The trends shown in Figures 1–5 should be seen in relation to the predictions of Equation 2 from the s esti-

Table 3
p Values From Analyses of Variance for Each Experiment

Experiment No.	Stimulus Type	Effect/Dependent Variable										
		ISI W_1	ISI W_2	ISI $W_1 - W_2$	$M \neq 0$ $W_1 - W_2$	$M \neq 0$ TOE%	ISI TOE%	Base Level*		ISI \times Base Level*		
								$W_1 - W_2$	TOE%	$W_1 - W_2$	TOE%	
Successive Stimulus Presentation												
1	tone duration	n.s. (.060)	n.s.	n.s. (.059)	<.0001	.041	.019	.023	<.0001	n.s.	n.s. (.093)	
2	visual duration	n.s. (.055)	.024	n.s.	<.0001	.016	<.0001	n.s.	<.0001	n.s.	n.s.	
3	line length	<.0001	<.0001	<.0001	n.s.	<.0001	<.0001	.042**	.018	n.s.	<.0001	
4	tone loudness	<.0001	<.0001	<.0001	<.0001	.024	n.s.	n.s.	.002	n.s.	<.0001	
Simultaneous Stimulus Presentation												
		Duration W_L	Duration W_R	Duration $W_L - W_R$	$M \neq 0$ $W_L - W_R$	$M \neq 0$ SOE%	Duration SOE%	Base Level*		Duration \times Base Level*		
								$W_L - W_R$	SOE%	$W_L - W_R$	SOE%	
5	line length	<.0001	<.0001	n.s.	<.0001	.0002	.033	n.s.	<.0001	n.s.	n.s.	

Note—L, left; R, right; n.s., not significant ($p > .05$); TOE%, time-order error percentage; SOE%, space-order error percentage. p values between .05 and .10 are given in parentheses. *In separate analyses with high and low base level. ** $W_1 - W_2$ more negative for high base level.

Table 4
Principal Components Analyses: Means Across the Four Temporal Conditions of First Two Eigenvalues (EV₁, EV₂) and Their Varimax Rotated Loadings for W₂ - W₁, W₁ + W₂, and U

Experiment No.	Stimulus Type	EV ₁	W ₂ - W ₁	W ₁ + W ₂	U	EV ₂	W ₂ - W ₁	W ₁ + W ₂	U
Successive Stimulus Presentation									
1	tone duration	2.483	.937	-.320	.943	0.511	-.331	.941	-.309
2	visual duration	2.007	.993	-.032	.992	0.972	-.042	.998	-.020
3	line length	2.080	.999	-.042	.993	0.919	-.092	.994	-.091
4	tone loudness	2.021	.996	-.049	.994	0.969	-.044	.998	-.055
Simultaneous Stimulus Presentation									
5	line length	2.103	.986	.124	.9925	0.888	.152	.992	.098

mates displayed in Figures 6–10 and should make it possible to evaluate the descriptive power of the SW model. According to Equation 1, if in an experiment ψ_{r1} and ψ_{r2} are constant across stimulus sets (ISIs or durations), a d value of $k(\psi_{r1} - \psi_{r2}) + b$ will result, irrespective of s_1 and s_2 , when $\psi_1 = \psi_{r1}$ and $\psi_2 = \psi_{r2}$. Equation 2 predicts that when $\psi_{r1} = \psi_{r2} = \psi_r$, the slope of a plot of TOE% or SOE% (= $D\%$ for $\psi_1 = \psi_2 = \psi$) against the magnitude of the base is proportional to $(s_1 - s_2)$. The plots for different ISIs will then intersect at $d = b$ for $\phi = \phi_r$. Non-intersecting plots for the stimulus sets in an experiment

therefore indicate a difference between ϕ_{r1} and ϕ_{r2} , or a variation in ϕ_r over these sets. Figures 1–5 may thus also be compared with Table 5.

JNDs

In the constant method with two response categories, the JND is defined (Luce & Galanter, 1963) as half the change in the Co stimulus that increases the percentage of correct judgments from 25% to 75% or—modeling the psychometric function with the cumulative normal distribution—that makes z accordingly increase by $\Delta z \approx$

Table 5
Results of Regression Analyses Across Participants, Using Equation 5 With U as Dependent Variable and (W₂ - W₁) and (W₁ + W₂) as Predictors

Successive Stimulus Presentation				
Experiment 1: Tone Duration				
ISI (msec)	500	1,000	2,000	4,000
Multiple R	.993	.997	.995	.994
Estimated ϕ_{r1} (msec)	943	1,038	1,171	860
Estimated ϕ_{r2} (msec)	1,024	1,073	1,258	865
p value for ($\phi_{r1} \neq \phi_{r2}$)	n.s. (.0636)	n.s.	n.s.	n.s.
Experiment 2: Visual Duration				
ISI (msec)	500	1,000	2,000	4,000
Multiple R	.966	.983	.980	.981
Estimated ϕ_{r1} (msec)	976	1,062	1,005	1,038
Estimated ϕ_{r2} (msec)	1,045	1,114	983	1,030
p value for ($\phi_{r1} \neq \phi_{r2}$)	n.s.	n.s.	n.s.	n.s.
Experiment 3: Line Length				
ISI (msec)	100	300	900	2,700
Multiple R	.999	.998	.999	.999
Estimated ϕ_{r1} (mm)	85.1	93.0	94.2	94.2
Estimated ϕ_{r2} (mm)	77.9	99.9	99.6	96.6
p value for ($\phi_{r1} \neq \phi_{r2}$)	.0007	n.s.	n.s. (.0814)	n.s.
Experiment 4: Tone Loudness				
ISI (msec)	100	300	900	2,700
Multiple R	.983	.992	.994	.989
Estimated ϕ_{r1} (dB)	74.3	72.8	74.9	75.5
Estimated ϕ_{r2} (dB)	75.0	72.3	74.4	73.3
p value for ($\phi_{r1} \neq \phi_{r2}$)	n.s. (.0562)	n.s. (.0501)	.0214	n.s.
Simultaneous Stimulus Presentation				
Experiment 5: Line Length				
Duration (msec)	100	200	400	800
Multiple R	.992	.985	.991	.996
Estimated ϕ_{rL} (mm)	51.8	60.0	62.4	54.9
Estimated ϕ_{rR} (mm)	51.7	56.5	57.1	49.6
p value for ($\phi_{rL} \neq \phi_{rR}$)	n.s.	.0378	.0078	.0011

Note—n.s., not significant ($p > .05$); ISI, interstimulus interval; L, left; R, right.

