Perception & Psychophysics
2007, 69 (2), 254-262

Small-sample characterization of stochastic approximation staircases in forced-choice adaptive threshold estimation

LUCA FAES
Università di Trento, Povo, Trento, Italy

GIANDOMENICO NOLLO AND FLAVIA RAVELLI
Università di Trento, Povo, Trento, Italy
and ITC, Trento, Italy

AND

LEONARDO RICCI, MASSIMO VESCOVI, MASSIMO TURATTO,
FRANCESCO PAVANI, AND RENZO ANTOLINI
Università di Trento, Trento, Italy

Despite the widespread use of up–down staircases in adaptive threshold estimation, their efficiency and usability in forced-choice experiments has been recently debated. In this study, simulation techniques were used to determine the small-sample convergence properties of stochastic approximation (SA) staircases as a function of several experimental parameters. We found that satisfying some general requirements (use of the accelerated SA algorithm, clear suprathreshold initial stimulus intensity, large initial step size) the convergence was accurate independently of the spread of the underlying psychometric function. SA staircases were also reliable for targeting percent-correct levels far from the midpoint of the psychometric function and performed better than classical up–down staircases with fixed step size. These results prompt the utilization of SA staircases in practical forced-choice estimation of sensory thresholds.

The measurement of sensory thresholds, which is intimately related to the estimation of specific points of the psychometric function underlying the perception of sensory stimuli, is a very common practice in psychophysics. For this reason, in recent decades, a variety of adaptive techniques has been developed to increase efficiency and accuracy of threshold measurement with respect to the pioneer method of constant stimuli (Fechner, 1860). These adaptive psychophysical procedures are excellently reviewed in the papers by Treutwein (1995) and Leek (2001). Among adaptive procedures, nonparametric staircases are largely used by psychophysicists in place of parametric Bayesian or maximum-likelihood techniques, mainly thanks to their simplicity and intuitiveness. Moreover, in the experimental design of adaptive procedures, forced-choice experiments are often preferred to yes–no paradigms, since they constitute a criterion-free approach to threshold estimation (Kershaw, 1985; Macmillan & Creelman, 1991). For these reasons, two-alternative forced-choice (2AFC) staircases are of widespread use in practical psychophysical experiments (García-Pérez, 1998; Leek, 2001). In up–down staircases, the stimulus intensity for the next trial is increased or decreased upon observing a subset of subject responses, according to a rule that determines the probability of correct response targeted by the procedure (Levitt, 1971; Wetherill & Levitt, 1965). The use of complicated up–down rules (Brown, 1996) and of weighted up–down approaches (Kaernbach, 1991) should theoretically allow one to target almost any performance level on the psychometric function.

Despite the extensive diffusion of up–down staircases, their effective convergence to the expected target probability has been recently debated. Indeed, the simulation studies of García-Pérez (1998, 2002) have demonstrated that neither the traditionally used transformed and weighted up–down staircases (Kaernbach, 1991; Wetherill & Levitt, 1965) nor some recent modifications of these methods (Kaernbach, 2001a; Zwislocki & Relkin, 2001) yield threshold estimates that correspond to the presumed probability of correct response. More importantly, the simulations in García-Pérez (1998, 2002) showed that unless particular conditions are met, the actual probability that is targeted varies greatly with the relative size of the placed intensity steps with respect to the spread of the underlying psychometric function. These asymptotic and small-sample findings should make 2AFC up–down staircases generally unfit for

L. Faes, faes@science.unitn.it

threshold estimation, since the spread of the psychometric function is an unknown parameter that often varies across subjects and stimuli. Thus, the same staircase may converge to different target probabilities as a consequence of different experimental conditions. Moreover, the particular parameter settings allowing for the implementation of traditional and modified up–down staircases severely limit the range of probability levels that can be targeted with sufficient reliability (García-Pérez, 1998, 2002).

The unreliability of 2AFC staircases might be related to the fact that in all these procedures, the upward and downward size of the stimulus intensity steps are set to fixed values. Consequently, only a limited number of points on the psychometric function can be visited during the development of the staircase. Hence, the utilization of adaptive procedures in which the step size is varied during the experiment could help mitigate the dependence of the target probability on the spread of the psychometric function. In this context, the method of stochastic approximation (SA) is a sequential up–down procedure in which the step size progressively decreases as successive trials are collected. The SA sequence has good asymptotic properties: Robbins and Monro (1951) formally proved that it converges in probability to the threshold intensity that corresponds to any predefined target probability. With regard to psychophysical applications, although the possibility of using SA staircases was mentioned by some authors (Kaernbach, 2001b; King-Smith, Grigsby, Vingrys, Benes, & Supowit, 1994) and explicitly advocated for adaptive threshold estimation (Treutwein, 1995), a thorough evaluation of the small-sample convergence properties of this adaptive method has not yet been provided. This might be a reason why this approach has been so far ignored in practical psychophysical applications.

The aim of the present study is to provide a validation of the convergence of adaptive SA staircases as a function of the various parameters that may affect the accuracy of threshold estimation in the small-sample setting that is typical of psychophysical applications. For this purpose, the empirical distribution of the estimated target probability is first characterized by means of simulations running SA staircases of limited length (up to a few hundred trials) and then compared with the expected target probability in order to evaluate the convergence error. The experimental conditions tested by the simulation approach include different values of the staircase length, the starting stimulus intensity, the expected target probability, and the spread of the imposed underlying psychometric function. Two different implementations of the SA algorithm and different criteria for estimating the threshold at the end of the staircase are considered. The analysis of the convergence properties over this broad range of conditions allows for the assessment of practical recommendations for the SA staircase utilization in common psychophysical applications.

