Perception & Psychophysics
1991, 49 (3), 227-229

Simple adaptive testing with the weighted up-down method

CHRISTIAN KAERNBACH
Unité Inserm, Bordeaux, France

This paper proposes a method for adaptive testing that is less complicated than the commonly used transformed up-down methods (1 up 2 down, 1 up 3 down, etc.). In addition, the weighted up-down method can converge to any desired point of the psychometric function. The rule is very simple: Each correct response leads to a decrease in signal level, each incorrect response to an increase. The only difference from the simple up-down method (1 up 1 down) is that the steps upward and the steps downward are of a different size. The straightforward construction of the novel procedure pays off in efficiency and stability: A Monte Carlo simulation reveals a definite advantage, though small, of the weighted up-down method over the 1-up-2-down rule.

The *simple up-down method* (1 up 1 down) converges to the X_{50} point of the psychometric function. This is not appropriate for tasks in which the chance performance is high. For two-interval forced-choice (2IFC) tasks (chance performance: 50%), X_{75} would be the halfway point. This point would be the most natural choice for a threshold estimate.

With *transformed up-down methods* (Levitt, 1971), level changes depend on the outcome of two or more of the preceding trials. For instance, the level is increased with each incorrect response and decreased after two successive correct responses (1 up 2 down, or the 2-step rule). The steps upward and the steps downward are of equal size. For each rule, there exists a distinct convergence point (e.g., $X_{70.7}$ for the 2-step rule, $X_{79.4}$ for the 3-step rule). Unfortunately, there exists no transformed up-down rule for X_{75}.

In many studies, transformed up-down methods have been compared with other adaptive procedures (see, e.g., Kaernbach, 1990, or Kollmeier, Gilkey, & Sieben, 1988, and references cited therein). In the present paper, I propose a novel modification of the simple up-down method that is much simpler than transformed up-down methods. Furthermore, it is more versatile and slightly more efficient than transformed up-down methods.

The restriction to equal step sizes for both directions is undesirable. If one drops it, the resulting procedure becomes very simple. The *weighted up-down method* proposed here can converge to any desired point on the psychometric function. In the first section, I will derive this algorithm and discuss its construction in relation to transformed up-down methods. In the next section, I compare the efficiency and the optimal step size of the 2-step rule with the weighted up-down method for X_{75}, by means of a Monte Carlo simulation.

I would like to thank Laurent Demany for his helpful suggestions and an anonymous reviewer for the reference to Tyler and Gorea (1986). Correspondence should be addressed to Christian Kaernbach, Laboratoire d'Audiologie Expérimentale, INSERM unité 229, Université de Bordeaux II, Hôpital Pellegrin, Place Amélie Raba Léon, 33 076 Bordeaux Cedex, France.

THE WEIGHTED UP-DOWN METHOD

The rule is quite similar to the simple up-down rule: Each correct response leads to a decrease in signal level, each incorrect response to an increase. But the step size, S_{up}, for upward steps may now differ from the step size, S_{down}, for downward steps. The equilibrium condition for convergence point X_p is

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

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. It is illuminating to compare this rule with the 3-step rule. Let us denote a sequence of trials as, for example, $++-$, with $+$ standing for a correct response and $-$ standing for an incorrect response. With transformed up-down methods, the possible outcomes of trials or series of trials are categorized into two groups: the down group (leading to a decrease in signal level) and the up group (leading to an increase). For the 3-step rule, the down group contains $+++$, and the up group contains $++-$, $+-$, and $-$. Let us assume that the actual step size is 3 dB. The subsequent adjustment would then follow column A of the scheme in Table 1.

It does not seem fair to treat $++-$ just as $-$, that is, not to acknowledge the first two correct responses of $++-$. It would be more logical to react as in column B: Sequences of the up group lead to a level increase that depends on the number of correct responses included. Column B corresponds to the reaction of the weighted up-down method for X_{75} ($+$ counts -1 dB, and $-$ counts $+3$ dB). And whereas the 3-step rule waits for the completion of a trial series, the weighted up-down method reacts immediately, because the level adjustments directly

Table 1
Comparison of Transformed and Weighted Up-Down Rules

Trials	Group	A	B
+++	down	−3 dB	−3 dB
++−	up	+3 dB	+1 dB
+−	up	+3 dB	+2 dB
−	up	+3 dB	+3 dB

follow the trials that have set them off. This straightforward structure makes the procedure faster and more stable.