Table 6
Results of Estimation of k and b Using Fits of Equation 5

Experiment No.	Stimulus Type	r	Estimated $k \pm SE$	Estimated $b \pm 95\%$ c.i.
Successive Stimulus Presentation				
1	tone duration	.961	670.6 \pm 137.2	-5.1 \pm 24.1
2	visual duration	.977	1,033.4 \pm 160.0	-98.8 \pm 118.3
3	line length	.997	2,764.7 \pm 148.8	-12.1 \pm 7.6
4	tone loudness	.995	646.1* \pm 55.5	0.0* (28.8**)
Simultaneous Stimulus Presentation				
5	line length	.997	1,515.6 \pm 81.4	20.7 \pm 12.1

Note— SE , standard error; c.i., confidence interval. *Estimated from data for ISI = 300–2,700 msec. **Estimated from data for ISI = 100 msec.

0.6745 (Guilford, 1954, chap. 6). In cases such as the present one with three categories (cf. Hellström, 2000), JND can instead be calculated as Δz times the slope in the regression of $(z_1 + z_2)/2$ on Co. Adapting this procedure to the present design, $(z_1 + z_2)/2$ was calculated for each stimulus pair and ISI or stimulus duration, using the group data, and the regression of this measure on $\log \phi_1$ and $\log \phi_2$ was computed. The logarithmic equivalent of JND was then estimated as the change in $\log \phi_1$ or $\log \phi_2$ that shifted $(z_1 + z_2)/2$ by Δz . The results, recalculated in terms of predicted Weber fractions (WFs) for pooled group data in an experiment with the constant method and presentation orders St–Co and Co–St, are given as percentages in Table 7.

DISCUSSION

As can be seen from Table 3 and from the figures, all the experiments showed differential SW in the sense that, for the individual stimulus sets (differing in ISI or stimulus duration), the effect of base magnitude on TOE% or SOE% was approximately linear (for one exception, see below) and that this effect corresponded to the difference between the two weight coefficients, as predicted by Equation 2. The fits of Equation 3 (for individual participants, within stimulus sets) and Equation 5 (for pooled results, within and between sets) are excellent. For each experiment and stimulus set, the coefficient of regression of U on $(W_1 - W_2)$ is highly significant ($p < .001$).

The ANOVAs of $(W_1 - W_2)$ (Table 3) show that the weighting pattern is virtually independent of the stimulus level: In no case was the ISI \times base level interaction significant; only for tone duration (Experiment 1) and line length, successive (Experiment 3), was the main effect significant; for tone duration, the weighting pattern was brought out more clearly for the high level; for line length, the irregularities discussed later became visible.

The results of the principal components and regression analyses also support the model; they indicate that stimulus weighting and reference levels are important factors for determining the outcomes of stimulus comparisons. In each case, almost all the variation in the parameter estimates can be attributed to two orthogonal

components: one (accounting for about two thirds of the variance) with high loadings of $(W_2 - W_1)$ and U , and the other with a high loading of $(W_1 + W_2)$. Thus, the interindividual variation in the relative weighting of the stimuli is by far the largest source of variability in the parameter estimates.

The results of fitting Equation 5 are shown in Table 5. For each experiment and each stimulus set, the coefficient for $(W_2 - W_1)$ is highly significant, as was predicted. The results for the individual experiments are discussed below.

Auditory and visual duration (Experiments 1–2). For auditory duration, the results, as shown in Figure 6, are very similar to those reported in Hellström (1977) and for the healthy controls in Hellström et al. (1989) and Hellström and Almkvist (1997). The results for visual duration (Figure 7) follow a similar pattern. Thus, the second duration was weighted more heavily than the first, and this weight pattern did not change much with the ISI (cf. Table 3). Accordingly, the TOE changed from positive to negative with increasing base duration (cf. Stott, 1935). The average TOE across base durations was mainly negative and approached zero for the longer ISIs. This is reminiscent of the results of Jamieson and Petrusic (1975a, 1975b) for fixed duration levels. The results for the shortest and longest base durations are also similar to those of Jamieson and Petrusic (1975a, 1975b) for roving duration levels, inasmuch as the TOEs tend to become less negative for long stimuli and short ISIs. However, the general tendency is that there are more positive TOEs in the present Experiments 1 and 2 than in those of Jamieson and Petrusic (1975a, 1975b). Many potentially important features distinguish their experiments from the present ones, such as the type of stimuli (red light diodes), the set of durations (outside the range of the present ones), 2, 4, or 16 experimental sessions instead of 1 (for the importance of the amount of practice, see Needham, 1934; Stott, 1935; Woodrow, 1935), and different ISIs being used in different sessions.

The nonintersecting plots in Figures 1 and 2 suggest that one factor behind the change of the mean TOE% toward zero with longer ISIs shown in Figures 6 and 7 is the changing values of ϕ_{r1} and ϕ_{r2} (see Table 5).

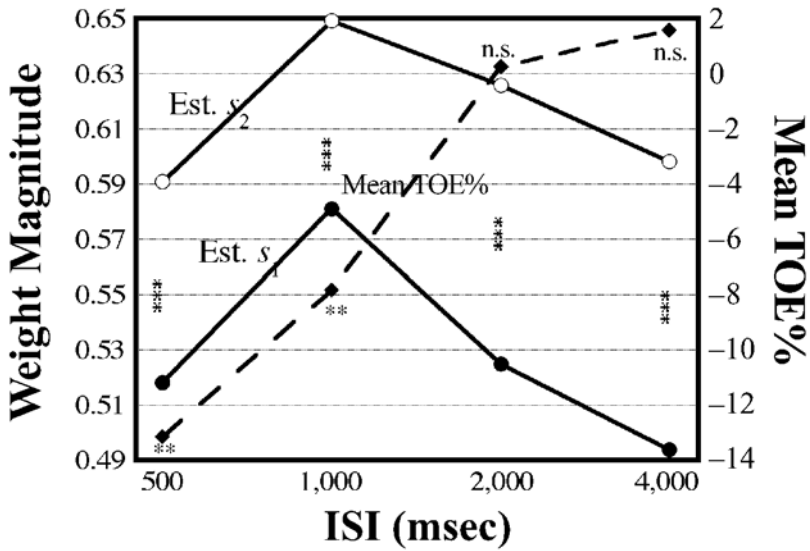


Figure 6. Experiment 1: tone duration. Means of s_{1est} , s_{2est} , and time-order error percentage (TOE%) plotted against interstimulus interval (ISI). Significance of differences ($s_{1est} - s_{2est}$) and (TOE% - 0) are indicated as follows: * $p < .05$; ** $p < .01$; *** $p < .001$; n.s., not significant ($p > .05$).

