A single-interval adjustment-matrix (SIAM) procedure for unbiased adaptive testing

Christian Kaernbach
Laboratoire d'Audiologie Expérimentale, INSERM unité 229, Université de Bordeaux II, Hôpital Pellegrin, Place Amélie Raba Léon, 33 076 Bordeaux Cedex, France

(Received 10 December 1989; accepted for publication 16 July 1990)

A new unbiased adaptive procedure is described that requires only half as many presentations in achieving the same precision as the well-known two-interval forced-choice (2IFC) 2-step procedure. The procedure is based on a yes–no task which avoids redundant presentation time. Furthermore, certain psychophysical studies can only be realized with yes–no tasks. Every trial contains randomly presented signals or noises and the answer is either yes or no. The outcome (hit, miss, false alarm, correct rejection) is taken into account by adjusting the signal level in a staircase manner. The adjustment matrix is set up to induce a neutral response criterion. Its convergence point can be adjusted at will. The single-interval adjustment-matrix (SIAM) procedure is compared to von Békésy and 2IFC transformed up–down methods using a Monte-Carlo simulation. The SIAM procedure proves to be the fastest of the unbiased procedures. A test on four subjects verified these results. Implications for optimum track length and the number of reversals to discard are discussed.

PACS numbers: 43.66.Yw[WAY]

INTRODUCTION

In psychophysics it is a very common procedure to determine the perception threshold for a signal, that is, to evaluate the level where the signal is just perceived. As detection is not a deterministic but a probabilistic process, evaluation requires a lot of redundant trials at appropriate signal levels to give a good estimate of the threshold. The appropriate signal levels are not known at the beginning of the test and so adaptive procedures are used to adjust the signal level. In such procedures the task is made more difficult when the subject performs well (as defined by the experimenter) and easier when the subject performs badly. The signal level will thus oscillate around its target value.

This paper was inspired by the desire to construct an unbiased adaptive procedure based on yes–no tasks rather than on forced-choice tasks. These two different sorts of task lead to quite different possibilities and results. A series of forced-choice trials allows for estimation of the area under the receiver operating characteristic (ROC, cf. Green and Swets, 1974) as a measure of sensitivity. On the other hand, a series of yes–no trials will give an estimate of the hit rate and the false alarm rate for a certain response criterion. For a neutral response criterion, the reduced hit rate (hit rate minus false alarm rate) is also a valid measure of the sensitivity (for a comparison of this measure to the area under the ROC see Sec. II A). Furthermore, both series will reveal an additional information. Yes–no tasks allow for differentiation of the subjects treatment of signal and noise, whereas the additional information of forced-choice tasks is spoiled by the symmetric design of these tasks: the difference in the treatment of the first versus the second interval is less interesting. Yes–no tasks will give different and perhaps more information than forced-choice tasks, and in addition will give comparable detectability information in less time (same trial number but less presentations). Given that the response criterion is controlled, yes–no tasks are highly recommended for the construction of efficient adaptive procedures.

Efficiency is a major criterion in choosing the adaptive procedure: one wants to get the maximum precision for the experimentation time invested. The relative importance of precision versus time, as well as the definition of precision, may vary according to particular situations. For example, one investigator may be interested in a specific target point of the psychometric function, whereas another may be interested in its spread. Staircase techniques have proven to be flexible enough to accomodate a variety of needs. In addition, they are easy to control and provide fast and stable data analysis. Section I discusses known adaptive staircase techniques. Section II introduces the single-interval adjustment-matrix (SIAM) procedure. Section III compares well-known adaptive procedures to the SIAM procedure using a Monte-Carlo simulation. Section IV presents an experimental test on human subjects to compare these procedures.