METHOD

Stochastic Approximation Staircases

The algorithm of SA accomplishes the following stepping rule for the intensity of the stimulus presented during a psychophysical task:

$$x(n+1) = x(n) - \frac{C}{n}\left[z(n) - \Phi\right], \tag{1}$$

where n is the number of trials that have occurred since the beginning of the experiment; $x(n+1)$ and $x(n)$ are the intensities at the $(n+1)$th and nth steps, respectively; Φ is the target probability, C the initial step size, and $z(n)$ a binary quantity, depending on the response at the nth trial as follows: z is equal to 1 in case of success and 0 in case of failure. Consequently, the stimulus level is decremented with a step size $\delta = C(1 - \Phi)/n$ in case of a correct response, or conversely incremented with a step size $\delta = C\Phi/n$ in case of an incorrect response. Since the step size δ is inversely related to the number of trials, both increments and decrements become smaller while running the experiment. Robbins and Monro (1951) proved that for any value of the probability of correct response (target probability) Φ in the range (0, 1), the SA staircase of Equation 1 converges with probability 1 to the stimulus intensity corresponding to Φ.

Kesten (1958) proposed a modification of the SA method named *accelerated SA*, in which the step size is varied only when a shift in response category occurs:

$$x(n+1) = x(n) - \frac{C}{n_{\text{shift}} + 1}\left[z(n) - \Phi\right]. \tag{2}$$

In Equation 2, n_{shift} is the number of reversal points in the staircase—that is, the number of trials at which the response changes from correct to incorrect or vice versa. Kesten proved that the sequence of Equation 2 converges more rapidly than the traditional SA of Equation 1.

After completion of a normal or accelerated SA staircase of N trials, two reasonable criteria to estimate the threshold intensity x_T corresponding to the target probability Φ are either to take the next stimulus level that would have been presented after the last performed trial—that is, $\hat{x}_T = x(N+1)$ (Anderson & Johnson, 2006)—or to average the last M collected stimulus intensities ($M > 1$)—that is,

$$\hat{x}_T = M^{-1}\sum_{m=0}^{M-1} x(N - m + 1).$$

Simulation Approach

In this study, the behavior of SA staircases was described by computer simulations in which the stimulus level was varied in accordance with the staircase rules, the predefined staircase parameters, and the response of a simulated observer in a 2AFC psychophysical experiment. The probability of a correct response of the simulated observer to a stimulus level x was modeled by a Weibull psychometric function (Treutwein, 1995),

$$\psi(x) = 0.98 - 0.48 \cdot e^{-0.736\left(\frac{x}{\mu}\right)^{\frac{2.265}{\sigma}}}, \tag{3}$$

where μ is the location of the psychometric function defined as the stimulus returning a 75% probability of correct response and σ is the spread of the function defined as the extent (expressed in logarithmic units) over which it displays a nonasymptotic behavior. The choice of the constants in Equation 3 has the following consequences: The low and high asymptotic probabilities of correct response are equal to .5 (so that a typical 2AFC task is reflected) and .98 (so that the probability of lapsing is .02), respectively; the psychometric function is .75 for $x = \mu$; the spread σ is defined as the difference between intensity values corresponding to 1% deviation from the asymptotic probabilities. A detailed calculation of these constants and an explanation of their relation to the parameters of the Weibull psychometric function is reported in García-Pérez (1998). Figure 1 shows examples of psychometric functions computed by means of Equation 3 assuming the same location and three different spread values. With these psychometric functions, SA staircases implemented according to Equations 1 and 2 converge to the stimulus intensity that corresponds to any target probability Φ belonging to the open interval between the low and high asymptotes of the curves—that is, $\Phi \in (.5, .98)$.

In each simulation run, the responses of the simulated observer were collected as Bernoulli trials with probability of success equal to the psychometric function sampled at the presented stimulus intensity. Specifically, at the nth trial, $1 \le n \le N$, the stimulus level $x(n)$ was entered in Equation 3 to obtain the probability $\psi[x(n)]$, which in turn was compared with a random number $u(n)$ drawn from uniform distribution in the range (0, 1) to obtain the nth simulated response according to the rule

$$z(n) = \begin{cases} 0 \text{ if } u(n) > \psi[x(n)], \\ 1 \text{ if } u(n) \le \psi[x(n)]. \end{cases} \quad (4)$$

The simulated binary responses were then used in Equations 1 or 2 according to the staircase settings in order to determine the stimulus level for the next trial, $x(n+1)$. After completion of the staircase, estimates $\hat{x}_T$ of the threshold intensity x_T corresponding to the target probability $\Phi = \psi(x_T)$ were evaluated according the two criteria mentioned above. Finally, the estimated target probability $\hat{\Phi} = \psi(\hat{x}_T)$ was compared with the expected value Φ to quantify the accuracy of convergence.

Simulation Design

In order to provide a complete description of the behavior of SA staircases, the simulations were performed by varying the relevant parameters that may affect the small-sample convergence. First of all, the possible effects of the main features of the psychometric function underlying the experiment were explored by considering three different values for the spread of the function ($\sigma = 0.5$, 1, 2 log units; see Figure 1). The spread of the psychometric function is related to the convergence of the staircase through the size of the intensity steps placed during the experiment. So, in all simulations, the initial step size C was expressed in terms of the parameter σ and was varied to obtain predefined values from 0.1 to 1.5, step 0.1, for the ratio C/σ. Henceforth, for the sake of brevity, the term *relative step size* will correspond to the ratio C/σ between the initial step size C and the spread σ of the underlying psychometric function. The location of the psychometric function may influence the small-sample convergence since it basically determines the distance (in intensity units) between the initial stimulus $x(1)$ and the threshold stimulus x_T. In our simulations, we set the location of the psychometric function at $\mu = 10^{-1.5}$ and studied its effects on convergence by varying the starting intensity $x(1)$ within the set {1, 0.1, 0.01, 0.001}. Furthermore, to characterize the small-sample drift of SA staircases, conditions with starting value at the convergence point (i.e., $\psi[x(1)] = \Phi$) were also considered. Three different expected target probabilities ($\Phi = .65$, $\Phi = .75$, $\Phi = .85$) were chosen and the SA staircase was run for each according to both normal and accelerated procedures (Equations 1 and 2). Stimuli beyond boundaries were treated according to the carry-on criterion: The staircase proceeded as if there were no boundaries, but off-limits values were presented at boundary levels. To study the small-sample convergence properties, the length N of the simulated staircases was varied within the set {10, 20, 30, 40, 50, 60, 80, 100, 125, 150, 175, 200, 250, 300}. With regard to threshold calculation, the mean of the last M values—with M within the set {1, 5, 10, 20, 30, 40, 50, 60, 70, 80}—was taken as the estimate of the final threshold intensity. Both the updating of the stimulus intensities during the staircase and the final threshold estimation were performed with stimuli expressed in logarithmic units.