In later years, TOEs in general have mainly been studied, discussed, and modeled in terms of TOEs for durations, a quite special case (e.g., stimulus magnitude and stimulus duration are the same thing). With consistently negative TOEs, such as those obtained in Experiments 1–2, explanations such as “disuse” (Pratt, 1936), “category effect” (Stevens, 1956), or “response bias” (Allan, 1977) might seem natural. However, as was mentioned above,

Hellström’s (1977) results for comparisons of tone durations, using different response modes, indicate that this type of explanation is implausible (cf. Hellström, 1978, 1985). The weight pattern $s_1 < s_2$ is in agreement with Vierordt’s (1868) law and later results (see Hellström, 1985) that indicate a *central tendency*—that is, a regression toward the mean—for past time intervals. However, again, time perception is special: It seems to depend on

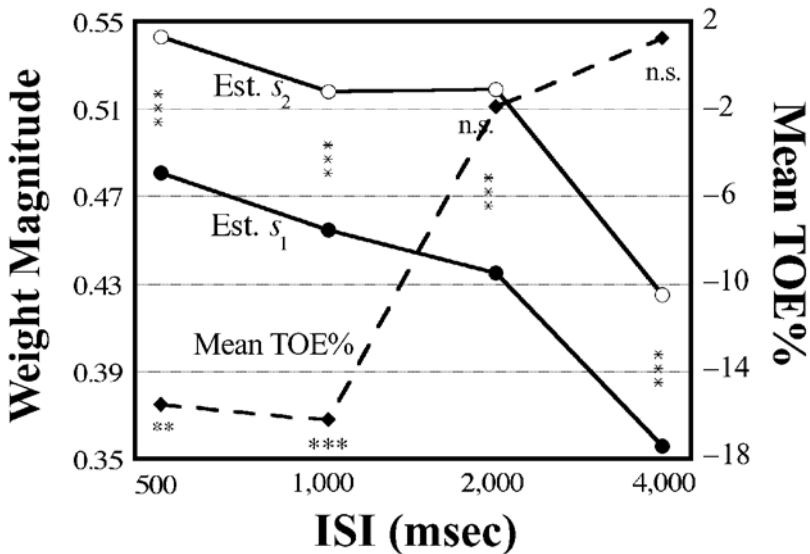


Figure 7. Experiment 2: visual duration. Means of s_{1est} , s_{2est} , and time-order error percentage (TOE%) plotted against interstimulus interval (ISI). Significance of differences ($s_{1est} - s_{2est}$) and (TOE% - 0) are indicated as follows: * $p < .05$; ** $p < .01$; *** $p < .001$; n.s., not significant ($p > .05$).

cognitive processing of experienced events, as well as on a biological *internal clock* or *pacemaker*: Short durations are processed at a low neural level, using the pacemaker; in contrast, the processing of durations of about 1 sec and longer depends mainly on cognitive factors (Ramm-sayer, 1999). For handling these longer durations, intact memory functions are essential (Nichelli, Venneri, Molinari, Tavani, & Grafman, 1993; Richards, 1973; Williams, Medwedeff, & Haban, 1989). This makes it plausible that the weight relation $s_1 < s_2$ and, possibly, the (non-significantly) negative b values are due to retention loss of information about the first duration.

Tone loudness (Experiment 4). Here, the picture is somewhat more complex than for duration. Still, Figure 4 and Tables 3–5 show the results to be quite well behaved in terms of the SW model, the plots intersecting within a small area. The mean TOE% as a function of the ISI (see Figure 9) follows the “classical” pattern with a change from positivity toward negativity with longer ISIs (Köhler, 1923; Needham, 1935; Postman, 1946; cf. Guilford, 1954, p. 309). The results in terms of weighting and TOEs are rather similar to those of Hellström (1979) for comparable conditions. Thus, the same weight pattern as that for duration ($s_1 < s_2$) is obtained for longer ISIs; however, for shorter ISIs, the stimulus weighting is reversed, and both s_1 and s_2 decrease in magnitude. Also, in terms of TOEs, the present results, like those of Hellström (1979), are similar to those in Needham’s (1935) Figure 2, and seem to extrapolate them to shorter ISIs. Needham’s figure shows plots of the TOE as a function of the ISI for different intensity levels: For an ISI of 6,000 msec Needham (1935) found a positive TOE for soft tones and a negative TOE for loud tones, but the curves converge for shorter ISIs and show no effect of

the loudness level for the shortest ISI used, 1,000 msec. A similar pattern of converging TOE curves was reported by Inomata (1963), although for the longer ISIs all TOEs were negative. Berliner, Durlach, and Braidia (1977) obtained results that were similar to the present ones in that the usual inverse relation between stimulus level and TOE vanished for short tone durations and short ISIs. (See their Figure 6, which may be compared with Hellström’s, 1979, Figure 2; their “bias factor” is opposite in sign to the TOE.) However, the reversal of the stimulus level effect for short ISIs (indicating, in terms of Equations 1–2, $s_1 > s_2$) did not occur.

Assuming the connection between stimulus weighting and information processing proposed by Hellström (1985, 1986, 1989a, 1989b), the reason for the pattern of results obtained is likely to be a shift, with increasing ISI, in the balance between the factors affecting the fate of loudness information: from mutual (but mainly proactive) blocking of processing to dissipation from memory.

Line length (Experiment 3). A similar pattern of weights and TOEs (see Figure 8) was obtained as for tone loudness. One difference is that the general decrease of the weights for short ISIs, which was obtained for tones, did not occur for lines. A possible explanation is that, in the processing of successive visual stimuli, only proactive, and no retroactive, blocking takes place. As regards the TOE pattern, although the stimulus duration in Experiment 3 was much shorter (100 msec), one comparison can be made with the results of Baranski and Petrusic (1992): in their Experiment 1 they used an ISI of 500 msec and a stimulus duration of 2,000 msec for the first stimulus; the second stimulus “remained on the screen until the completion of the comparative judgment” (p. 256). Baranski and Petrusic found TOEs that

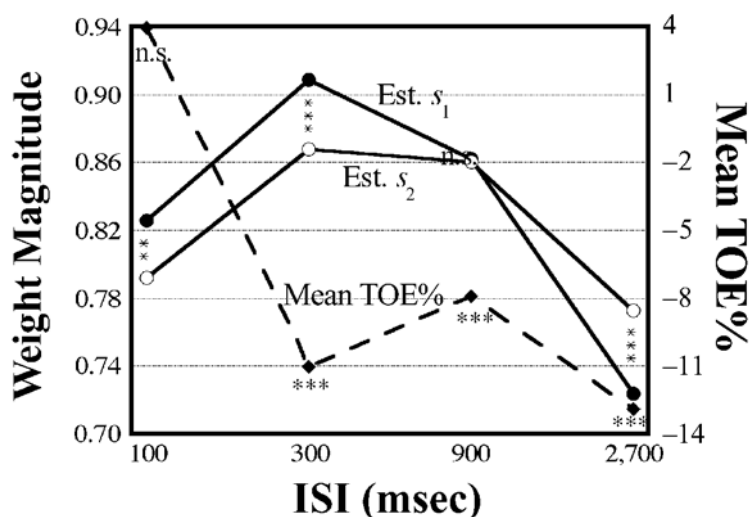


Figure 8. Experiment 3: line length, successive. Means of s_{1est} , s_{2est} , and time-order error percentage (TOE%) plotted against interstimulus interval (ISI). Significance of differences ($s_{1est} - s_{2est}$) and (TOE% - 0) are indicated as follows: * $p < .05$; ** $p < .01$; *** $p < .001$; n.s., not significant ($p > .05$).

