

Mental processes in magnitude estimation of length and loudness

STEPHEN M. KERST and JAMES H. HOWARD, JR.
Catholic University of America, Washington, D.C. 20064

Subjects made magnitude estimates of line length in a short-term memory (STM) task in which the standard and comparison lines were presented for 1 sec each, separated by a 2-sec interval (Experiment 1). Response time (RT) increased systematically with judged magnitude. These findings are consistent with previous results for perceptual magnitude estimation of length (Hartley, 1977, 1981) and show that the effect takes place in STM as well. To explain this effect, Hartley has proposed that subjects repeatedly lay off a mental image of the standard line along the line to be judged. Similar results were found in Experiment 2, in which subjects made magnitude estimates of the loudness of a 1-kHz tone preceded by a standard. Since the relation between RT and judged magnitude appears to hold for the nonspatial continuum of loudness, Hartley's visual imagery model may be too specific to account for the RT effect in the general case. Two alternative explanations are considered: (1) a sensory image model based on Hartley's proposal and (2) an abstract analog model with a nonsensory representation.

The psychophysical function relating actual and estimated length or distance has been studied for some time (e.g., Ekman & Junge, 1961; Stevens & Galanter, 1957), but only recently have the processes involved in making these spatial estimates been investigated. Hartley (1977, 1981) developed an image-based mental measurement model from the introspective reports of pilot subjects in a line-length estimation task. His subjects reported that they laid off a mental image of the standard length along the line to be estimated. Hartley's results were consistent with this model in that he found a linear relation between the time required to make a magnitude estimate (RT) and the value of the estimate itself. Since the standard was always the shortest stimulus, these results suggested that RT depends on the number of times the standard was mentally laid off against the comparison time.

In another study, in which subjects estimated the widths of rectangles, Uhlarik, Pringle, Jordan, and Misceo (1980) found a similar positive relation between judgment RT and the ratio of the standard to comparison stimuli. These authors noted that their results were comparable to those found in tasks that require same-different judgments of shapes that differ in size (e.g., Bundesen & Larsen, 1975; Larsen & Bundesen, 1978; Sekuler & Nash, 1972). In the same-different task, sub-

jects appear to normalize or equate the size of the test figures before making the shape comparison, a process that requires additional time. Uhlarik et al. suggest that the same kind of transformational process may be involved in spatial magnitude estimation.

Hartley's (1977, 1981) image-based mental measurement interpretation seems particularly appropriate for spatial dimensions such as line length, in which the mental ruler metaphor is intuitively appealing. On the other hand, it raises the question of whether the mental measurement effect occurs only for such spatial dimensions or whether the effect is a more general property of the magnitude estimation process. If the effect is not restricted to spatial judgments, a visual imagery explanation would need to be revised.

The purpose of the present experiments was to test whether the relation between RT and judged magnitude holds for both spatial and nonspatial continua under comparable conditions. Experiment 1 tested whether the mental measurement effect is found when estimates are based on sequentially presented lines held in short-term memory (STM), rather than on perceived lines. This procedure provided comparability with Experiment 2. Experiment 2 used sequentially presented tones to test whether the relation between judged magnitude and RT holds for the nonspatial continuum of loudness.

EXPERIMENT 1

Method

Subjects. Nine subjects responded to advertisements at the Catholic University campus and were paid \$8 for serving in two 1-h sessions.

Apparatus. The experiment was controlled by a PDP-11/03 computer. The standard and the comparison lines whose lengths were to be judged were displayed on a Tektronix 4006-1 graphic

This research was supported by grants from the National Institute of Mental Health and the U.S. Army Research Institute to the Catholic University of America. Since the contribution of the authors to this research was equivalent, the order of authorship is arbitrary. The authors acknowledge the helpful comments of Jim Ballas, Leo Gugerty, Kevin Bennett, David Russo, and Darlene Howard. Requests for reprints should be addressed to Stephen M. Kerst, Human Performance Laboratory, Catholic University of America, Washington, D.C. 20064.

display terminal. The terminal screen was 21.4 cm wide by 16.5 cm high. The lines were bright green on a dark background and were .5 mm wide.

