HOW STABLE IS THE STIMULUS MAGNITUDE EFFECT ON PERCEIVED DURATION?

Martin Verner and Thomas Rammsayer
Institute for Psychology, University of Bern, Switzerland
martin.verner@psy.unibe.ch

Abstract

In the present study, temporal reproduction was used to further elucidate the effect of stimulus magnitude on perceived duration. More specifically, we investigated split-half and test-retest reliabilities of perceived duration as obtained by the temporal reproduction task as well as the effect of stimulus magnitude on perceived duration. For this purpose, 36 participants reproduced the duration of visual stimuli presented for 800, 1000, and 1200 msec, respectively, varying in non-temporal stimulus magnitude. There were two experimental sessions separated by a one-week test-retest interval. A statistically significant main effect of stimulus magnitude on perceived duration could be established with larger stimuli being perceived as longer than smaller ones. As indicated by split-half and test-retest coefficients, perceived durations proved to be highly reliable. Only rather low test-retest coefficients, however, could be demonstrated for the effect of stimulus magnitude on perceived duration indicating poor stability for this latter effect.

Several studies provided empirical evidence for a positive relationship between perceived duration and nontemporal stimulus magnitude. For instance, the perceived duration of visually presented stimuli becomes longer with increasing stimulus magnitude such as size or stimulus complexity (e.g., Long & Beaton, 1980; Ono & Kawahara, 2007, Thomas & Cantor, 1975; Xuan, Zhang, He & Chen, 2007). This effect is referred to as the stimulus magnitude effect on perceived duration. One of the most suitable psychophysical approaches to measure perceived duration represents the so-called temporal reproduction task. Unfortunately, the amount of literature on the reliability of temporal reproduction tasks or the stimulus magnitude effect on perceived duration is very limited. The present study, therefore, was designed to systematically investigate split-half and test-retest reliability of reproduced durations as well as the temporal stability of the stimulus magnitude effect.

Method

Participants

Thirty-six female volunteers, ranging in age from 18 to 28 years (mean age ± SD: 21.3 ± 2.1 years), participated in the experiment. All were undergraduate psychology students from the University of Bern and had normal or corrected-to-normal vision. Informed consent was obtained from each participant prior to the experiment.

Apparatus and Stimuli

The presentation of the stimuli was controlled by E-Prime 2.0 experimental software running on a Dell Optiplex 760 Computer. Participants' responses were logged by means of a Cedrus RB-730 response box. Target stimuli were either filled squares or circles presented in two different sizes subtending a visual angle of 1.2° and 10.0°, respectively. Using a 17" monitor, all stimuli were presented in black color on a white background

Procedure

Participants performed two different visual temporal reproduction tasks a couple of times with a test-retest interval of one week. Order of tasks was randomized across participants. On each task, participants were required to reproduce three different target durations (TD; 800 msec, 1000 msec, and 1200 msec). There were 16 presentations of each TD, resulting in a total of 48 trials per task. Within each trial, the TD was followed by a blank screen for 900 msec (pre-reproduction delay). The start of the reproduction interval was marked by the appearance of a fixation cross. The participants were instructed to end the presentation of the fixation cross by pressing a designated button when its display duration was temporally identical to the corresponding TD. A blank screen between the trials was displayed either for 1000 msec or 1400 msec to prevent a rhythmic response during duration reproduction. To assure the perception of stimulus magnitude information, participants were required to indicate, in addition to the temporal reproduction task, whether the nontemporal target stimulus was a circle or a square (Task A) or whether it was small or large (Task B). More specifically, the participants had to press the left button for duration reproduction (e.g., to terminate the reproduction interval marked by a fixation cross) if the stimulus indicating the TD was small; if a large stimulus was displayed, the right button had to be pressed (see Figure 1). The assignment of response button to hand was held constant within each participant but balanced across participants. Within each task, presentation order of TD, target stimulus magnitude and target stimulus shape was randomized. Prior to each task, instructions were given followed by five practice trials.

