Bulletin of the Psychonomic Society
1992, 30 (5), 425-427

An experimental comparison of the weighted up-down method and the transformed up-down method

THOMAS H. RAMMSAYER
University of Giessen, Giessen, Germany

As a simple procedure for adaptive testing, Kaernbach (1991) proposed the weighted up-down method (WUDM). In a temporal discrimination task with 24 subjects, differences between the WUDM and the transformed up-down method (TUDM) were evaluated. The results of the present experiment confirm Kaernbach's claim, based on computer simulations, that the WUDM is more efficient than the TUDM.

In a recent paper, Kaernbach (1991) introduced the weighted up-down method (WUDM) as a simple procedure for adaptive testing. The WUDM can be seen as a modification of Derman's (1957) version of the simple up-down procedure. As with the simple up-down procedure, the WUDM follows the general principle of decreasing the signal level after a correct response and increasing the signal level after an incorrect response. However, unlike the simple up-down procedure, the WUDM allows a different step size for the upward steps (S_{up}) than for the downward steps (S_{down}). According to Kaernbach, "the equilibrium condition for convergence point X_p is

$$S_{up}\,p = S_{down}\,(1-p).$$

For X_{75}, it follows that $S_{up}/S_{down} = 1/3$. The rule for a convergence to the X_{75} point would thus read: Decrease the Level 1 step after each correct response, and increase it 3 steps after each incorrect response" (p. 227).

Unlike the simple up-down method, which is designed to converge to the X_{50} point of a psychometric function, the transformed up-down method (TUDM) represents a more versatile procedure for estimating points on a psychometric function (Levitt, 1971). In contrast to the WUDM, the TUDM uses the same step size for upward and downward steps. Also, with the TUDM, changes in stimulus level depend on the outcome of a sequence of observations. Furthermore, there is a certain number of convergence points for the TUDM, whereas with the WUDM any desired target level can be used.

In a Monte Carlo simulation (Kaernbach, 1991), the WUDM (converging to a target level of X_{75}) proved to be slightly more efficient than the TUDM (converging to a target level of $X_{70.7}$). That is, for a given level of precision, fewer trials were required with the WUDM and, furthermore, the WUDM appeared to be more stable in that it was less prone to deviations from the optimal step size for adjustment of the testing level (Kaernbach, 1991).

The purpose of the experiment reported here was to evaluate the differences between the two adaptive procedures by comparing human subjects' threshold estimates obtained via the WUDM and the TUDM. For this purpose, 24 subjects were tested twice with an auditory temporal discrimination task. This task allowed for a comparison of the difference threshold estimates obtained by both procedures. The idea was to be closely parallel to Kaernbach's (1991) simulation study. Therefore, the same convergence points as in his study were used: $X_{70.7}$ for the TUDM and X_{75} for the WUDM.

METHOD

Subjects

Subjects were 12 male and 12 female students ranging in age from 22 to 33 years (M=23.8, SD=2.7). All of these subjects had normal hearing and had no previous experience with temporal discrimination tasks.

Apparatus and Stimuli

Stimulus presentation and recording of the subjects' responses were controlled by an IBM-AT-compatible computer. The stimuli consisted of filled auditory intervals that were generated by a computer-controlled sound generator. The frequency of the auditory intervals was 1000 Hz, and the intensity was 67 dB.

Procedure

The intervals were presented through headphones (Vivanco Model SR85). An experimental session consisted of one block with the TUDM and one block with the WUDM; the order of blocks was counterbalanced across subjects. Each block consisted of 50 trials, and each trial consisted of two stimuli, one 50-msec standard interval and one comparison interval. The comparison interval varied in duration from trial to trial depending on the subject's previous responses according to either the TUDM or the WUDM, which converge on a probability of hits of 70.7% and 75%, respectively. These convergence levels were chosen because they were the ones that Kaernbach (1991) used in his computer simulations.

I thank Susan D. Lima and Rolf Ulrich for helpful discussions. Correspondence should be addressed to T. H. Rammsayer, Department of Psychology, University of Giessen, Otto-Behaghel-Str. 10F, D-6300 Giessen, Germany.

With the TUDM, the duration of the comparison interval was decreased by step size S after two correct responses and increased by step size S after each incorrect response. Based on the results of prior studies, the duration of the comparison interval changed with a constant step size of S=8 msec for Trials 1-10, S=4 msec for Trials 11-30, and S=2 msec for Trials 31-50. The results of the prior studies indicated that these values provided the optimal strategy for using the TUDM to estimate threshold values in the temporal discrimination task. The rule for convergence to the 75%-difference threshold with the WUDM, given a basic step size of S=8 msec, resulted in a downward step size of S_{down} = 2 msec and an upward step size of S_{up} = 6 msec. According to the outcome of a pilot study using 8 subjects, this basic step size for the WUDM proved to be the best choice when no information on the psychometric function of a subject is available. The initial difference in duration between the standard and comparison intervals was chosen to allow the threshold region to be reached within 10 trials under optimal conditions. This consideration resulted in an initial comparison-interval duration of 70 msec for the WUDM and 98 msec for the TUDM.