Figure 2 shows a representative sample of $N = 50$ trials of a normal and an accelerated SA staircase with initial step size $C = 1$, initial level $x(1) = 0.1$, target probability $\Phi = .75$, and run over a psychometric function with $\mu = -1.5$ log units and spread $\sigma = 1$ log unit. For selected combinations of the parameters described above, 1,000 independent realizations of the simulation were generated. Specifically, the convergence of SA staircases was tested as a function of the following parameters: starting intensity $x(1)$ with $N = 100$ and $M = 1$ (Figure 3); staircase length N, with $M = 1$ and $x(1)$ such that $\psi[x(1)] = \Phi$ (Figures 4 and 5); and number M of final intensities averaged to estimate the threshold, with $N = 100$ and $x(1)$ such that $\psi[x(1)] = \Phi$ (Figure 6). In all cases, the complete ranges for the parameters σ, C/σ, and Φ were explored. The distribution of the resulting estimated target probability $\hat{\Phi}$ was characterized in terms of median and interquartile range (*IQR*, defined as the difference between the 75th percentile and 25th percentile of the distribution) and compared with the expected target probability Φ.

To provide a direct comparison between the proposed SA staircase and the traditional fixed step size (FSS) staircases largely used in psychophysics, we simulated on the same psychometric function ($\mu = -1.5$ log units, $\sigma = 1$ log unit) staircases of 100 trials by using the accelerated SA method and the traditional transformed up–down method (Wetherill & Levitt, 1965). Specifically, we considered the three-down/one-up FSS staircase with equal step size that is claimed to converge at the 79.37%-correct point of the psychometric function (Levitt, 1971) and the transformed and weighted three-down/one-up FSS staircase having a step-down to step-up ratio equal to 0.734, which has been demonstrated to have an asymptotically stable convergence point at 83.15% correct responses (García-Pérez, 1998). To allow comparison, in the two cases the expected target probabilities of the accelerated SA method were set to $\Phi = .7937$ and $\Phi = .8315$, respectively. In both cases, the initial stimulus intensity was $x(1) = 1$. Both the initial step size C of the SA staircase and the upward step size Δ^+ of the FSS staircase were varied from 0.1 to 1.5. The final threshold estimate for the FSS staircase was obtained by averaging the last 20 reversal points of the staircase. As was the case in the other simulations, 1,000 realizations of each staircase were generated and the results were presented in terms of median and IQR of the estimated target probability (Figure 7).

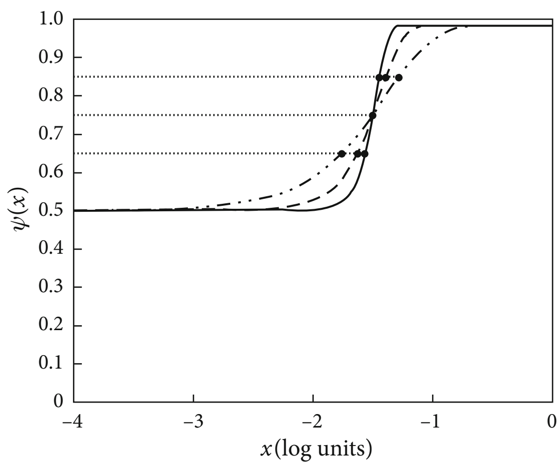

Figure 1. Weibull psychometric functions used in the simulations. The three functions are all located at $\mu = 10^{-1.5}$ with respect to a probability of correct response equal to .75, and have spread $\sigma = 0.5$ (continuous line), $\sigma = 1$ (dashed line), and $\sigma = 2$ (dash-dotted line) log units. The circles indicate the points targeted on the psychometric functions, corresponding to the three target probabilities $\Phi = .65$, $\Phi = .75$, and $\Phi = .85$ (dotted horizontal lines).

RESULTS

All the analyses presented in this section were repeated for the three values chosen for the spread of the psychometric function underlying the simulated experiments—that is, $\sigma = 0.5$, 1, and 2 log units. We found that the convergence of small-sample staircases was substantially unaffected by the specific value of the spread used in the simulations. Indeed, as far as the initial step size C was set

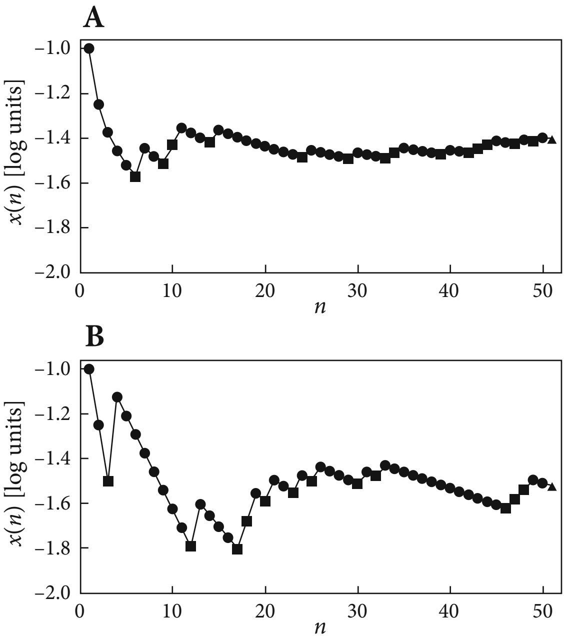

Figure 2. Example of simulated stochastic approximation staircases where the stimulus intensity $x(n)$ at the trial n was set according to the normal (A) and accelerated (B) algorithm. Circles and squares correspond to correct and incorrect responses, respectively. Triangles at the end of the staircases indicate the next stimulus level that would have been presented if the procedure had not terminated. See text for details on the selection of parameters.

to obtain the same value for the relative step size C/σ, the behavior of the estimated target probability was the same for any value of the spread σ. Thus, henceforth, we will present only the results for $\sigma = 1$.