An anonymous reviewer brought my attention to the article of Tyler and Gorea (1986), who used a quite similar method. Incorrect responses led to an upward step, and correct responses had a 33% chance to lead to a downward step. Sixty-seven percent of the correct responses led to no change in signal level. Tyler and Gorea considered their method comparable to "1 up 3 down," and they concluded that it should converge to 79% correct responses. In reality, however, it is comparable to the weighted up-down method described above, and it converges to 75% correct responses. The additional arbitrary randomness should make it slightly less efficient. It could be an alternative in situations in which fractional steps are hard to produce.

MONTE CARLO SIMULATION

The psychometric function was assumed to be a tanh-like function. It leveled off at 1.0 for high intensities and at 0.5 for low intensities (see Figure 1):

$$p(X) = [e^X/(e^X+e^{-X}) + 1]/2. \qquad (2)$$

This could, for example, describe the percentage of correct responses in a 2IFC experiment. The spread of this psychometric function could be defined as $X_{90}-X_{60}$, a difference that would then amount to 1.4. The signal level was set at 2.5 (p = .9967) at the beginning of each track. The first two reversals were discarded. At every even number of reversals, the median intensity of the reversal points was calculated. These so-called midrun estimates were used here as they are used commonly. Any other analysis (e.g., mean of the reversals) would work as well. After 24 further reversals, the track was stopped. Ten thousand tracks were simulated for each condition. The simulation procedure is nearly identical to that used by Kaernbach (in press), which was verified by comparison with experimental data.

The trial number n was set to 8 at the last discarded trial—that is, at the second reversal. This was done to take into account the initial phase. In the inital phase, the rules are usually modified, and other step sizes are applied. A careful choice of the starting level should allow for reaching the threshold region (more precisely, the second reversal) within 8 trials.

To evaluate the efficiency of an adaptive procedure, one must determine the error as a function of the trial number. The error consists of two parts: the statistical error, corresponding to the fluctuations of the estimates around their mean value, and the systematic error, corresponding to the systematic deviations of the mean values from the convergence level, generally in the direction of the starting point. The total error is equal to the orthogonal sum: $E^2_{total} = E^2_{stat} + E^2_{sys}$. It behaves at first approximation like $1/\sqrt{n}$. The normalized total error $E_{total}(n)\sqrt{n}$ moves much less as function of n. This construction differs from the sweat factor used by some authors (see, e.g., Taylor & Creelman, 1967), in that it does not claim to be independent from the track length and covers the systematic error too. It is convenient for the comparison of slight differences in the efficiency of adaptive procedures.

Figure 2 shows the normalized total error as a function of the trial number n. The solid lines correspond to the weighted up-down rule for X_{75}, and the dotted lines correspond to the 2-step rule. For the weighted up-down method, the step sizes given in the figure legend correspond to the geometric mean S_w of S_{up} and S_{down}:

$$S_w = \sqrt{S_{up}\, S_{down}}\,, \qquad S_{up} = S_w\sqrt{3}, \quad S_{down} = S_w/\sqrt{3}. \qquad (3)$$

For both procedures, the optimal step size is 0.23 (circles). This is about one sixth of the spread of the psycho-

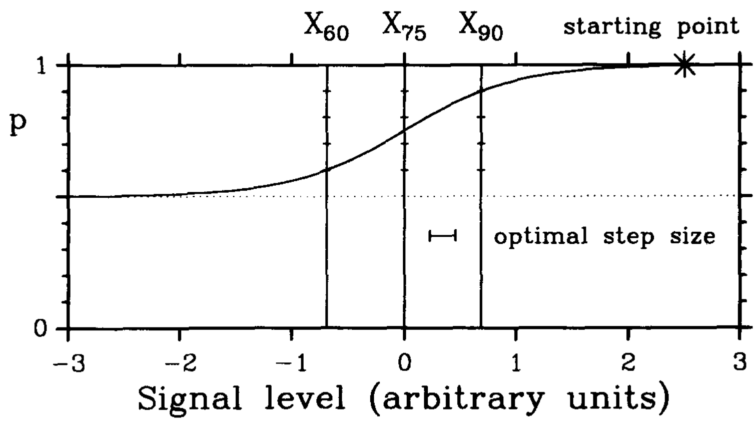

Figure 1. The psychometric function of the simulation model. The spread can be defined as $X_{90}-X_{60}$. The starting point was at 2.5.