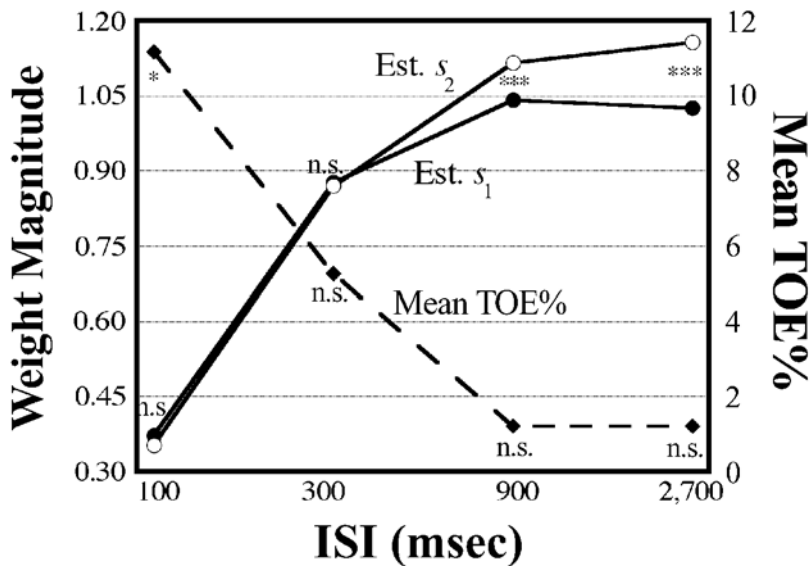


Figure 9. Experiment 4: tone loudness. Means of s_{1est} , s_{2est} , and time-order error percentage (TOE%) plotted against interstimulus interval (ISI). Significance of differences ($s_{1est} - s_{2est}$) and (TOE% - 0) are indicated as follows: * $p < .05$; ** $p < .01$; *** $p < .001$; n.s., not significant ($p > .05$).

were close to zero for the longest lines and became more negative the shorter the lines. This resembles the present results for ISI = 300 msec and can be interpreted in terms of a higher weight for the first than for the second line. Petrusic, Harrison, and Baranski (in press), using the same temporal conditions, found consistently negative TOEs. Another comparison can be made with the results of Masin and Agostini (1990), who used (vertical) lines with an ISI of 5,000 msec and a stimulus duration of 1,500 msec and reported more negative TOEs the longer the lines and a steeper psychometric function for the second line. This resembles the present results for the 2,700-msec ISI and can be interpreted in terms of a higher weight for the second line. Finally, Inomata (1959) studied TOEs for line length systematically. With a stimulus duration of 1,000 msec and ISIs of 500–4,000 msec, he obtained results (see his Figure 2) that are very similar to the present ones, displayed in Figure 3. However, the results of Inomata's experiments also attest to the fact that other stimulus conditions (duration, brightness) interact with the ISI and stimulus level, so that changing these conditions can bring about great changes in TOE patterns.

Figure 3 shows a more complex pattern than does Figure 4. For ISI = 2,700 msec, the plot of TOE% against stimulus level is nonmonotonic, for unknown reasons. The shortest ISIs, 100 and 300 msec, yield rising TOE% curves typical of conditions with $s_1 > s_2$, but for ISI = 100 msec, the TOE% values are more positive. As can be seen from Equation 1, as long as the s values are below 1, the relation $\phi_{r1} > \phi_{r2}$ obtained for this ISI will tend to yield positive TOE% values. As can be seen from Table 6 and confirmed by the significant intercepts (Table 5), the

estimated b value is significantly negative—that is, contributing to a negative TOE.

Simultaneous Stimulus Presentation

In the experiment with simultaneous line length, significant lateralization effects were found, showing up as differential weighting and SOEs. As can be seen from Figure 10, there was a larger weight for the left than for the right stimulus ($W_{left} > W_{right}$)—that is, a left-hemifield advantage that suggests a corresponding right-hemisphere advantage (Beaton, 1986). Similar results were obtained by Masin and Agostini (1991) for vertical lines. Table 2 shows that this effect was statistically significant ($p < .0001$). However, the presence of the fixation cross does not guarantee that the results were not influenced by attentional left-to-right scanning (cf. Masin & Agostini, 1991). There was a positive average SOE—that is, one that tended to “favor” the left line. It was also found that $\phi_{rL} > \phi_{rR}$; the ϕ_r values were clearly above M_{base} , which may be due to the participants' taking the combined length of the lines into account. As can be seen from Table 6, the estimated b value is significantly positive—making the left line tend to seem longer than the right line.

WFs

As can be seen from Table 7, the size of the WF depends on which stimulus is varied; a comparison with Figures 6–10 shows that the ratio of the estimated WFs with orders of St–Co and Co–St is very close to the ratio s_1/s_2 (mean r across stimulus sets within experiments = .9995). As was discussed at length in Hellström (2000), this is a simple consequence of the differential weight-

Table 7
Estimated Weber Fractions in Percentages for Each Experiment and Stimulus Set with Both Presentation Orders of Standard and Comparison Stimuli

Experiment No.	Stimulus Type	Set No.			
		1	2	3	4
Successive Stimulus Presentation					
1	tone duration	16.1/18.9	13.9/15.8	14.7/17.8	14.4/18.0
2	visual duration	23.5/26.5	24.1/28.2	25.0/30.2	32.6/40.0
3	line length	9.9/9.5	9.1/8.5	8.8/8.8	10.5/11.1
4	tone loudness*	46.1/43.4	15.0/15.0	10.5/11.3	9.5/10.8
Simultaneous Stimulus Presentation					
5	line length	7.1/7.0	7.0/6.8	6.1/5.9	6.2/5.9

*Weber fractions are given in terms of sound pressure.

ing: Differing values of s_1 and s_2 make the two stimuli exert different impacts on the subjective difference between them. Also, the mean WFs are negatively related to the corresponding mean s values within each experiment across the four stimulus sets (mean $r = -.99$), as well as across the five experiments ($r = -.79$).