I. ADAPTIVE STAIRCASE TECHNIQUES

With staircase techniques, the signal level is not changed continously but in discrete steps. The restricted set of possible signals may have advantages in certain experimental situations. The control of such an experiment is made easier by the fact that one need not transfer a continous signal level to the sound producing apparatus but only a small integer number indicating the stair of the next trial. A series of trials forms a track. The trials leading to a change of direction in the variations in signal level are called reversals. In most applications, the track is continued until a certain prescribed number of reversals is reached. Data analysis is commonly accomplished by averaging the reversal points or by taking the median of them. To avoid a bias in the estimates, the number of reversal points requested should be even. This

 0001-4966/90/122645-11$00.80 © 1990 Acoustical Society of America

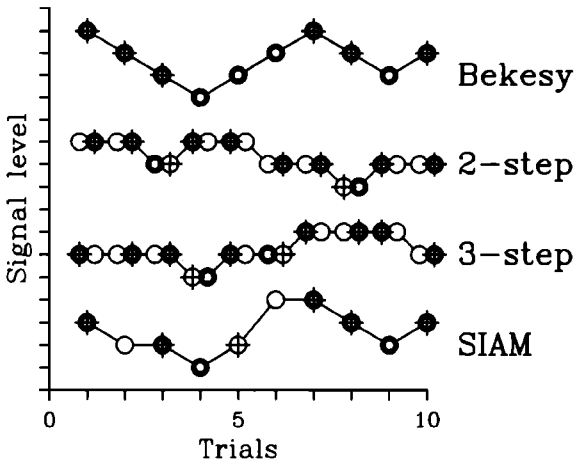

FIG. 1. Portions of possible tracks for von Békésy tracking, 2-step rule, 3-step rule, and the SIAM procedure. Heavy circles indicate signal presentation, whereas light circles represent noise. A cross indicates that the subject heard the signal in that presentation. Note the "silent change" following a false alarm at trial 5 of the SIAM track.

type of data analysis is very simple and has proven to be robust, efficient, and precise.

The classical staircase technique is the "1 up 1 down" rule of von Békésy: In each trial the signal is presented, and with positive responses the signal level is reduced and with negative responses it is increased. A portion of a possible track is given in Fig. 1. This procedure should converge to the 50% performance point of the psychometric function. But the subject will anticipate the trial, and this will allow him to use lower response criteria and will lead to significantly lower signal levels. As the response criterion of the subject is not under the control of the experimenter and may generally change markedly, it is capable of introducing noticeable errors. The method of von Békésy is thus a fast but not objective adaptive procedure, because it strongly depends on the subject's response criterion.

A forced-choice task will overcome this problem, due to the fact that the subject does not know in which interval the signal will be presented. Two or three intervals are commonly presented on each trial. The two-interval forced-choice (2IFC)[1] task will lead to a chance performance of 50%, whereas the 3IFC task will lead to 33.3%. This creates problems for the simple up–down method of von Békésy, as the latter will converge to the 50% performance point. For the 2IFC task a 75% performance point would be acceptable for most applications, as the performance in a 2IFC task varies between 50% and 100%. For 3IFC tasks, the halfway performance is 66.7%.

To use forced-choice tasks in adaptive procedures, several modifications to von Békésy's rule were developed. Levitt (1971) could show that a large part of them belong to one class: the "transformed up–down methods." Among these, the 2-step (1 up 2 down) and the 3-step (1 up 3 down) rule are the best known and commonly used procedures. Following these rules the signal level will be increased with each incorrect response and decreased after two or three successive correct responses, respectively. Portions of possible tracks are given in Fig. 1. These procedures lead to 70.7% or 79.4% performance points. In combination with the forced-choice tasks this leads to unbiased procedures that converge in the neighborhood of the halfway performance points of 67% or 75%. Another approach to adjust the target point of the adaptive procedure was introduced by Taylor and Creelman (1967; Taylor *et al.*, 1983). The parameter estimation by sequential testing (PEST) uses variable step sizes that within one track lead to decreasing steps as the threshold is approached. The target point of this procedure can be adjusted to any desired value. Different rules to adjust the signal level are given by maximum likelihood techniques (Hall, 1968). These rules can be combined with 2IFC tasks as well as with 3IFC tasks. The 2IFC 2-step procedure has become a standard procedure in psychophysics. It is a criterion-free procedure, easy to control, and simple and precise for data analysis. On the other hand, it is costly in experimentation time as it needs three to four times more presentations than the simple up–down method of von Békésy.