Staircase Type

Figure 3 shows the median over 1,000 simulation runs of the target probability estimated by the normal and the accelerated SA staircases as a function of the relative step size C/σ for different values of the initial stimulus intensity. The figure indicates that with the typical staircase length used ($N = 100$), there is a significant bias of the target probability at low relative step size values, which is then reduced by increasing the relative step size. However, the normal SA procedure failed in reaching the target probability unless particular conditions were met—for example, the initial stimulus was close to the convergence point [$x(1) = -1$ log units], and the relative step size was larger than unity. On the contrary, the accelerated version of the SA method was more efficient in reaching the target probability, with a bias that became negligible provided that the relative step size was sufficiently high ($C/\sigma \geq 0.4$) and that the initial stimulus intensity was set above target. Indeed, for starting values equal to 0 and −1 log units, all of the three investigated target probabilities were easily reached; on the other hand, when $x(1) = -2$ log units and $x(1) = -3$ log units, the negative bias was more persistent and in some conditions difficult to eliminate even with large relative step size. Figure 3 also shows a slight negative bias that—independent of the initial conditions—is present at high C/σ values. This bias, which seems to vary with the expected target probability, will be discussed in the next subsection.

Apart from the different dependence on the initial stimulus intensity depicted in Figure 3, the behavior of normal and accelerated SA staircases was substantially the same. Therefore, in the following subsections, we provide the characterization of the accelerated SA staircase, which was more efficient in reaching the target probability starting from different stimulus intensities.

Notwithstanding the above recommendations about the optimal values for initial stimulus intensity and initial step size, the convergence to the target region was obtained in any case, provided that a sufficiently long staircase was run. We evaluated the minimum number of trials required in order for the central tendency of the target probability to fall within a 5% confidence interval with respect to the expected value ($|\text{median}(\hat{\Phi}) - \Phi| \leq .05$) by using an accelerated SA staircase with the expected target probability $\Phi = .75$. Starting with initial intensities $x(1)$ equal to 0, −1, −2, and −3 log units, respectively, this evaluation yielded: 10, 5, 14, and 53 trials in case of $C/\sigma = 1.4$; 9, 4, 16, and 103 trials in case of $C/\sigma = 1$; 13, 6, 33, and 183 trials in case of $C/\sigma = 0.6$; 86, 22, 223, and 360 trials in case of $C/\sigma = 0.2$. Although it is clear that convergence is always achieved for a sufficiently

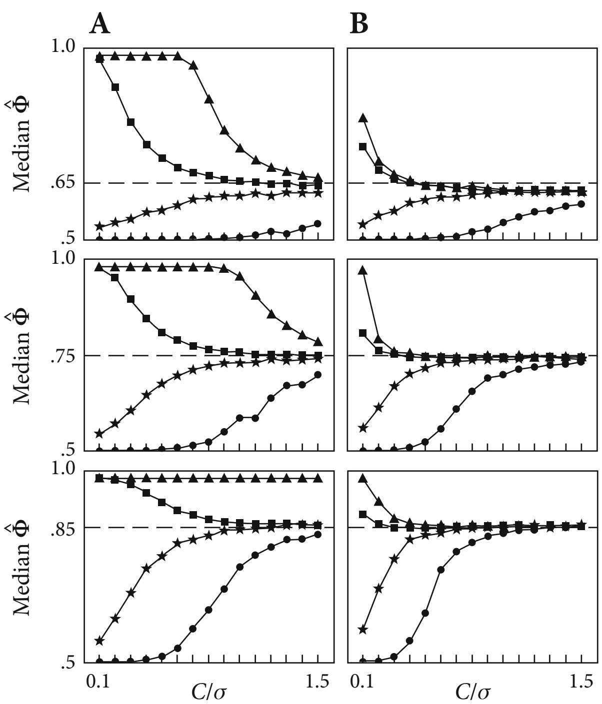

Figure 3. Median over 1,000 simulation runs (each lasting $N = 100$ trials) of the estimated target probability ($\hat{\Phi}$), plotted as a function of the relative step size C/σ for the normal (A) and accelerated (B) SA procedure and for different expected target probabilities: $\Phi = .65$ (up), $\Phi = .75$ (middle), and $\Phi = .85$ (down). The four curves in each graph correspond to different stimulus intensities tested at the first presentation: $x(1) = 1$ (triangles), $x(1) = 0.1$ (squares), $x(1) = 0.01$ (stars), and $x(1) = 0.001$ (circles).