Stimuli and Procedure. Subjects were tested individually. Subjects first read instructions that explained that they were to assign a length of 100 units to the standard line and estimate the length of the comparison line relative to the standard. The instructions directed the subjects to answer as rapidly as possible while still giving an accurate response. Each trial was preceded by the message, "THE FIRST LINE IS 100 UNITS IN LENGTH." After the subject pressed the "5" key to start the trial, the standard and the comparison lines were presented sequentially for 1 sec each (standard first) and were separated by a 2-sec interval. RT was measured from the onset of the comparison line. This procedure was adopted because it could be used for loudness estimates in Experiment 3, so that comparable data could be obtained for spatial and nonspatial continua. Subjects responded by pressing the appropriate numbered keys, followed by an "enter" key. The subject's response was displayed on the screen and he/she either pressed a key labeled "C" to confirm it or one labeled "D" to delete it (i.e., if he/she had made a typing mistake). Deleted trials were presented again at a randomly chosen point later in the session.

RT was measured from stimulus onset to the first keypress of the response, with an accuracy of 1 msec. The numbered keys on the Tektronix terminal were rearranged in four rows, with 7, 8, 9 in the top row, 4, 5, 6 in the second row, 1, 2, 3 in the third row, and 0 in the bottom row. This approximates the standard keypad layout of an electronic calculator and ensures that after pressing "5" to begin the trial, the subject's finger is centered among the keys.

The standard and comparison lines were horizontal and were located the same distance from the bottom of the screen. The standard lines were presented on the right side of the screen, with their midpoints centered around a point 19.0 cm over and 8.8 cm up from the lower left-hand corner of the screen. The comparison lines were presented on the left-hand corner of the screen, with their midpoints centered around a point 9.5 cm over and 8.8 cm up from the lower left-hand corner of the screen. There were six different lengths of comparison lines equally spaced in the log steps (1.6, 2.5, 3.8, 5.7, 8.7, and 13.2 cm). Each of these was displayed 32 times with the standard line (1.6 cm), for a total of 192 randomized trials in each of the two experimental sessions. The first session was preceded by 30 randomized practice trials.

Results and Discussion

Preliminary analysis. To eliminate extreme RTs, latencies greater than ± 2.6 SDs from the appropriate cell mean for the subject were excluded. Sixty-seven extreme RTs (2.4%) were discarded.

Length estimates. Power functions were fit to the data for individual subjects. The median exponent was .97. The median correlation of log estimate and log actual length was .99, which shows a good fit for the power function. For the group data, the exponent was .97 and the correlation was .99. These results are similar to those usually found for perceived length (Baird, 1970), for which an exponent of 1.0 is typical. The group data and power function were plotted in log/log coordinates. Visual inspection showed that the data did not deviate from linearity.

Reaction time. Individual subjects' correlations between RT and estimated length ranged from .34 to .99, with a median of .97. The relation between the

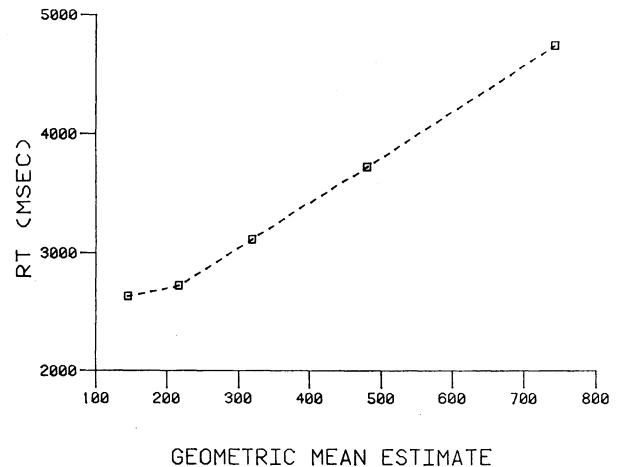


Figure 1. Mean response time as a function of estimated length in Experiment 1. The standard is not shown.

group geometric mean of estimated length and mean RT is shown in Figure 1 ($r = .99$, $p < .001$, one-tailed test). This strong linear relation is similar to the results found with perceptual comparisons (Hartley, 1977, 1981).