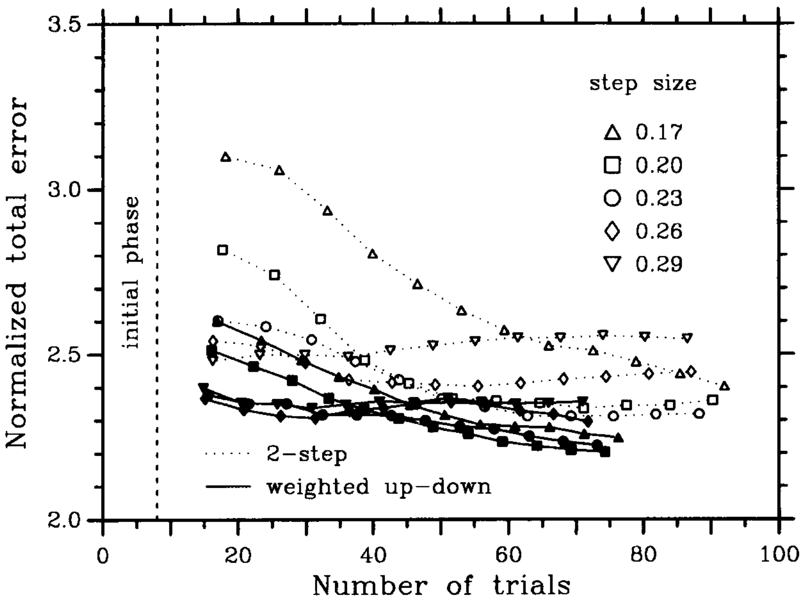

Figure 2. The normalized total error, as a function of the trial number. The first two reversals were discarded. The single data points represent stop criteria after 2, 4, 6, ... 24 further reversals. The trial number corresponds to the mean number of trials required to get the respective number of reversals.

metric function (see Figure 1). Smaller step sizes produce markedly more errors for short tracks, which fail to reach the threshold region (systematic error). Bigger step sizes are ideal for short tracks, because they reach the threshold region quickly, but they compare badly for long tracks (random walk). Step sizes bigger than optimal are less critical than step sizes that are too small. With the optimal step size, the weighted up-down method produces about 5% less error; or, for a given precision, it needs about 10% less experimentation time than does the 1-up-2-down method. Moreover, the novel method is more stable: The deviations from the optimal step size in the range of about 30% are much better dealt with by the weighted up-down method.

CONCLUSION

The weighted up-down method is markedly simpler than the transformed up-down methods. This makes it easy to implement. In addition, the level changes depend only on the outcome of the last trial. This may ease the implementation of interleaving tracks, and—if they are given feedback—this may help the subjects to understand the tracks better. Finally, the uncomplicated construction of this procedure pays off with a faster and more stable convergence toward the desired point on the psychometric function.

The weighted up-down method is not restricted to forced-choice tasks. For instance, just like 2IFC tasks, a series of yes/no tasks with 50% noise presentations will allow for a chance performance (a probability of correct answer) of 50%. The weighted up-down method, then, corresponds to the symmetric SIAM procedure described by Kaernbach (1990).

An experimental verification of the efficiency gain of the weighted up-down rule has yet to come. Since this rule does not introduce essential new elements to adaptive psychophysics, but only reduces the complexity of the applied rule, it is to be expected that human subjects would work at least as efficiently with it as with transformed up-down methods. The efficiency gain is anyhow not extreme (5% less error, or 10% more speed), so that the conceptual advantages of the weighted up-down rule are more important.

REFERENCES

Kaernbach, C. (1990). A single-interval adjustment-matrix (SIAM) for unbiased adaptive testing. *Journal of the Acoustical Society of America*, **88**, 2645-2655.

Kollmeier, B., Gilkey, R. H., & Sieben, U. K. (1988). Adaptive staircase techniques in psychoacoustics: A comparison of human data and a mathematical model. Journal of the Acoustical Society of America, **83**, 1852-1862.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, **49**, 467-477.

Taylor, M. M., & Creelman, C. D. (1967). PEST: Efficient estimates on probability functions. *Journal of the Acoustical Society of America*, **41**, 782-787.

Tyler, C. W., & Gorea, A. (1986). Different encoding mechanisms for phase and contrast. *Vision Research*, **26**, 1073-1082.

(Manuscript received June 7, 1990;
revision accepted for publication September 21, 1990.)