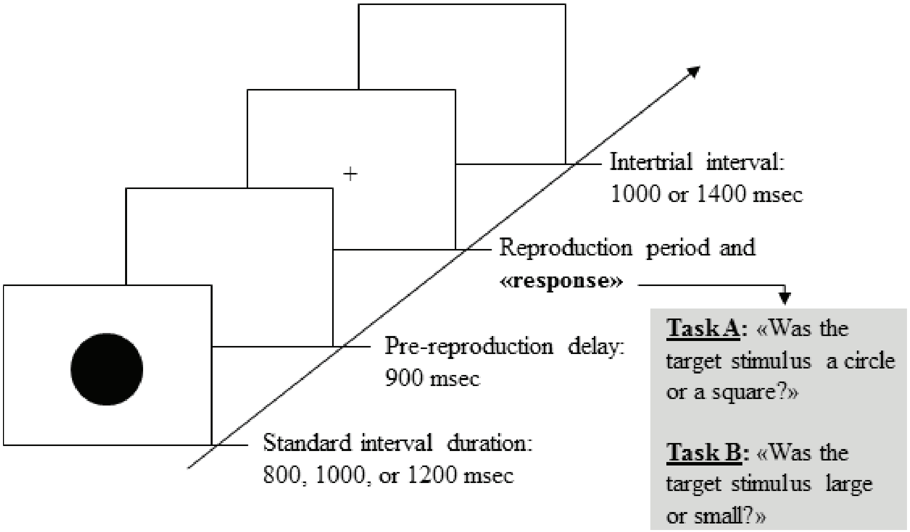

Figure 1. The sequence of events within an experimental trial.

Data Analysis

For each trial, the reproduced duration was logged. The mean reproduced durations (MRD) were compared by means of a three-way analysis of variance (ANOVA) with Magnitude (small and large target stimuli), TD (800 msec, 1000 msec, and 1200 msec) and Time (first and second experimental session) as three repeated-measurement factors. The stimulus magnitude effect was quantified as the difference score between the MRD for large target stimuli and the MRD for small target stimuli. In order to assess the reliability of MRD and of the stimulus magnitude effect, split-half coefficients as derived from the odd-even method were calculated for each experimental session. Resulting coefficients were corrected by means of the Spearman-Brown formula to predict reliability of the whole test length. As to test-retest reliabilities of MRD and the stimulus magnitude effect, coefficients based on the two experimental sessions were calculated by means of Pearson correlations.

Results

Effects on perceived duration

Initial analysis of the data indicated that neither geometrical shape of the target stimuli (squares or circles) nor type of the nontemporal secondary task (Task A and Task B) had an influence on duration reproduction. Hence, for subsequent statistical analyses, MRDs were collapsed across shape of the target stimuli and type of the nontemporal task (Table 1). Greenhouse-Geisser corrected p values are reported where appropriate (for all main effects of TD and the interactions with TD) to protect against violations of sphericity (Geisser & Greenhouse, 1958).

As one would expect, three-way ANOVA yielded a statistically significant main effect TD [$F(1.40, 48.94) = 220.94$; $p < .001$; $\eta_p^2 = .86$] indicating that longer TDs were reproduced longer than shorter TDs. There also was a significant main effect of Magnitude [$F(1, 35) = 34.77$; $p < .001$; $\eta_p^2 = .50$]; large target stimuli were reproduced longer than small target stimuli. Eventually, a significant main effect of Time [$F(1,35) = 21.19$; $p < .001$; $\eta_p^2 = .38$] revealed longer reproduced intervals (i.e., perceived durations in the second experimental session conducted one week after the first session).

All three two-way interactions yielded statistical significance. The interaction between Time and TD [$F(1.80,62.94) = 6.74$; $p < .01$; $\eta_p^2 = .16$] was caused by non-significant differences between MRDs from different TDs (non-orthogonal comparisons). Differences between the two experimental sessions were statistically significant for all three TDs ($p < .001$) as revealed by a post-hoc Scheffé test.

The significant interaction between Magnitude and Time [$F(1,35) = 7.25$; $p < .05$; $\eta_p^2 = .17$] revealed a larger stimulus magnitude effect in the second than in the first experimental session.

Finally, a significant interaction between Magnitude and TD [$F(1.71,59.71) = 8.62$; $p < .01$; $\eta_p^2 = .20$] indicated a more pronounced effect of stimulus magnitude on perceived duration with increasing TD; all orthogonal comparisons yielded statistical significance (Scheffé post hoc test: $p < .01$ for 800 msec; $p < .001$ for 1000 and 1200 msec).