The WFs given in Table 7 are larger than those typically reported for practiced observers. It should first be noted that the present WFs were estimated from pooled group data. However, Hellström (2000) found individual WFs for successively lifted weights to be close to those for pooled data (only the latter were reported in the tables). Also, the number of judgment categories (two, three, or six) made no difference as regards WFs. An important factor, however, was roving versus fixed St; with roving St, the WFs were about 50% larger than with fixed St. For *line length*, Teghtsoonian (1971) listed a typical WF

of 0.029. Baranski and Petrusic (1992) found WFs of 0.032–0.051. As was already noted, the stimulus duration (100 msec) in the present study was much shorter than that used by Baranski and Petrusic, whose subjects also received much more training. For *loudness*, Teghtsoonian listed a typical WF of 0.048. For 1000-Hz tones at 100 dB SPL, Stevens (1948) reported a WF of 0.088, and Holway and Pratt (1936) found values of about 0.10 from reanalyses of older data. Hellström (1977) found WFs of about 0.11 for 1000-Hz, 80-dB tones with an ISI of 1.5 sec and an St (first in the pair) that alternated between 1,000 and 2,000 msec. Hellström (1978), for 1000-Hz, 1,000-msec tones [order, St–Co; St alternating between 70 and 50 dB] found a WF of about 0.16 for the 70-dB level. Hellström (1979) varied both compared tones in the range of 66–73.5 dB with constant duration and ISI. The WFs estimated from his results for a duration

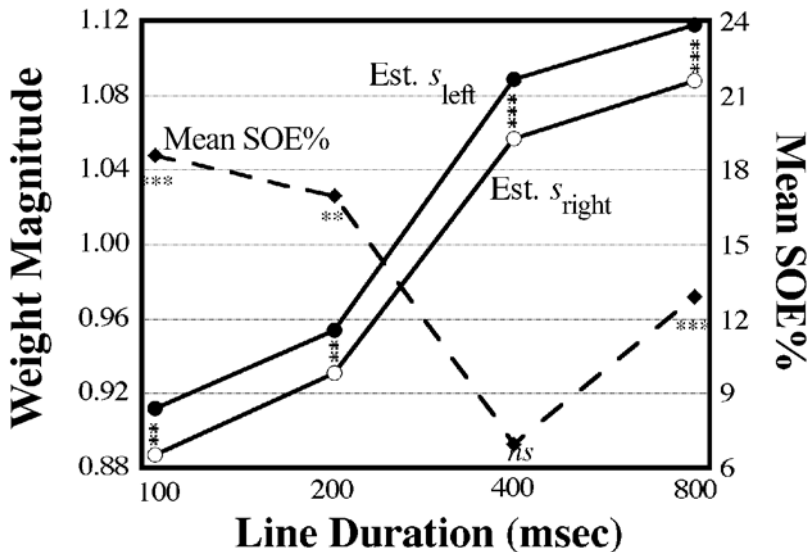


Figure 10. Experiment 5: line length, simultaneous. Means of $s_{left\ est}$, $s_{right\ est}$, and space-order error percentage (SOE%) plotted against stimulus duration. Significance of differences ($s_{left\ est} - s_{right\ est}$) and (SOE% - 0) are indicated as follows: * $p < .05$; ** $p < .01$; * $p < .001$; n.s., not significant ($p > .05$).**

of 100 msec and ISIs of 500–4,500 msec are around 0.16. For ISI = 100 msec, the WF is 0.39 for the equivalent of the order St–Co and 1.97 for Co–St (reflecting the weight ratio for the tones). For *tone duration*, Stott (1935), using 1000-Hz tones, with St stimuli of 200–4,000 msec and an ISI of 1,500 msec, found—in line with the present results—smaller WFs with the order of St–Co than with Co–St; for St = 1,000 msec, the WFs were 0.11 and 0.38, respectively. Rammsayer (2001), using an adaptive method with correctness feedback, obtained a mean WF of 0.16 for his control participants with 1000-Hz, 65-dB tones, St = 1,000 msec, and ISI = 900 msec. Westheimer (1999), using 600-Hz tones with ISI = 600 msec, found the WFs of 2 observers for 150–1,500 msec to be 0.040 and 0.039. For *visual duration* (a bright circle), Westheimer obtained WFs of 0.048 and 0.059 for the same 2 observers and 0.066 for a third. Grondin (1999), using an adaptive procedure and an St of 250 msec, found average WFs of 0.13 for auditory and 0.35 for visual durations.

One important methodological factor behind the relatively large values obtained in the present author's studies was probably the roving-standard designs, and another was the low degree of training, as compared with most other studies. Westheimer (1999) reported that with training, "thresholds improved to about 40% of the initial value" (p. 125). In each of the present experiments, there was only one session with 96 stimulus pairs; Hellström (1977, 1978, 1979) used, in order, 121, 122, and 129 pairs. The results of Berliner et al. (1977) attest to the importance for loudness discrimination of factors such as the range of roving stimulus levels, the WF increasing with the range. (It is worth noting that although these results show improved discrimination for shorter ISIs, for the shortest tones, of 200-msec duration, there is a tendency toward impaired discrimination for very short ISIs, as in the present results and those of Hellström, 1977, for 100-msec tones.)

Psychophysical Exponents

As was mentioned above, power functions were fitted to the stimulus magnitudes. The obtained exponents, given in Table 7, are lower than those typically obtained for the same continua, using magnitude estimation, where, for duration, values near 1.0 and, for loudness, values around 0.6 are usually found (Teghtsoonian, 1971). For duration, a typical exponent is about 0.9 (Eisler, 1976). The present low exponents are probably due partly to the nonnumerical nature of the response measure (cf. Petrusic, Baranski, & Kennedy, 1998) and partly to the compression, discussed above, of the $D\%$ measure of stimulus difference. Lower exponents than those typical for magnitude estimation, although larger than the present ones, were obtained by Hellström (1977) for tone duration and by Hellström (1978, 1979) for tone loudness, using three-category comparison of paired successive stimuli with the $(D/T)^*$ measure. Hellström (1979) remarked that "[the fitted exponent .367] is well within the range of loudness exponents—typically in the .25–.30

range—obtained by interval scaling (Marks, 1974) as well as by nonmetric scaling of judged ratios (Schneider, Parker, Farrell, & Kanow, 1976)" (p. 466).