Each subject was seated at a table with a keyboard and a computer monitor in a sound-attenuated room. To start a trial, the subject pressed the space bar; the auditory presentation began 900 msec later. Two auditory intervals were presented with an interstimulus interval of 900 msec. The subject's task was to decide which of the two intervals was longer and to indicate his/her decision by pressing one of two designated keys on the keyboard. After each response, visual feedback (+ = correct; − = false) was displayed on the screen. The next trial was started by pressing the space bar again. For each block, the subjects were given 10 practice trials. After the practice trials, the subjects were asked whether they understood the procedure. No subject requested additional practice.

As a measure of performance, mean differences between standard intervals and comparison intervals were computed for the three decades of trials starting with Trials 21-30 and ending with Trials 41-50. Data from Trials 1-20 were not analyzed because the initial stimulus level (difference between standard and comparison intervals) was far above the threshold range and, therefore, did not contribute to a threshold estimate. Furthermore, differences in performance between the TUDM and the WUDM for the early trials could not be compared because of the dissimilar initial values of the comparison intervals, which were chosen because of the different step sizes used by the TUDM and the WUDM.

RESULTS AND DISCUSSION

Threshold estimates calculated separately for Trials 21-30, 31-40, and 41-50 were 14.6, 12.8, and 10.8 msec for the TUDM and 10.1, 10.8, and 9.4 msec for the WUDM. These data are presented in Figure 1. A two-way analysis of variance with method (TUDM and WUDM) and decade (Trials 21-30, 31-40, and 41-50) confirmed a significant main effect of method [$F(1,23)$ = 4.57, $p < .05$] and a significant main effect of decade [$F(2,46)$ = 5.86, $p < .01$]. These findings indicate differences in threshold values between the TUDM and the WUDM and a significant progression toward lower threshold values as more trials are completed. A Tukey *HSD* test revealed a significant decrease in threshold values for Trials 41-50 as compared to Trials 21-30 ($p < .05$). However, there was no significant interaction between method and decade [$F(2,46)$ = 2.32, n.s.]. Inspection of the data presented in Figure 1 suggests that the main effect of method is primarily due to the high threshold estimates for Trials 21-30 with the TUDM. When only threshold estimates based on Trials 41-50 were compared, no difference was found [$t(23)$ = 1.11, n.s.].

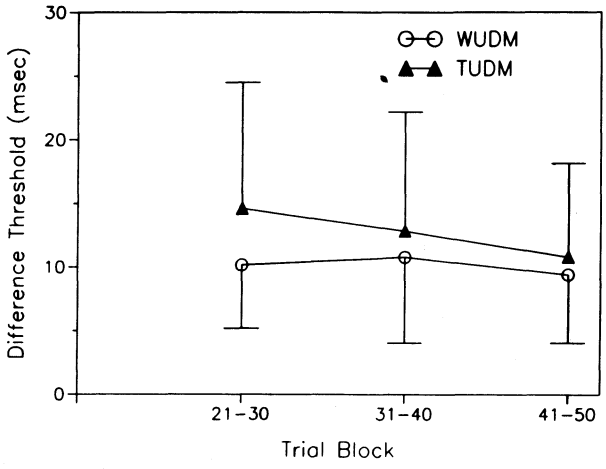

Figure 1. Threshold estimates and standard deviations calculated separately for Trials 21-30, 31-40, and 41-50.

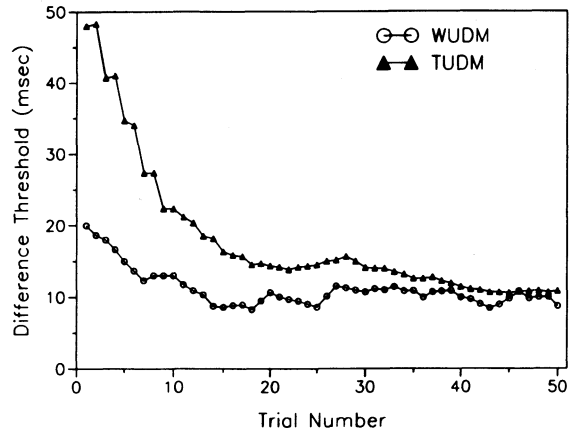

Figure 2. Average threshold estimates based on mean differences between comparison and standard interval plotted as a function of trial number.