Reliability of reproduced durations and the stimulus magnitude effect

Corrected split-half reliabilities as well as test-retest coefficients for MRDs are presented in Table 2. The split-half reliabilities reached high values, particularly for the second experimental session where coefficients ranged from .94 to .98. The test-retest reliabilities for MRDs were moderate with coefficients ranging from .60 to .71.

Table 1. Mean reproduced durations (MRD) and standard error of mean (SEM) for each target duration, stimulus size, and experimental session.

	Target duration					
	800 msec		1000 msec		1200 msec	
	MRD	SEM	MRD	SEM	MRD	SEM
Measurement 1						
Small stimuli	895	25	1032	27	1148	28
Large stimuli	899	27	1077	31	1223	30
Measurement 2						
Small stimuli	1008	35	1144	37	1226	38
Large stimuli	1077	45	1218	42	1329	46

Table 2. Split-half (odd-even) and test-retest reliability (one-week interval) for mean reproduced durations of the three target durations.

		Target duration		
		800 msec	1000 msec	1200 msec
Measurement 1 (split-half)	Small stimuli	.81	.87	.92
	Large stimuli	.94	.90	.94
Measurement 2 (split-half)	Small stimuli	.98	.94	.96
	Large stimuli	.97	.94	.96
Test-retest reliability	Small stimuli	.70	.65	.60
	Large stimuli	.68	.71	.70

Table 3. Split-half (odd-even) and test-retest reliability for the stimulus magnitude effect obtained with the three target durations.

	Target duration		
	800 msec	1000 msec	1200 msec
Measurement 1 (split-half)	.25	.26	.47
Measurement 2 (split-half)	.43	.71	.72
Test-retest reliability	.10	.01	.13

In contrast to perceived duration, reliability measures of the stimulus magnitude effect were much lower (see Table 3). Despite the reliable MRDs, the stimulus magnitude effect, defined as the difference between MRDs for large and small stimuli, seemed quite unreliable. This especially holds for the first experimental session with split-half coefficients ranging from .25 to .47. Moreover, temporal stability of the stimulus magnitude effect was even worse as indicated by test-retest coefficients not exceeding .13.

Discussion

The present study systematically investigated split-half and test-retest reliability of the measurement of perceived duration and of the stimulus magnitude effect on perceived duration. For this reason, performance on two temporal reproduction tasks was assessed twice in two experimental sessions separated by a one-week test-retest interval. Although, the focus of this work lay on reliability, several effects that could exert some influence on perceived duration will be discussed first.

In both experimental sessions, a large effect of stimulus magnitude on perceived duration was found as indicated by MRDs. Physically larger target stimuli led to longer MRDs compared to smaller target stimuli. There were two types of secondary task employed in the present study. In task A, the participants were instructed to additionally focus on stimulus shape (irrespective of the physical size the target stimulus), whereas in task B, they had to indicate if a presented target stimulus were either physically large or small. The stimulus magnitude effect could be observed independently of the type of the secondary task. These results support the idea of a generalized magnitude system (Walsh, 2003) as well as the notion that magnitude information of a stimulus is processed in an automatic manner (Dehaene & Akhavein, 1995; Schwarz & Ischebeck, 2003).

Moreover, this effect seemed to become more pronounced with increasing TD. To further elucidate to what extent this interaction could be explained by a proportional increase of the stimulus magnitude effect across TDs, MRDs were standardized for TD in an additional analysis. The interaction between Magnitude and TD remained significant ($p < .01$) indicating that the stimulus magnitude effect increases more than just proportionally to TD.

Perceived duration was also influenced by the factor Time. Consistently, the reproduced intervals were longer in the second than in the first experimental session. This result provides additional converging evidence for a repetition effect as described by Hicks and Allen (1979) or Wearden, Pilkington and Carter (1999). These authors explained this effect by declining arousal as consequence of task monotony. Indeed, the tasks used in the present study might have been experienced as rather monotonous and boring. During the second experimental session, this phenomenon might have been particularly strong since participants could anticipate the procedure.