GENERAL DISCUSSION

The results for the diverse kinds of stimuli used in the present study strengthen the evidence for the view that, regardless of how one conceptualizes the underlying mechanism, differential weighting of sensory magnitudes is a basic principle of the comparison process. Thus, the study has demonstrated the need to use, in an experiment with stimulus comparisons, a design in which both stimuli in the pairs (first and second or left and right) are varied. Doing this, one encounters a phenomenon of much greater importance than the TOE/SOE, which turns out to be mainly a by-product of the differential impact of the two stimuli on the judgment. This relative impact, in turn, depends on the temporal and/or spatial conditions of stimulus presentation. Had comparisons of successive stimuli been studied in this way in the first place, the TOE would hardly have been thought to be due to "bias;" other explanations, of wider scope, would have seemed more plausible (Hellström, 1985).

Although the basic principles involved in stimulus comparison seem to be the same across modalities, the plots and tables in the present article show that the detailed patterns of results differ; this is as should be expected if the model parameters reflect the processing of stimulus information. In the future, therefore, such experiments as the present ones are likely to become tools for studying this processing.

Time-order effects in comparisons of successive stimuli are repeatedly "rediscovered" in various contexts by unsuspecting researchers who are not aware of the generality of the phenomenon and make no references to the TOE literature. The list of modalities for which such "discoveries" have been made includes tonal loudness (Gol'dburt, 1963), vowel pitch (Chuang & Wang, 1978), color purity (Uchikawa, 1983), tactile motion (Essick, Franzén, & Whitsel, 1988), tonal pitch (Jaroszewski, 1992), vibrotactile frequency (Sinclair & Burton, 1996), and, probably, voice onset time (Uselding, 1977).

As this list indicates and as Schab and Crowder (1988) have pointed out, the TOE is a neglected phenomenon. This neglect is probably partly due to the hegemony of the signal detectability paradigm, which has caused stimulus comparison and discrimination to be studied exclusively in terms of resolution and bias, leaving no room for the TOE, except perhaps as a kind of bias (Berliner et al. 1977; cf. Hellström, 2000).

According to Link's (1992) wave theory of subjective difference, the basis for the comparative response is the unweighted difference, cumulated over time and compared with fixed criteria, between the momentary representations (Poisson magnitudes, or "waves") of the two stimuli. Although the theory seems useful in many respects, as a TOE theory it fails to fulfill the minimum re-

quirement of accounting for the differential impacts the compared stimuli have on the judgment. Link made no explicit mention of TOEs (neither did he recognize their dependence on, e.g., ISI or stimulus magnitude) and accounted for the “constant error” by biased response criteria. Using, for illustration purposes, classic data—in particular, those of Guilford (1931)—showing only negative TOEs, Link disregarded the fact that applying the wave theory to more recent experimental data would result in a complicated pattern of bias parameter values needing to be explained (cf. Hellström, 1985). Wave theory needs to be modified to include a model for the integration of the stimulus representations that is more realistic than simple subtraction. In its present form, the theory takes us no closer to an explanation of the TOE.

The research of Masin’s group (e.g., Masin, 1993; Masin & Agostini, 1990; Masin & Fanton, 1989) does address the TOE, as well as the differential effects of the two stimuli on the comparison outcome. In a series of studies using the method of paired comparisons, Masin and his collaborators have demonstrated TOEs and SOEs for length of vertical lines (successive or simultaneous). From the results, they have concluded that the comparison process is based on categorizing the second (or left) stimulus as long (short) and inferring from this that it is longer (shorter) than the first (or right). (The theory is somewhat similar to that of John, 1975, which was discussed at length in Hellström, 1979.) The problem with this theory is that, as regards successive stimuli, it is built exclusively on results obtained for (vertical) line length in such temporal conditions (e.g., an ISI of 5 sec and a stimulus duration of 1.5 sec) as can be expected from the present results to yield $s_1 < s_2$ (this is indeed the case in the data of Masin’s group, although the results were never described in terms of weights). The present results for shorter ISIs and durations (see Figures 8 and 9) require another explanation.

Masin and Agostini (1990) found that explicit length comparison of successive lines yielded (using the present terminology) weighting effects. In contrast, rating the length of the individual lines—serial but not paired—arbitrarily designating successive stimuli as standard and variable stimuli and interpreting the results as “potential comparisons” did not yield such effects. Masin’s (1993) explanation for this was that “actual comparisons involve guesses of comparative responses while potential comparisons do not” (pp. 63–64). He also stated that SW theory cannot explain this result. However, as was proposed by Hellström (e.g., 1985) differential SW for two stimuli occurs in direct comparison of these stimuli. Judging an individual stimulus in a series is a different task, which is by no means predicted by SW theory to yield analogous results when considered as a *potential comparison* with the previous stimulus.

However, as regards judgments of single stimuli (such as visual size, grayness, and line length), Huttenlocher, Hedges, and Vevea (2000) demonstrated that imperfectly remembered stimulus information is dynamically weighted together with information about the category

the stimulus belongs to; this increases the overall accuracy of the judgments but also creates errors that are systematic and related to the true magnitudes. This is a theoretical standpoint that is closely related to the present one and employs modeling similar to SW—in principle, the model used is one half of Equation 1.

CONCLUSIONS

Strangely enough, replication of Hellström’s (1979) experiment seems never to have been attempted by any other researcher. The present results confirm the results of that experiment for a variety of stimuli, thus strengthening the basis for the SW model.

Schab and Crowder (1988) were right in their astonishment at the “modern neglect of the TOE” (p. 233; cf. Allan & Gibbon, 1994). However, the value of studies with the TOE as the sole dependent variable is limited. The neglect of the TOE is astonishing, but still more so is the dominance of oversimplified conceptualizations of stimulus comparison in terms of simple subtraction of one stimulus magnitude from another. Without addressing the differential impact of compared stimuli on the perceived difference and the dependence of this phenomenon on the temporal conditions, research on stimulus comparison falls short of realism and fails to promote psychophysics as an instrument for studying the processing of sensory information.