In Experiment 1, however, STM is involved in the comparison process. The standard and comparison stimuli were presented sequentially for 1 sec each with an interstimulus interval (ISI) of 2 sec, so that both stimuli were not simultaneously present for comparison. Mean RTs ranged from 2,635 msec for the shortest comparison line to 4,743 msec for the longest, measured from the onset of the comparison (second) line. Consequently, estimates were completed when neither the standard nor the comparison line was visible. The mental measurement effect was thus obtained when a purely perceptual process of "laying off" the standard against the comparison was not possible. In fact, both lines were in STM before the comparison process was completed. Hartley (1981) found similar results when the comparison line was perceptually available and the standard was either imagined or omitted completely. Going beyond Hartley's results, Experiment 1 showed that the perceptual presence of the comparison line itself is not required for the mental measurement effect to take place. Jonides and Baum (Note 1) found a somewhat comparable effect with campus distances judged entirely from long-term memory (LTM), although the relation between subjective magnitude and RT was considerably less linear than that found in Experiment 2.

EXPERIMENT 2

Overall, the results of Experiment 1 were consistent with the earlier findings of Hartley (1977, 1981) and Jonides and Baum (Note 1) in supporting an image-based mental measurement model. This interpretation seems

particularly appropriate for spatial dimensions such as line length or campus distance, for which the mental ruler metaphor is intuitively appealing. On the other hand, it raises the question of whether the mental measurement effect occurs only for such spatial dimensions or whether the effect is a more general property of the magnitude estimation process. In Experiment 2, we tested the generality of the mental measurement effect by having subjects make memory-based judgments of a nonspatial continuum under conditions similar to those of Experiment 1. Subjects evaluated the subjective magnitude of six comparison intensities of a 1-kHz pure tone relative to a standard intensity. The demonstration of a mental measurement effect for loudness would establish the effect as a more general property of the process of magnitude estimation.

Method

Subjects. Twelve student volunteers served as subjects in the experiment. None of the subjects reported any history of hearing disorders. Each served in two sessions of 1-1.5 h and received \$9 for participating.

Apparatus. A 1-kHz pure tone was synthesized on a PDP-8/e laboratory computer using digital techniques. The tone was output on a 12-bit digital-to-analog converter at a 10-kHz sampling rate. The signal was low-pass filtered at 3 kHz (Khron-Hite Model 3550), attenuated by a programmable attenuator (Charybdis Model D), and delivered to subjects binaurally over Telephonics TDH-49 headphones with MX-41/AR cushions. A television monitor displayed trial messages to subjects.

Stimuli. The 1-kHz tone was attenuated in equal 6-dB steps to produce a series of six different intensities. The resulting intensities measured at the headphones (B & K Type 2203 sound-level meter with a Type 4152 artificial ear) ranged from 67 to 96 dB SPL, in steps of 5.6-6.0 dB SPL. This physical intensity range was selected to produce sounds that could be easily discriminated in loudness. The first, or standard, stimulus was the softest (67 dB). The second, comparison, stimulus could be any of the six intensities. Each sound was presented for 1 sec with an ISI of 2 sec.

Procedure. The experimental procedure was generally comparable to that of Experiment 1. The number of sessions, practice trials, and experimental trials was the same as in that study. The instructions and trial-by-trial procedure were identical to those of Experiment 1, with the exception that the modulus was changed from 100 units to 1 unit to simplify the subjects' task of entering their estimates. Each trial was preceded by the message "REMEMBER, THE FIRST SOUND IS 1 UNIT OF LOUDNESS." RT was measured from the onset of the comparison (second) tone.