In addition, the significant interaction between Magnitude and Time indicated a more pronounced stimulus magnitude effect in the second compared to the first experimental session. This stronger effect might be explained by a habituation effect of sorts. If one assumes that participants were not familiar with temporal reproduction tasks in the first experimental session, they might have tried out different strategies to memorize and reproduce the presented TDs. In the second experimental session, however, it is reasonable to assume that participants benefited from their experience during the previous session and, thus, applied the strategy that showed the best prove of value during the first experimental session. This notion receives support from our finding of increasing *inter*-individual variability from the first to the second experimental session (see Table 1.) and a concomitant decrease in *intra*-individual variance as reflected by the respective coefficients of variation (standard deviation of reproduced duration divided by MRD).

For MRD, the split-half coefficients ranging from .81 to .98 revealed high reliability within an experimental session. Although being lower than split-half reliabilities, test-retest coefficients ranging from .60 to .70 suggested MRDs to be moderately stable across the one-week test-retest interval. Neither split-half coefficients nor test-retest coefficients appeared to be differentially affected by the type of the secondary task.

The stimulus magnitude effect on perceived duration - quantified as the difference between MRDs for large and small stimuli – was shown to be less reliable than the MRDs. Split-half reliabilities were rather inconsistent and ranged from .25 to .72 while the stimulus magnitude effect seemed to be more consistent in the second compared to the first experimental session. Such a difference might be explained by the participants' higher precision in reproduced durations (lower intra-individual variance) in the second experimental session. However, with test-retest coefficients between .01 and .13, the stimulus magnitude effect proved not to be stable across time. A possible explanation for this obvious lack of temporal stability could be seen in the statistical method applied. It is well known, that the reliability of difference scores is rather low, especially if underlying variables are highly correlated to each other (Cohen & Cohen, 1983). Because the MRDs for large and small stimuli were highly correlated in the present study, the difference scores may include a large amount of measurement error. Thus, the stability of the magnitude effect on perceived duration could be somewhat underestimated. Nevertheless, split-half coefficients in the second experimental session reached moderate values for the 1000- and 1200-msec TDs, respectively, despite the use of difference scores. Thus, the large discrepancy between split-half and test-retest coefficients could be additional support for a weak stability of the stimulus magnitude effect across time. Furthermore, the available data do not support the idea of certain individuals being more sensitive to stimulus magnitude compared to others.

In conclusion, the tasks employed in the present study proved to be well suited to investigate the stimulus magnitude effect on perceived duration. The mean reproduced durations were highly reliable within one testing session and moderately stable across time. However, even if partially consistent within a given experimental session the effect of stimulus magnitudes on perceived duration does not seem to be stable across time.

References

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Dehaene, S., & Akhavein, R. (1995). Attention, automaticity, and levels of representation in number processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 314-326.

Geisser, S., & Greenhouse, S.W. (1958). An extension of Box's results on the use of the *F* distribution in multivariate analysis. *Annals of Mathematical Statistics 29*, 885-891.

Hicks, R., & Allen, D. (1979). The repletion effect in judgments of temporal duration across minutes, days, and months. *American Journal of Psychology, 92*, 323-333.

Long, G. M., & Beaton, R. J. (1980). The contribution of visual persistence to the perceived duration of brief targets. *Perception & Psychophysics, 28*, 422-430.

Ono, F., & Kawahara J.-I. (2007). The subjective size of visual stimuli affects the perceived duration of their presentation. *Perception & Psychophysics, 69*, 952-957.

Schwarz, W., & Ischebeck, A. (2003). On the relative speed account of number-size interference in comparative judgments of numerals. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 507-522.

Thomas, E. A. C., & Cantor, N. E. (1975). On the duality of simultaneous time and size perception. *Perception & Psychophysics, 18*, 44-48.

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences, 7*, 483-488.

Wearden, J. H., Pilkington, R., & Carter, E. (1999). “Subjective lengthening” during repeated testing of a simple temporal discrimination. *Behavioral Processes, 46*, 25-38.

Xuan, B., Zhang, D., He, S., & Chen, X. (2007). Larger stimuli are judged to last longer. *Journal of Vision, 7*, 1-5.