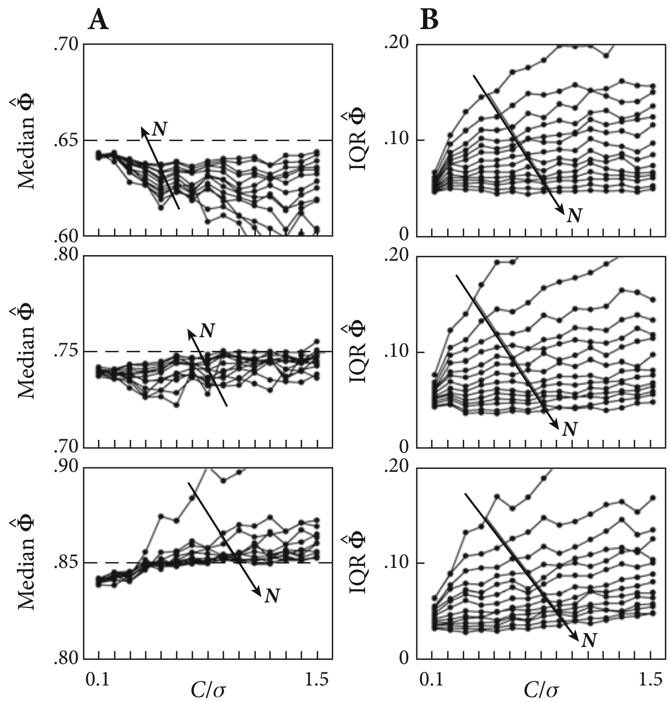

Figure 4. Median (A) and interquartile range (IQR; B) of the empirical distribution of the target probability $\hat{\Phi}$, estimated over 1,000 simulation runs (each starting at its convergence point) and plotted as a function of the relative step size C/σ for different expected target probabilities: $\Phi = .65$ (up), $\Phi = .75$ (middle), and $\Phi = .85$ (down). The curves in each graph correspond to different values of the staircase length (N = 10, 20, 30, 40, 50, 60, 80, 100, 125, 150, 175, 200, 250, 300), where shifts of the curves with increasing N are indicated by the arrows.

long staircase, these results confirm that a large initial step size and an initial stimulus intensity higher than the threshold lead to a faster convergence to the target.

Effects of Staircase Length

To investigate the effects of the staircase length on the convergence of the SA procedure, the central tendency (median) and variation (IQR) of the target probability were estimated over 1,000 realizations of staircases. These staircases started at the convergence point and lasted for a number of trials, varying from 10 to 300. The results for the accelerated SA procedure are shown in Figure 4. The analysis of the median revealed a slight negative bias for all three targeted probabilities, which was progressively reduced upon increase of the number of trials. Increasing the staircase length also contributed to decreasing the random component of the error, as documented by the reduction of the IQR of the estimated target probability. Variations of both median and IQR with respect to the relative step size C/σ were small and were further reduced by increasing the staircase length.

For any given staircase length, the estimates became more accurate when higher points of the psychometric function were targeted. This aspect can be seen in Figure 5, where the reduction of bias and variance corresponding to the increase of the estimated target probability from .65 to .75 and then to .85 can be easily appreciated. As an example, with the length $N = 100$, the negative bias and the variability averaged over all the tested C/σ values decreased while moving the target probability from $\Phi = .65$ (bias = −0.018, IQR = 0.077) to $\Phi = .75$ (bias = −0.007, IQR = 0.071) and then to $\Phi = .85$ (bias = −0.001, IQR = 0.061). As expected, with $N = 200$, the error was further reduced ($\Phi = .65$: bias = −0.012, IQR = 0.057; $\Phi = .75$: bias = −0.04, IQR = 0.051; $\Phi = .85$: bias = −0.0005, IQR = 0.045). In any case, Figure 5 indicates the overall good performance of the SA staircase in targeting separate points of the psychometric function, which results from the substantial independence of the target probability on the initial step size as well as the good separation achieved between the distributions of the different estimated target probabilities.

Criteria for Threshold Estimate

All the results presented in the previous subsections were obtained by taking the stimulus level that would have been presented after the last trial if the staircase had not terminated, $x(N + 1)$, as a final estimate of the threshold intensity. The accuracy of the threshold estimation was also assessed by taking the average of a given number M of intensities as a final estimate of the threshold intensity. As shown in Figure 6 for the accelerated SA staircase with $N = 100$ trials, increasing the number M led to a slight worsening of both

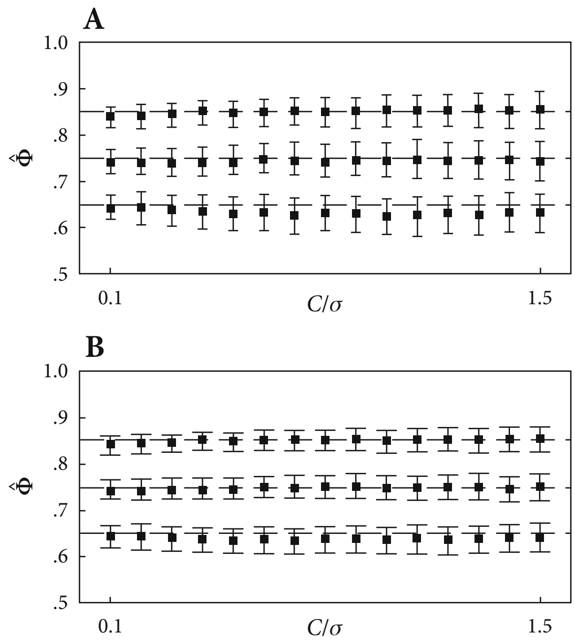

Figure 5. Estimated target probability $\hat{\Phi}$, expressed as median and interquartile range over 1,000 simulation runs (each starting at its convergence point) and plotted as a function of the relative step size C/σ for different expected target probabilities ($\Phi = .65$, $\Phi = .75$, $\Phi = .85$; dashed lines). Evaluations are shown for staircases lasting $N = 100$ trials (A) and $N = 200$ trials (B).

bias and variance of the target probability. For this reason, we eventually took the stimulus to be presented after the last tested value as the threshold intensity.

Comparison With Fixed-Step-Size Staircases

Figure 7 shows the estimated target probability for the three-down/one-up FSS staircases and the accelerated SA staircases aiming at the 79.37% and 83.15% correct responses, expressed as median and IQR over 1,000 simulation runs and plotted as a function of the relative step size (Δ^+/σ for the up–down staircase and C/σ for the SA staircase). As shown in Figure 7A, the convergence of the FSS staircase with equal step size is largely affected by the step size itself, with large fluctuations of both the mean and the variance of the estimated target probability; this last frequently turns out to be very far from the expected value. On the contrary, the accelerated SA procedure converged well to the expected target probability. Apart from deviations for low C/σ values resulting from the distance between the starting and the target points, this convergence occurs independently of the relative step size.