Many studies have focused on comparison of the efficiency of adaptive staircase rules and possible improvements. Several approaches were used to determine the relative efficiency of the various adaptive methods: computer simulations (Emerson, 1984; Findlay, 1978; Hall, 1968, 1974, 1981; Pentland, 1980; Schlauch and Rose, 1986; Taylor and Creelman, 1967), a theoretical Markov-chain model (Kollmeier and Gilkey, 1983), as well as empirical data of human subjects (Shelton *et al.*, 1982; Shelton and Scarrow, 1984). Kollmeier *et al.* (1988) compared the prediction of a Markov-chain model with empirical data. They considered different combinations of tasks (2IFC/3IFC) and rules (2-step/3-step/PEST) and found the most commonly used combination "2IFC 2-step" to be slightly less efficient than the other combinations. De Boer and van Breugel (1984) enlarged the range of transformed up–down methods, introducing a dependence on the actual direction. Upward and downward runs are considered separately and the reversal rule (when to terminate a run by a reversal) is conditional upon the direction. They could develop a rule that would avoid reversals in the center region, where random walk characteristics dominate. The reversals obtained in this way would then give more reliable information about the psychometric function. Green *et al.* (1989) studied the effect of initial and final step size as well as stimulus heterogeneity (assuming a set of stimuli producing different psychometric functions) on the variability of the threshold estimates. They recommended an initial step size of 1/4 and a final step size of 1/8 of the useful range (60%–90%) of the psychometric function. The stimulus heterogeneity seems to have only a small effect on the variability.

The studies cited above concentrated on the rules applied in adaptive procedures, but little attention has been given to the task. Apart from von Békésy's classical procedure, yes–no tasks are generally not used in adaptive psychophysics, and are rarely considered for the construction of unbiased procedures. Adaptive testing concentrated on forced-choice tasks. However, experimental conditions can sometimes recommend yes–no tasks more than forced-choice tasks. For example, Moore *et al.* (1986) found yes–no

tasks more appropriate than forced-choice tasks for the measuring of the thresholds for hearing mistuned partials as separate tones. They developed an adaptive yes–no task procedure that fullfilled their requirements but was not extremely efficient. Furthermore experiments with long-lasting stimulus intervals (e.g., concerning the perception of rhythmic structures) will prevent the subjects from comparing the intervals directly. Besides the conceptual advantages that yes–no tasks may have for certain experimental situations, they improve efficiency by avoiding redundant presentation time. The following section introduces an unbiased adaptive staircase technique based on a yes–no task that leads to a completely new method of adjusting signal level.

II. A SINGLE-INTERVAL ADJUSTMENT-MATRIX (SIAM) PROCEDURE

The multiple-interval tasks can not be optimally efficient as they give too much information to the subject, thus deliberately reducing the information gained from the answer. The 2IFC task obtains one bit of information with two presentations. Two bits could be obtained, if the subject did not know that exactly one signal was present in the two presentations. Yes–no tasks are optimal from that point of view: They do not offer additional information. This recommends them for the construction of a very efficient adaptive procedure. The yes–no task should, however, contain noise presentations to control the subject's response strategy. The use of yes–no tasks for adaptive procedures as well as the interpretation of the results requires careful examination of the underlying signal detection theory (Green and Swets, 1974; Kaernbach, 1990). Sections A and B will prepare the theoretical background for the construction of an unbiased adaptive procedure based on yes–no tasks. Section C will introduce the single-interval adjustment-matrix (SIAM) procedure and will discuss its construction.

A. Measuring the sensitivity

In signal detection theory it is assumed that the subject evaluates a likelihood ratio $\mathbf{l} = \mathbf{P}(\mathbf{e}|\mathbf{s})/\mathbf{P}(\mathbf{e}|\mathbf{n})$ of the probabilities for the observed event **e** given the hypothesis **s** "signal was present" or given the hypothesis **n** "noise alone." In a yes–no task the subject fixes a criterion β and decides to answer yes if $\mathbf{l} > \beta$. In a 2IFC task the subject will choose the interval with the largest likelihood ratio, i.e., he will answer "first interval" if $\mathbf{l}_1 > \mathbf{l}_2$. This implies a symmetry in the observer's decision to the effect that there is no bias in the selection of one interval over the other. More generally, the decision rule for a 2IFC task is: reply "first interval" if $\mathbf{l}_1 > \mathbf{c} \cdot \mathbf{l}_2$, where **c** is the response bias parameter (cf. Green and Swets, 1974).