Results and Discussion

Preliminary analysis. To eliminate extreme RTs, latencies greater than ± 3 SDs from the appropriate cell mean for the subject were excluded. Ninety-three RTs (2.4%) were thus discarded.

Loudness estimates. Power functions were fit to the data for individual subjects. The median exponent was .61. The median correlation of log estimates and decibel sound pressure was .99, which shows a good fit for the power function. For group data, the exponent was .64 and the correlation was .99. The group data and power function were plotted in log/log coordinates (decibel

SPL is a logarithmic scale). Visual inspection showed that the data did not deviate from linearity. The obtained exponent was close to the .60 level typically found for loudness (Marks, 1974).

Reaction time. Ten of the 12 individual subjects' correlations between RT and judged loudness were positive, with a range of .27 to .95. Two subjects showed negative correlations (-.46 and -.28). The median individual correlation was .68. The relation between the group geometric mean of estimated loudness and mean RT is shown in Figure 2 ($r = .82, p < .05$), one-tailed test).

The two negative correlations show that not all subjects performed the task in the same way. These two subjects, however, did not show a complete reversal of the trend for the group. Instead, they showed a rapid monotonic increase in RT over a small range of estimates for the three faintest stimuli, followed by a large drop in RT over a large range of estimates for the two loudest tones. The drop in RT in the group data for estimates of the loudest stimulus is due to the performance of these subjects and three others who showed a drop in RT for the loudest tone (total N = 12). This drop may be related to the range of estimates selected. Although there may be other differences involved, subjects whose range of mean estimates was less than about sixfold did not show a drop in RT, whereas those who used a larger range did show such a drop, starting with estimates near six. This observation suggests that small estimates may involve a process different from that involved in large ones. Small estimates are made by a strategy in which RT is related to subjective magnitude, but this strategy may be abandoned for larger estimates and replaced with a rapid process in which RT does not increase with subjective magnitude.

Although the median individual correlation of judged magnitude and RT for loudness (.68) is less than that

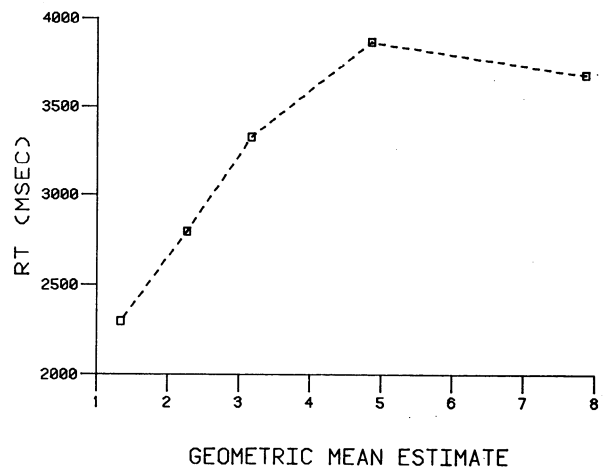


Figure 2. Mean response time as a function of estimated loudness in Experiment 2. The standard is not shown.

found for length in Experiment 1 (.97), it is greater than the figure found by Hartley (1977) for length (.52) and compares favorably with his figures of .83, .86, .75, and .69 from four length experiments (Hartley, 1981). The present data support the conclusion that RT and judged magnitude are related for the nonspatial continuum of loudness, much as they are related for the spatial continuum of length. The implications of this finding for an image-based mental measurement model of magnitude estimation are considered in the General Discussion.

GENERAL DISCUSSION

We conclude that the increase in RT with judged magnitude occurs when the standard and comparison stimuli are held in STM (Experiment 1) and that the effect holds for the nonspatial dimension of loudness (Experiment 2). These findings suggest that similar processes underlie relative magnitude judgment for spatial and nonspatial continua.