The accelerated SA method was also found to be more stable when the target probability was set at one of the values suggested in García-Pérez (1998) to reduce fluctuations of the outcome of FSS staircases. Indeed, excluding very low relative step size in which the FSS staircase performs better, Figure 7B shows that the SA procedure yielded lower bias and variance values than did the FSS rule when the 83.15%-correct point of the psychometric function was targeted.

DISCUSSION

In this study, the small-sample convergence properties of SA adaptive procedures were characterized by describing the empirical distribution of the target probability, which was estimated for different combinations of the parameters involved in the staircase definition and development. When the effects of the starting stimulus intensity on convergence were investigated, simulations run at varying values of $x(1)$, and according to the normal SA procedure of Equation 1, revealed that relatively short staircases often had no time to reach the target region if the step size was low and/or the starting point was far from the threshold intensity (Figure 3). This result strongly suggests the use of the accelerated SA version of Equation 2 that—thanks to the less frequent decreases of the step size, which is reduced only in correspondence of shifts in response category—has less inertia and may thus reach the target point much more rapidly than the normal procedure. Nevertheless, some actions should also be taken with regard to the accelerated SA so as to ensure cancellation of the occurrence of bias because of uncertainty in the placement of the initial stimulus. First, the experimenter should set an initial stimulus intensity above target: In 2AFC tasks, incorrect responses starting below threshold (that are necessary to increase the stimulus intensity toward the target) occur less frequently ($p < .5$) than correct responses starting above threshold ($p > .5$), thus making staircases starting below threshold less efficient than staircases starting above. Second, in order to guarantee that the threshold is actually targeted, a sufficiently high initial step size C has to be chosen, typically at least 0.4 times higher than the spread of the psychometric function for staircases of 100 trials. Note that these recommendations—though based on parameters that might be a priori unknown (i.e., the location and the spread of the psychometric function)—should be easily met, since they do not set strict values for the starting intensity and the initial step size. In any case, failing to meet these recommendations has only an impact on the efficiency but does not compromise the applicability of SA staircases. Indeed, we found that the threshold could be targeted even with low initial step size and/or stimulus intensity, at the cost of longer staircases. However, since the use of clear suprathreshold starting intensity and large initial step size allows the stimulus intensity to fall within the target region by saving a substantial number of trials, in practical implementations, one may find it useful to take into account these recommendations. Moreover, the use of large step size that makes the staircase more agile (i.e., with less inertia) and initial values well above threshold that help the observer to familiarize him- or herself with the task have been previously suggested in other studies implementing different psychophysical procedures (Green, 1990; King-Smith et al., 1994).

Provided that the recommendations above are met, the accelerated SA staircases seem to be suitable for guaranteeing convergence independently of the chosen relative step size. Indeed, the psychometric function slope was found not to affect the small-sample convergence of SA, since the estimated target probabilities remained un-

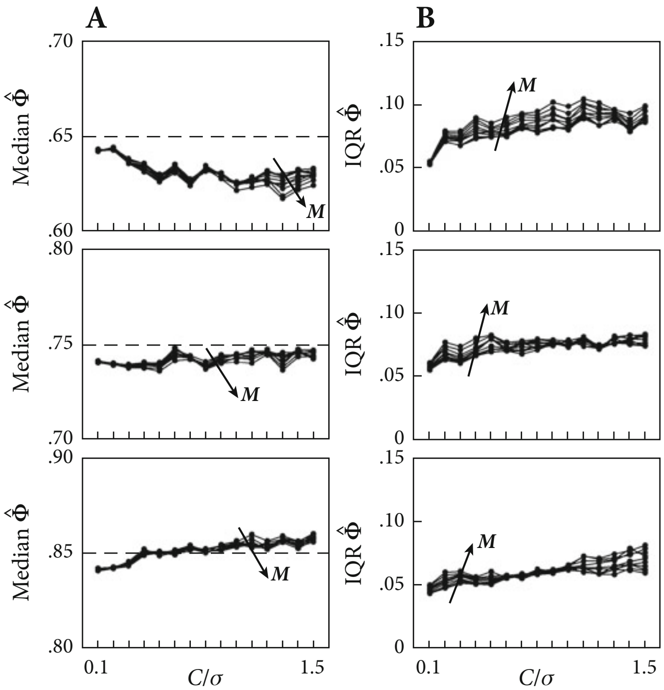

Figure 6. Median (A) and interquartile range (IQR; B) of the empirical distribution of the target probability $\hat{\Phi}$, estimated over 1,000 simulation runs (each starting at its convergence point and lasting $N = 100$ trials) and plotted as a function of the relative step size C/σ for different expected target probabilities: $\Phi = .65$ (up), $\Phi = .75$ (middle), and $\Phi = .85$ (down). The curves in each graph correspond to different values of the number of intensity values averaged at the end of the staircase ($M = 1, 5, 10, 20, 30, 40, 50, 60, 70, 80$), where shifts of the curves with increasing M are indicated by the arrows.

changed while varying the absolute value of the spread σ from 0.5 to 1 and then to 2 log units. Moreover, the target probability was substantially unaffected by the relative step size as long as this last parameter was sufficiently high to guarantee convergence (see Figures 4 and 5). This desirable property is not encountered in the traditional FSS staircases, as documented by García-Pérez (1998), who demonstrated that the target probability, estimated with classical k-down/one-up rules (with $k = 1, 2, 3$, and 4), can largely fluctuate upon variations of the relative step size. In addition, our simulations confirmed the expectations that the accuracy of small-sample convergence is closely related to the length of the staircase. As shown in Figures 4 and 5, increasing the length of accelerated SA staircases resulted in a reduction of both bias and variability of the estimated target probability, as well as a minimization of bias effects that were due to differences between starting and target stimulus intensities. This fact is documented by the data reported in the Results section: Longer staircases require a lower initial step size to allow for the convergence of the estimated target probability within 5% of its expected value. In practical implementations of these staircases, a trade-off has to be reached between the necessities of maximizing the accuracy of threshold estimation (requiring long staircases) and limiting the duration of the experiment (requiring short staircases).