The outcome of a series of trials can be described by two values for both tasks. For a yes–no task this will be the probabilities $\mathbf{P}(\text{yes}|\mathbf{n})$ and $\mathbf{P}(\text{yes}|\mathbf{s})$ to answer "yes" when presented "noise" or "signal." For a 2IFC task this will be the probabilities $\mathbf{P}(\mathbf{R1}|\langle\mathbf{ns}\rangle)$ and $\mathbf{P}(\mathbf{R1}|\langle\mathbf{sn}\rangle)$ to answer "first interval" (**R1**) when presented "noise first, then signal" ($\langle\mathbf{ns}\rangle$) or when presented "signal first, then noise" ($\langle\mathbf{sn}\rangle$). The probabilities to answer "no" or to answer "second interval" (**R2**) follow from the fact that the sums for all possible

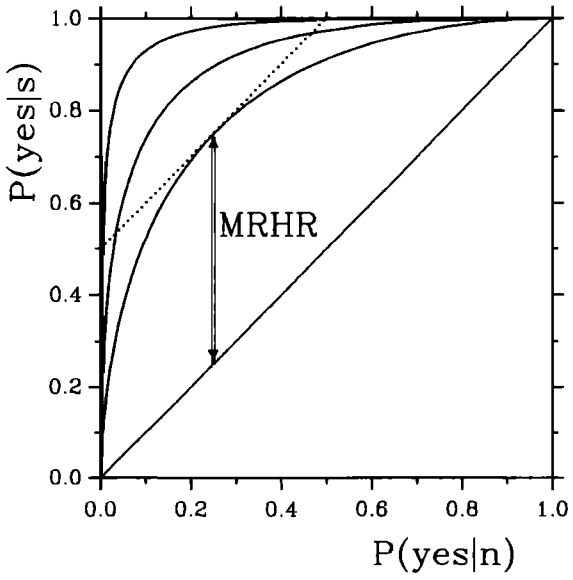

FIG. 2. The yes–no task outcome plotted in the probability unit square. The solid lines represent ROC curves for several signal levels. The lowest ROC curve has a maximum RHR of 0.5. It touches a line of equal payoff for a neutral payoff matrix (dotted line).

answers must be one. This makes it possible to describe the outcome of these tasks as a point in a probability unit square. Varying the criterion β or the response bias parameter **c** will move this point through the square. The resulting plot is called receiver operating characteristic (ROC). The solid lines in Fig. 2 give an example for the yes–no task at different signal levels. For weak signals the ROC is close to the diagonal, whereas for stronger signals it approaches the upper left corner.

The ROC describes all the subject's possible behaviors toward a stimulus of a certain strength. But for many applications it would be sufficient to obtain a single number that describes the sensitivity for this stimulus. For this purpose the maximum distance of the ROC from the diagonal can be evaluated. At the ROC position with the maximum distance from the diagonal the slope of the ROC is 1 (or, for discrete models, it may jump from a value greater than 1 to a value less than 1). At this point β or c equals 1 (Green and Swets, 1974). Section II B will treat the possibilities to induce the peak position of the ROC. Then the distance from the diagonal $\mathbf{P}(\text{yes}|\mathbf{s}) - \mathbf{P}(\text{yes}|\mathbf{n})$ or $\mathbf{P}(\mathbf{R1}|\langle\mathbf{sn}\rangle) - \mathbf{P}(\mathbf{R1}|\langle\mathbf{ns}\rangle)$ is a single-number measure of the subjects sensitivity. For yes–no tasks the difference $\mathbf{P}(\text{yes}|\mathbf{s}) - \mathbf{P}(\text{yes}|\mathbf{n})$ of the hit rate $\mathbf{P}(\text{yes}|\mathbf{s})$ and the false alarm rate $\mathbf{P}(\text{yes}|\mathbf{n})$ is commonly called reduced hit rate (RHR). At the "$\beta = 1$" ROC position the RHR is maximal. The maximum RHR (MRHR) is illustrated for the lowest ROC curve of Fig. 2. For 2IFC tasks the difference $\mathbf{D} = \mathbf{P}(\mathbf{R1}|\langle\mathbf{sn}\rangle) - \mathbf{P}(\mathbf{R1}|\langle\mathbf{ns}\rangle)$ is commonly transformed to the average percentage correct:

$$(\mathbf{D}+1)/2 = \{\mathbf{P}(\mathbf{R1}|\langle\mathbf{sn}\rangle) + [1 - \mathbf{P}(\mathbf{R1}|\langle\mathbf{ns}\rangle)]\}/2$$
$$= [\mathbf{P}(\mathbf{R1}|\langle\mathbf{sn}\rangle) + \mathbf{P}(\mathbf{R2}|\langle\mathbf{ns}\rangle)]/2. \quad (1)$$

For the $\mathbf{c} = 1$ position of the ROC, this is equal to the area under the ROC (AUROC) for the corresponding yes–no task (Green and Swets, 1974).