REFERENCES

- ALLAN, L. G. (1977). The time-order error in judgments of duration. *Canadian Journal of Psychology*, *31*, 24-31.
- ALLAN, L. G., & GIBBON, J. (1994). A new temporal illusion or the TOE once again? *Perception & Psychophysics*, *55*, 227-229.
- ALMKVIST, O., HELLSTRÖM, Å., HÅRDEMARK, H.-G., & JOHANSSON, T. (1996). *Neuropsychological dysfunction two years after subarachnoidal haemorrhage: Relation to acute and long-term brain damage severity* (Rep. No. 809). Stockholm: Stockholm University, Department of Psychology.
- BARANSKI, J. V., & PETRUSIC, W. M. (1992). The discriminability of remembered magnitudes. *Memory & Cognition*, *20*, 254-270.
- BEATON, A. (1986). *Left side, right side: A review of laterality research*. New Haven, CT: Yale University Press.
- BERLINER, J. E., DURLACH, N. I., & BRAIDA, L. D. (1977). Intensity perception: VII. Further data on roving-level discrimination and the resolution and bias edge effects. *Journal of the Acoustical Society of America*, *61*, 1577-1585.
- BOAS, F. (1882). Über die verschiedenen Formen des Unterschiedschwelligkeitswertes [On the different forms of the difference limen value]. *Pflügers Archiv für Physiologie*, *27*, 214-222.
- CHUANG, C.-K., & WANG, W. S.-Y. (1978). Psychophysical pitch biases related to vowel quality, intensity difference, and sequential order. *Journal of the Acoustical Society of America*, *64*, 1004-1014.
- EISLER, H. (1976). Experiments on subjective duration 1868-1975: A collection of power function exponents. *Psychological Bulletin*, *83*, 1154-1171.
- EISLER, H. (1982). On the nature of subjective scales. *Scandinavian Journal of Psychology*, *23*, 161-171.
- ESSICK, G. K., FRANZÉN, O., & WHITSEL, B. L. (1988). Discrimination and scaling of velocity of stimulus motion across the skin. *Somatosensory & Motor Research*, *6*, 21-40.
- FECHNER, G. T. (1860). *Elemente der Psychophysik* [Elements of psychophysics]. Leipzig: Breitkopf & Härtel.
- GOLDBURT, S. N. (1963). Discrimination between the order and loudness

- of two short tones in relation to their duration and the interval between them. *Fiziologicheskij Zhurnal SSSR imeni I. M. Sechenova*, **49**, 1410-1413.
- GRONDIN, S. (1999). Duration discrimination of empty and filled intervals marked by auditory and visual signals. *Perception & Psychophysics*, **54**, 383-394.
- GUILFORD, J. P. (1931). Some empirical tests of the method of paired comparisons. *Journal of General Psychology*, **5**, 64-77.
- GUILFORD, J. P. (1954). *Psychometric methods* (2nd ed.). New York: McGraw-Hill.
- HELLSTRÖM, Å. (1977). Time errors are perceptual: An experimental investigation of duration and a quantitative successive-comparison model. *Psychological Research*, **39**, 345-388.
- HELLSTRÖM, Å. (1978). Factors producing and factors not producing time errors: An experiment with loudness comparisons. *Perception & Psychophysics*, **23**, 433-444.
- HELLSTRÖM, Å. (1979). Time errors and differential sensation weighting. *Journal of Experimental Psychology: Human Perception & Performance*, **5**, 460-477.
- HELLSTRÖM, Å. (1985). The time-order error and its relatives: Mirrors of cognitive processes in comparing. *Psychological Bulletin*, **97**, 35-61.
- HELLSTRÖM, Å. (1986). Sensation weighting in comparing: A tool for optimizing discrimination. In B. Berglund, U. Berglund, & R. Teghtsoonian (Eds.), *Fechner Day 86* (pp. 89-94). Stockholm: International Society for Psychophysics.
- HELLSTRÖM, Å. (1989a). Discrimination optimization: A governing principle for psychophysics. In D. Vickers & P. L. Smith (Eds.), *Human information processing: Measures, mechanisms, and models. Proceedings of the XXIV International Congress of Psychology, Sydney, Australia, 28 August-2 September, 1988* (Vol. 2, pp. 217-227). Amsterdam: Elsevier, North-Holland.
- HELLSTRÖM, Å. (1989b). What happens when we compare two successive stimuli? In G. Ljunggren & S. Dornic (Eds.), *Psychophysics in action* (pp. 25-39). Berlin: Springer-Verlag.
- HELLSTRÖM, Å. (1990). Neuropsychological applications of stimulus comparison tasks. In F. Müller (Ed.), *Fechner Day 90* (pp. 205-211). Würzburg: International Society for Psychophysics.
- HELLSTRÖM, Å. (1992). Comparative judgments of successive and simultaneous lateralized stimuli. In G. Borg & G. Neely (Eds.), *Fechner Day 92: Proceedings of the Eighth Annual Meeting of the International Society for Psychophysics* (pp. 99-104). Stockholm: International Society for Psychophysics.
- HELLSTRÖM, Å. (1993). The normal distribution in scaling subjective stimulus differences: Less "normal" than we think? *Perception & Psychophysics*, **54**, 82-92.
- HELLSTRÖM, Å. (2000). Sensation weighting in comparison and discrimination of heaviness. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 6-17.
- HELLSTRÖM, Å., & ALMKVIST, O. (1997). Tone duration discrimination in demented, memory impaired, and healthy elderly. *Dementia & Geriatric Cognitive Disorders*, **8**, 49-54.
- HELLSTRÖM, Å., FORSELL, L. G., & FERNAEUS, S. E. (1989). Early stages of late onset Alzheimer's disease: V. Psychometric evaluation of perceptual/cognitive processes. *Acta Neurologica Scandinavica*, **79**(Suppl. 121), 87-92.
- HELSON, H. (1964). *Adaptation-level theory*. New York: Harper & Row.
- HOLWAY, A. H., & PRATT, C. C. (1936). The Weber-ratio for intensive discrimination. *Psychological Review*, **43**, 322-340.
- HUTTENLOCHER, J., HEDGES, L. V., & VEVEA, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General*, **129**, 220-241.
- INOMATA, S. (1959). The analysis of conditions behind the time-error in the successive comparison of visual length. *Japanese Psychological Research*, **1**, 36-54.
- INOMATA, S. (1963). Comparison of time-error phenomena in different sense modalities. *Psychologia*, **6**, 207-214.
- JAMIESON, D. G., & PETRUSIC, W. M. (1975a). The dependence of time-error direction on stimulus range. *Canadian Journal of Psychology*, **29**, 175-182.
- JAMIESON, D. G., & PETRUSIC, W. M. (1975b). Presentation order effects in duration discrimination. *Perception & Psychophysics*, **17**, 197-202.
- JAMIESON, D. G., & PETRUSIC, W. M. (1976). On a bias induced by the provision of feedback in psychophysical experiments. *Acta Psychologica*, **40**, 199-206.
- JAMIESON, D. G., & PETRUSIC, W. M. (1978). Feedback versus an illusion in time. *Perception*, **7**, 91-96.
- JAROSZEWSKI, A. (1992). A study of constant (systematic) errors in pitch discrimination of short tone pulses. *Acustica*, **77**, 106-110.
- JOHN, I. D. (1975). A common mechanism mediating the time-order error and the cross-over effect in comparative judgments of loudness. *Australian Journal of Psychology*, **27**, 51-60.
- KÖHLER, W. (1923). Zur Theorie des Sukzessivvergleichs und der Zeitfehler [On the theory of successive comparison and time-errors]. *Psychologische Forschung*, **4**, 115-175.
- LINK, S. W. (1992). *The wave theory of difference and similarity*. Hillsdale, NJ: Erlbaum.
- LUCE, R. D., & GALANTER, E. (1963). Discrimination. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 191-243). New York: Wiley.
- MARKS, L. E. (1974). On scales of sensation: Prolegomena to any future psychophysics that will be able to come forth as science. *Perception & Psychophysics*, **16**, 358-376.
- MASIN, S. C. (1993). The inferential theory of errors in comparative judgment. *Ricerche di Psicologia*, **17**, 57-71.
- MASIN, S. C., & AGOSTINI, A. (1990). Time errors in the method of pair comparisons. *American Journal of Psychology*, **103**, 487-494.
- MASIN, S. C., & AGOSTINI, A. (1991). Attentional scanning and space errors. *Perception & Psychophysics*, **50**, 285-289.
- MASIN, S. C., & FANTON, V. (1989). An explanation for the presentation-order effect in the method of constant stimuli. *Perception & Psychophysics*, **46**, 483-486.
- MICHEL, W. C., & HELSON, H. (1954). A quantitative theory of time-order effects. *American Journal of Psychology*, **67**, 327-334.
- NEEDHAM, J. G. (1934). The time-error as a function of continued experimentation. *American Journal of Psychology*, **46**, 558-567.
- NEEDHAM, J. G. (1935). The effect of the time interval upon the time error at different intensive levels. *Journal of Experimental Psychology*, **18**, 539-543.
- NICHELLI, P., VENNERI, A., MOLINARI, M., TAVANI, F., & GRAFMAN, J. (1993). Precision and accuracy of subjective time estimation in different memory disorders. *Cognitive Brain Research*, **1**, 87-93.
- PETRUSIC, W. M., BARANSKI, J. V., & KENNEDY, R. (1998). Similarity comparisons with remembered and perceived magnitudes: Memory psychophysics and fundamental measurement. *Memory & Cognition*, **26**, 1041-1055.
- PETRUSIC, W. M., HARRISON, D. H., & BARANSKI, J. V. (in press). Long-term memory for elementary visual percepts: Memory psychophysics of context and acquisition effects. *Perception & Psychophysics*.
- POSTMAN, L. (1946). The time-error in auditory perception. *American Journal of Psychology*, **59**, 193-219.
- PRATT, C. C. (1936). The law of disuse. *Psychological Review*, **43**, 83-93.
- RAMMSAYER, T. (1999). Neuropharmacological evidence for different timing mechanisms in humans. *Quarterly Journal of Experimental Psychology*, **52B**, 273-286.
- RAMMSAYER, T. (2001). Effects of noradrenergic activity on temporal information processes in humans. *Quarterly Journal of Experimental Psychology*, **54B**, 247-258.
- RESTLE, F. (1961). *Psychology of judgment and choice*. New York: Wiley.
- RICHARDS, W. (1973). Time reproductions by H.M. *Acta Psychologica*, **37**, 279-282.
- SCHAB, F. R., & CROWDER, R. G. (1988). The role of succession in temporal cognition: Is the time-order error a recency effect of memory? *Perception & Psychophysics*, **44**, 233-242.
- SCHNEIDER, B., PARKER, S., FARRELL, G., & KANOW, G. (1976). The perceptual basis of loudness ratio judgments. *Perception & Psychophysics*, **19**, 309-320.