With regard to the criterion for optimal threshold intensity determination, estimates of the target probability relying on the last tested value were more accurate than estimates obtained by averaging over several intensity levels at the end of the staircase. This result differentiates SA from FSS staircases for which it was shown that averaging over the reversal points as well as all the levels produced more stable threshold estimates (Klein, 2001; Levitt, 1971). The difference could arise again from the fact that in SA procedures, the step size is not constant; thus, it is better to consider only the last intensity (occurring after the smallest step of the staircase) in the threshold calculation, since the previous points result from larger intensity steps and can therefore bring instability.

As is documented by the good separation among the empirical distributions of the target probability estimated for different expected targets (Figure 5), a good setting of the adjustable parameters leads to high performance of the SA procedure in the estimation of different points of the psychometric function. Moreover, the SA method was found to be reliable for estimating low percent-correct points on the psychometric function, although the accuracy decreased with the probability targeted by the staircase. In our simulations, the negative bias of accelerated SA staircases of 100 trials targeting the 65% correct point was about 3%, and it decreased for longer staircases. Hence, the accuracy of

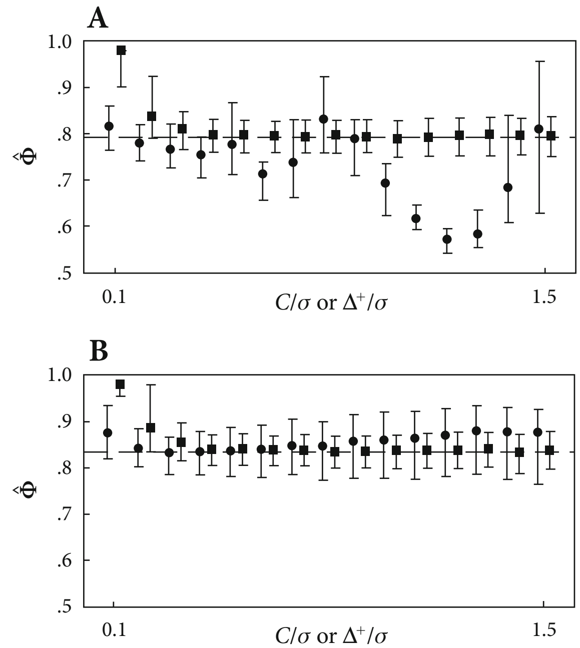

Figure 7. Estimated target probability $\hat{\Phi}$, expressed as median and interquartile range over 1,000 simulation runs (each starting with unitary stimulus intensity and lasting $N = 100$ trials) for the expected target probabilities $\Phi = .7937$ (A) and $\Phi = .8315$ (B). The two target probabilities were estimated by the traditional weighted up–down staircase (circles) and by the accelerated SA staircase (squares) and are plotted as a function of the relative step size (Δ^+/σ for the up–down staircase and C/σ for the SA staircase).

SA staircases offers the opportunity to safely target points far from the midpoint of the psychometric function, as is required in some psychophysical experiments. By contrast, the widely used FSS staircases have undesirable statistical properties, particularly when performance levels below the midpoint of the psychometric function have to be targeted (Green, 1990; Leek, 2001). For instance, FSS staircases were demonstrated to be strongly unfit—even asymptotically—for estimation of thresholds corresponding to lower target probabilities (García-Pérez, 1998).

The comparison with the very common FSS staircases using equal size for the upward and downward steps (Wetherill & Levitt, 1965) demonstrated the superiority of SA procedures in targeting the desired percent-correct point (Figure 7A). The better behavior can be explained in part by the use of different size for the upward and downward steps, since the SA procedure actually implements a weighted up–down rule (see Equations 1 and 2) that is quite similar to that proposed by Kaernbach (1991). The simulations by Kaernbach (1991) and García-Pérez (1998) revealed the good convergence properties of the weighted up–down method with respect to transformed up–down methods having equal upward and downward step size. In addition, SA staircases seem to work better even when compared with weighted up–down methods in which the FSS is in the optimal ratio for the adopted rule (as was suggested in García-Pérez, 1998; see, e.g., Figure 7B). In this case, the better performance has to be ascribed to the progressive reduction of the step size performed by Equations 1 and 2. Hence, reducing the size of the variations in stimulus intensity carried out during the experiment seems to favor the convergence of the staircase toward the desired target probability. Nevertheless, note that this aspect is somewhat perceived by researchers performing practical staircase thresholding: The practice of reducing the step size according to predefined heuristic rules is indeed very common in psychophysical experiments (see, e.g., Amitay, Hawkey, & Moore, 2005; Kaernbach, 2001a; Takeuchi, 2005). With consideration of these practical adjustments, FSS staircases would probably yield more encouraging results than those reported in García-Pérez (1998) and also in the present study (Figure 7). In addition, more complicated techniques that combine sequential statistics methods for deciding when to change stimulus intensity and sophisticated heuristic rules for setting the step size, such as PEST and subsequent refinements (Hall, 1981; Taylor & Creelman, 1967), actually perform step-size modifications during the development of the staircase. Given the similar conceptual framework, it would be interesting for one to compare SA staircases with methods performing non-Monro modifications of the step size, such as PEST or other heuristic approaches (e.g., varying the step size only after even or odd reversals, or halving it after each reversal). Qualitatively, the lower inertia of PEST and other rules less severe than SA in reducing the step size should probably result in a higher instability on the one hand, but a more efficient response against time variations of the psychometric function (e.g., induced by learning or attentional changes) on the other hand. In any case, systematic tests by means of computer simulations are needed to quantitatively compare the performance of the different staircase procedures based on progressive step-size modifications.