Both AUROC and MRHR are valid measures of the subjects sensitivity. Their relation is determined by the sig-

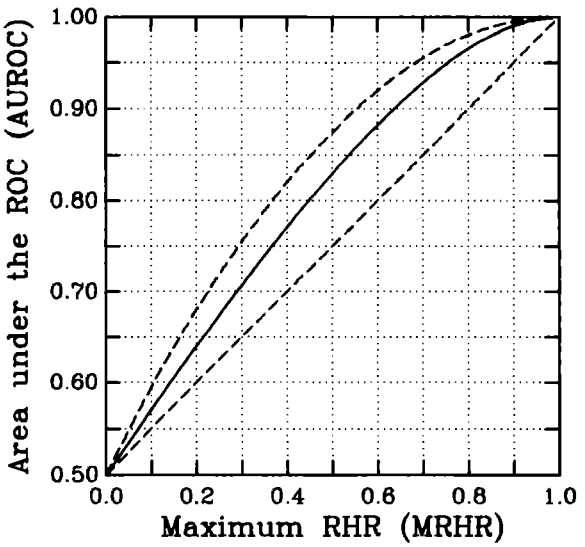

FIG. 3. Relation between the area under the ROC (AUROC) and the maximum reduced-hit rate (MRHR). The dashed lines give the range of possible values for any possible signal detection model. The solid line corresponds to the Gaussian model.

nal detection model that is assumed to produce the ROC. Figure 3 gives the AUROC as a function of the MRHR. The dashed lines give the maximum and the minimum AUROC for any possible signal detection model. The lower border corresponds to the low-threshold model (Luce, 1963a,b). The upper border is given by the total area under the corresponding parallel of the diagonal. These borders follow power functions with the exponent being equal to 1 for the lower border and 2 for the upper border. The solid line gives the relation for the Gaussian observer. It is well approximated by a power function with the exponent 1.55.

B. Controlling the bias and the criterion

The measures of sensitivity discussed above require that the response bias and the response criterion be directed toward the ROC position with slope one to maximize the distance to the diagonal. For forced-choice tasks this is done simply by instruction. The subject is told not to prefer one interval over the other. Furthermore the subject knows that the preference of one interval will lower his performance: The threshold estimates will indicate a lower sensitivity. As the subject is motivated to show a high sensitivity, this too will direct the subject to adopt a neutral attitude ($c = 1$).

For yes–no tasks the criterion β can be influenced with a so-called payoff matrix. The four possible outcomes of a yes–no trial are rewarded or punished with small monetary amounts. If a miss is punished with a relatively high amount, the subject will lower his criterion β, i.e., he will more easily answer yes. The average payoff of the subject depends linearly on his hit rate $P(\text{yes}|\mathbf{s})$ (p_s) and his false alarm rate $P(\text{yes}|\mathbf{n})$ (p_n): All points (p_n,p_s) that give a certain average payoff form a straight line with a certain slope. This is illustrated by the dotted line in Fig. 2. The ROC point whose slope is equal to this slope will receive maximum average payoff (see, e.g., the touching point of the lowest ROC curve of Fig. 2). To induce a neutral response criterion the slope of the lines giving equal payoff should be equal to 1. For a signal quota of 50%, this is the case if the average reward for the answer yes is equal to the average reward for the answer no.