To evaluate the characteristic changes in stimulus presentation in each of the two procedures, mean differences between comparison and standard values and the corresponding standard deviations on each of the 50 trials were computed across subjects within each condition. In Figure 2, mean differences between comparison and standard value are plotted as a function of trial number for the TUDM and the WUDM. It can be seen that the WUDM converges to a target stimulus level within 20 trials, whereas for the TUDM, it takes approximately 40 trials until the decreasing trend in the threshold values ceases and a stable asymptotic level is reached. The slow convergence rate toward the target stimulus level with the TUDM cannot be attributed exclusively to the higher initial stimulus level as compared to the WUDM, because under optimal conditions, that is, no incorrect responses, a hypothetical difference threshold value of 0 msec could be reached after 16 trials even with the TUDM. Thus, the results of this experimental investigation suggest that

the convergence rate of the WUDM as compared to the TUDM is actually even faster than the predicted increase in speed of approximately 10% based on the Monte Carlo simulations by Kaernbach (1991).

A critical difference between the TUDM and the WUDM is the reduction of step size of the TUDM in the course of a testing session, as compared to the constant step size of the WUDM. Although a few trials may be saved by starting with a stimulus level close to the threshold region, it is helpful for a subject if the testing session begins with easy discriminations above the threshold level. With an initial testing level above the threshold region, the use of too small a step size necessitates too large a number of trials in converging to the threshold region. Furthermore, if neither the spread nor the location of the psychometric function is known, a large initial step size that decreases in the course of the experiment is recommended (Levitt, 1971). However, after reaching the threshold range, a smaller step size results in a more reliable and less variable threshold estimate. Therefore, reducing step size in the course of a testing session tends to improve the efficiency of an adaptive procedure. Based on pilot work, the variable step size of the TUDM provides a highly efficient strategy for estimation of threshold values in temporal discrimination tasks. However, for the WUDM, a further increase in efficiency by also using a variable step size cannot be ruled out by the present data. This points to the possibility of an even higher superiority of the WUDM as compared to the TUDM.

In the present experiment, between-subject variability was estimated by calculating the standard deviation between the subjects' stimulus levels on a trial-by-trial basis (see Figure 3). The initial step size of 8 msec for Trials 1-10 and a step size of 4 msec for Trials 11-30 resulted in greater standard deviations with the TUDM than with the WUDM. This effect leveled off after Trial 31, at which point the step size was reduced to 2 msec for the TUDM. Accordingly, from Trial 31 to Trial 50 the standard deviation decreased. Obviously, the WUDM benefits from the different step sizes for upward and downward steps: the smaller step sizes of the downward steps contribute to a smaller between-subject variability.

In conclusion, the results of the present experiment with human subjects confirm Kaernbach's (1991) claim, based on computer simulations, that the WUDM is more efficient than the TUDM. In addition, the WUDM has the advantage of converging to any desired target level, and the underlying rule is very simple. Therefore, it appears that in psychophysical testing, the WUDM is the preferred method. The only drawback of the WUDM observed in the present experiment was that because the rule for controlling the stimulus level was so simple and because the step size of an upward step was relatively large, all the subjects tested in the present study reported awareness of the sequential rule applied with the WUDM, whereas not a single subject reported awareness of any kind of rule when tested with the TUDM. However, this problem could be easily solved by randomly interleaving trials from two WUDM series. For example, a series of trials converging to X_{60} could be interleaved with a series of trials converging to X_{90}. This method should eliminate the problem of subject awareness of the sequential rule applied with the WUDM.

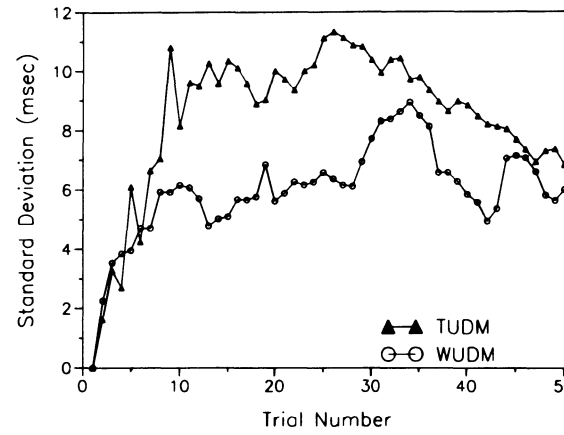

Figure 3. Average between-subject standard deviation plotted as a function of trial number.

REFERENCES

DERMAN, C. (1957). Non-parametric up-and-down experimentation. *Annals of Mathematical Statistics*, **28**, 795-797.

KAERNBACH, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics*, **49**, 227-229.

LEVITT, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, **83**, 1852-1862.

(Manuscript received May 20, 1992.)