- SINCLAIR, R. J., & BURTON, H. (1996). Discrimination of vibrotactile frequencies in a delayed pair comparison task. *Perception & Psychophysics*, **58**, 680-692.
- STEVENS, S. S. (1948). Sensation and psychological measurement. In E. G. Boring, H. S. Langfeld, & H. P. Weld (Eds.), *Foundations of psychology* (pp. 250-268). New York: Wiley.
- STEVENS, S. S. (1956). The direct estimation of sensory magnitudes—loudness. *American Journal of Psychology*, **69**, 1-25.
- STEVENS, S. S. (1957). On the psychophysical law. *Psychological Review*, **64**, 153-181.
- STOTT, L. H. (1935). Time-order errors in the discrimination of short tonal durations. *Journal of Experimental Psychology*, **18**, 741-766.
- TEGHTSOONIAN, R. (1971). On the exponents in Stevens' law and the constant in Ekman's law. *Psychological Review*, **78**, 71-80.
- UCHIKAWA, K. (1983). Purity discrimination: Successive vs simultaneous comparison method. *Vision Research*, **23**, 53-58.
- USELDING, D. K. (1977). A temporal order effect in voice onset time discrimination. *Language & Speech*, **20**, 366-376.
- VIERORDT, K. (1868). *Der Zeitsinn nach Versuchen* [The time-sense according to experiments]. Tübingen: Laupp.
- WESTHEIMER, G. (1999). Discrimination of short time intervals by the human observer. *Experimental Brain Research*, **129**, 121-126.
- WILLIAMS, J. M., MEDWEDEFF, C. H., & HABAN, G. (1989). Memory disorders and subjective time estimation. *Journal of Clinical & Experimental Neuropsychology*, **11**, 713-723.
- WOODROW, H. (1933). Weight-discrimination with a varying standard. *American Journal of Psychology*, **45**, 391-416.
- WOODROW, H. (1935). The effect of practice upon time-order errors in the comparison of temporal intervals. *Psychological Review*, **42**, 127-152.

NOTES

1. Fifty-nine participants also took part in one or more of six additional experiments, which are not reported in this article. These experiments employed aesthetic, alphanumeric, and social stimuli (see Hellström, 1992).

2. The computer program was originally written for testing purposes. Using a single pseudorandom stimulus sequence facilitates between-subjects and between-experiments comparisons and eliminates the risk of pathological sequences. However, it could also decrease the generality of the results, if it were to create sequence effects that differed between sets (temporal conditions). Checks of the stimulus sequences revealed no likely sources of such effects, except for a certain variation between the sets in the mean position in the sequence—that is, the amount of previous practice. This variation (in Experiments 1–4, the mean position for Sets 1–4 was 61.3, 49.0, 37.3, and 46.4; in Experiment 5, it was 64.8, 46.0, 40.9, and 42.3) is unlikely to have had any effect on the comparison process. One possible indicator of such an effect would be the ReL. Partialing out the effect of set number (from low to high ISI or duration), the mean ReL estimates for the sets (see Table 5) showed no significant correlation with the mean sequence positions of the sets for any experiment.

3. For Experiment 4 (tone loudness), the results indicated a deviant b value for ISI = 100 sec; therefore, the k estimate was based on data for the remaining ISIs.

(Manuscript received April 21, 2001;
revision accepted for publication April 2, 2003.)