This conclusion implies that Hartley's (1977, 1981) visual imagery model for line-length estimation may be too specific to account for the RT effect in the general case. Two general models of the mental measurement effect were considered in our analysis. The sensory image model was based on a direct extension of Hartley's proposal. It argues that in relative magnitude estimation, the subject mentally lays off a sensory representation of one stimulus against a representation of the other. In the spatial, length estimation case, the subject works with visual codes, whereas in the intensive, loudness case, the subject works with auditory codes. A similar proposal was offered by Krantz (1972) when he introspected that his own loudness magnitude estimates were based on the number of times that one loudness would "fit into" another (p. 190). Our findings, however, suggest that Krantz's introspections are not typical, since none of the subjects in Experiment 2 described this strategy in a postexperimental interview.

The abstract analog model also proposes that magnitude estimates are determined by "fitting" one internal representation into another, but unlike the sensory image model, a common nonsensory abstract code is thought to underlie the mental representations for all continua. Palmer (1975) has described a similar model of visual representation. Since the abstract analog and sensory image models propose that subjects manipulate informationally equivalent representations (Palmer, 1975), we will make no attempt to distinguish between them. However, one can speculate on the kinds of processes that are used to compare stimulus representations. Two alternatives come to mind. On one hand, subjects may literally "mentally fit" one representation of a stimulus into another, as suggested by Hartley (1977, 1981) and Pringle and Fedorchuk (Note 2). On the other hand, subjects may use some kind of continuous normalization or adjustment process to bring one representa-

tion into congruence with the other (Dixon & Just, 1978; Larsen & Bundesen, 1978). Any attempt to distinguish between these processes extends beyond the scope of the present study and will be left for future research.

REFERENCE NOTES

1. Jonides, J., & Baum, D. R. *Cognitive maps as revealed by distance estimates*. Paper presented at the 50th meeting of the Midwestern Psychological Association, May 1978.
2. Pringle, R., & Fedorchuk, R. A. *Laying-off processes in magnitude estimation of length, area, and numerosity*. Manuscript submitted for publication, 1983.

REFERENCES

- BAIRD, J. C. *Psychophysical analysis of visual space*. London: Pergamon Press, 1970.
- BUNDESEN, C., & LARSEN, A. Visual transformation of size. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, 1, 214-220.
- DIXON, P., & JUST, M. A. Normalization of irrelevant dimensions in stimulus comparisons. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, 4, 36-46.
- EKMAN, G., & JUNGHE, K. Psychophysical relations in visual perception of length, area, and volume. *Scandinavian Journal of Psychology*, 1961, 2, 1-10.
- HARTLEY, A. A. Mental measurement in the magnitude estimation of length. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, 3, 622-628.
- HARTLEY, A. A. Mental measurement of line length: The role of the standard. *Journal of Experimental Psychology: Human Perception and Performance*, 1981, 7, 309-317.
- KRANTZ, D. H. A theory of magnitude estimation and cross-modality matching. *Journal of Mathematical Psychology*, 1972, 9, 168-199.
- LARSEN, A., & BUNDESEN, C. Size scaling in visual pattern recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, 4, 1-20.
- MARKS, L. E. On scales of sensation: Prologomena to any future psychophysics that will be able to come forth as a science. *Perception & Psychophysics*, 1974, 16, 358-376.
- PALMER, S. E. Visual perception and world knowledge. In D. A. Norman, D. E. Rumelhart, & the LNR Research Group (Eds.), *Explorations in cognition*. San Francisco: Freeman, 1975.
- SEKULER, R., & NASH, D. Speed of size scaling in human vision. *Psychonomic Science*, 1972, 27, 93-94.
- STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, 1957, 54, 377-411.
- UHLARIK, J., PRINGLE, R., JORDAN, K., & MISCEO, G. Size scaling in two dimensional pictorial arrays. *Perception & Psychophysics*, 1980, 27, 60-70.

(Received for publication March 2, 1983.)