Since no parametric model of the psychometric function is required to implement the staircase, the SA procedure is included in the category of nonparametric methods for the adaptive estimation of sensory thresholds (Treutwein, 1995). For this reason, SA was compared in this study with other widely used nonparametric methods (Kaernbach, 1991; Levitt, 1971) rather than more sophisticated Bayesian and maximum likelihood procedures (King-Smith et al., 1994; Kontsevich & Tyler, 1999; Watson & Pelli, 1983). Although a direct comparison with parametric techniques goes beyond the scope of this study, we remark that nonparametric staircases are mostly used in practical psychophysical researches, thanks to their simplicity and relaxed assumptions (e.g., about shape and slope of the psychometric function); nonparametric staircases have been also recently reconsidered from a statistical point of view (Klein, 2001).

Our simulations were conducted by using the Weibull function for the definition of the underlying psychometric functions. However, as in the case of FSS staircases (García-Pérez, 1998), with the utilization of other shapes for the psychometric function (e.g., cumulative normal, or logistic), the results would presumably not change in any respect. Indeed, the only requirement about the psychometric functions used in nonparametric threshold estimation methods is monotonicity, a property satisfied by the Weibull function as well as by other functions commonly exploited in simulation experiments.

CONCLUSIONS

The small-sample simulation analyses conducted in this study led us to infer the expected performance of SA staircase methods in forced-choice adaptive threshold estimation and consequently establish an optimal experimental design for maximizing such performance in practical applications. Our results indicate that the initial stimulus intensity should be set well above threshold and that the accelerated version of the SA procedure should be implemented to facilitate the convergence toward the threshold region. With regard to the length of the staircase, in case of sequences of 50 trials, the bias that occurs because of the shortness of the staircase is substantially absent, provided that the initial step size *C* is larger than half the spread of the psychometric function. A length of 100 trials is suggested to reduce the IQR of the estimated target probability to about .07. The best criterion to infer the threshold estimate is to take the stimulus value calculated upon the last performed trial. Although some of these recommendations are made on the basis of parameters that are a priori unknown—such as the threshold intensity and the spread of the psychometric function—very rough estimates are sufficient to meet the indications proposed. Moreover, as SA staircases are asymptotically convergent, failing to satisfy these requirements will not result in missing the threshold, provided that a sufficient number of trials is performed.

With these settings and practical recommendations, SA staircases appear to make up a very dependable adaptive nonparametric technique for the accurate measurement of sensory thresholds in forced-choice experiments. In particular, our results are evidence that SA staircases should be used in place of traditional FSS staircases: In comparison with these last, SA staircases offer a greater reliability while being equally simple to implement and free of heavy assumptions about the psychometric function underlying the experiment. As opposed to FSS staircases, the SA method can be also used—though with an accuracy that decreases with the estimated target probability—to target performance levels below the midpoint of the psychometric function.

AUTHOR NOTE

Address correspondence to L. Faes, Lab. Biosegnali, Dipartimento di Fisica, Università di Trento, via Sommarive 14, Povo, Trento 38050 Italy (e-mail: faes@science.unitn.it).

REFERENCES

Amitay, S., Hawkey, D. J. C., & Moore, D. R. (2005). Auditory frequency discrimination learning is affected by stimulus variability. *Perception & Psychophysics*, **67**, 691-698.

Anderson, A. J., & Johnson, C. A. (2006). Comparison of the ASA, MOBS, and ZEST threshold methods. *Vision Research*, **46**, 2403-2411.

Brown, L. G. (1996). Additional rules for the transformed up–down method in psychophysics. *Perception & Psychophysics*, **58**, 959-962.

Fechner, G. T. (1860). *Elemente der Psychophysik* [Elements of psychophysics]. Leipzig: Breitkopf & Härtel.

García-Pérez, M. A. (1998). Forced-choice staircases with fixed step size: Asymptotic and small-sample properties. *Vision Research*, **38**, 1861-1881.

García-Pérez, M. A. (2002). Properties of some variants of adaptive staircases with fixed step sizes. *Spatial Vision*, **15**, 303-321.

Green, D. M. (1990). Stimulus selection in adaptive psychophysical procedures. *Journal of the Acoustical Society of America*, **87**, 2662-2674.

Hall, J. L. (1981). Hybrid adaptive procedure for estimation of psychometric functions. *Journal of the Acoustical Society of America*, **69**, 1763-1769.

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

Kaernbach, C. (2001a). Adaptive threshold estimation with unforced-choice tasks. *Perception & Psychophysics*, **63**, 1377-1388.

Kaernbach, C. (2001b). Slope bias of psychometric functions derived from adaptive data. *Perception & Psychophysics*, **63**, 1389-1398.

Kershaw, C. D. (1985). Statistical properties of staircase estimates from two interval forced choice experiments. *British Journal of Mathematical & Statistical Psychology*, **38**, 35-43.

Kesten, H. (1958). Accelerated stochastic approximation. *Annals of Mathematical Statistics*, **29**, 41-59.

King-Smith, P. E., Grigsby, S. S., Vingrys, A. J., Benes, S. C., & Supowit, A. (1994). Efficient and unbiased modifications of the QUEST threshold method: Theory, simulations, experimental evaluation and practical implementation. *Vision Research*, **34**, 885-912.

Klein, S. A. (2001). Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & Psychophysics*, **63**, 1421-1455.

Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, **39**, 2729-2737.

Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception & Psychophysics*, **63**, 1279-1292.

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

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.

Robbins, H., & Monro, S. (1951). A stochastic approximation method. *Annals of Mathematical Statistics*, **29**, 400-407.

Takeuchi, T. (2005). The effect of eccentricity and the adapting level on the café wall illusion. *Perception & Psychophysics*, **67**, 1113-1127.

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

Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Research*, **35**, 2503-2522.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, **33**, 113-120.

Wetherill, G. B., & Levitt, H. (1965). Sequential estimation of points on a psychometric function. *British Journal of Mathematical & Statistical Psychology*, **18**, 1-10.

Zwislocki, J. J., & Relkin, E. M. (2001). On a psychophysical transformed-rule up and down method converging on a 75% level of correct responses. *Proceedings of the National Academy of Sciences*, **98**, 4811-4814.

(Manuscript received February 2, 2006;
accepted for publication May 8, 2006.)