In adaptive psychophysics the payoff matrix can be substituted by an adjustment matrix. As the subject is motivated to show a high sensitivity (a low threshold estimate), each increase of the signal level is equivalent to a punishment, and each lowering is a reward. As with payoff matrices, a feedback about the resulting changes will help the subjects to maintain the optimal criterion. Imagine, e.g., an adaptive procedure in which the outcome of each trial is taken into account by adjusting the signal level (adjustment matrix), and by simultaneously paying a proportional monetary amount (payoff matrix). For example, lowering the signal level for 1 dB could be attended by receiving \$1, and increasing it for 1 dB could be attended by being eased of \$1. The overall lowering of the level at the end of this track (i.e., the starting value minus the final value) would then correspond exactly to the overall payoff. It would hence be sufficient to regard only the final value. That is, if the subject is motivated to show a maximum sensitivity (e.g., by an increasing monetary reward for decreasing threshold estimates), the effect of the adjustment matrix is equivalent to the effect of a proportional payoff matrix. Fortunately, for psychophysical research, monetary rewards are not the only possibility to motivate the subjects to show high sensitivity measures.

Imagine the dotted line of Fig. 2 to represent the line of an average adjustment of zero. The signal level corresponding to the lowest ROC curve would allow the subject to keep this level constant—at least in the average. Lower signal levels lead to an increase, whereas higher levels allow the subject to achieve a lowering in the average. If the subject ever fails to maintain a neutral response criterion, this will only increase the threshold estimates, never decrease as with the von Békésy method.

C. Task and rules

The single-interval adjustment-matrix (SIAM) procedure is based on a yes–no task. The latter consists of a single presentation of either a signal or a noise presented randomly. The subject is asked whether there was a signal or not. As in signal detection theory four events are possible: hit, miss, false alarm, or correct rejection. Each outcome is taken into account by decreasing or increasing the signal level according to the adjustment matrix. This matrix is set up so as to lead to the desired target performance.

The experimental situation is defined by the adjustment matrix and the probability of signal presentation (signal quota). The average adjustment of the signal level over a series of trials follows then from the subject's average behavior, expressed as false alarm rate p_n and hit rate p_s. For the behavior at the target performance, the average adjustment should be zero. The points (p_n,p_s) that will give a zero adjustment form a straight line in the probability unit square, as the adjustment is a linear function of the response probabilities p_n and p_s. This equilibrium line (EQL) has to be parallel to the diagonal so as to induce a neutral response criterion (see previous section):

$$p_s = p_n + t, \tag{2}$$

where t is the desired target performance. The latter can take any value between 0 and 1. It describes the maximal difference between hit rate and false alarm rate, the maximum reduced-hit rate (MRHR). A target performance of $t = 0.5$ seems to give a good estimate of the threshold. The following paragraph derives the adjustment matrix for any value of t.

Let us denote the adjustment for the four possible outcomes of each trial (hit, miss, false alarm, correct rejection) by M_h, M_m, M_{fa}, and M_{cr}, respectively. They describe for each individual outcome the corresponding amount of change in the signal level. Together with the signal quota s one has five parameters influencing the average adjustment τ. The latter is a linear function of p_n and p_s:

$$\tau = s[M_h\ p_s + M_m(1 - p_s)] + (1 - s)[M_{fa} \cdot p_n + M_{cr}(1 - p_n)]. \quad (3)$$

One needs to know a feasible constellation of the parameters that give an EQL ($\tau = 0$) with the desired target performance. Let us first introduce reasonable restrictions to the set of the five parameters: M_{cr} can be set to zero, as this is a plausible reaction to an outcome that is not a mistake, but is also not an achievement.[2] Let us set M_h to -1. This may be changed later by multiplying the entire matrix appropriately to avoid fractional steps. Setting $s = 0.5$ (equal probability for noise and signal) will reduce the set of free parameters to two, controlling the two parameters of the EQL:

$$0 = [-p_s + M_m \cdot (1 - p_s) + M_{fa} \cdot p_n]/2$$
$$= [M_m - (M_m + 1) \cdot p_s + M_{fa} \cdot p_n]/2. \quad (4)$$

The slope of the EQL should be 1 [Eq. (2)], this leads to

$$0 = M_m - (M_m + 1) \cdot (p_n + t) + M_{fa} \cdot p_n$$
$$= p_n \cdot (M_{fa} - M_m - 1) + M_m \cdot (1 - t) - t. \quad (5)$$

This is valid for any value of p_n. Setting both the absolute and linear terms to zero leads to $M_m = t/(1 - t)$ and $M_{fa} = M_m + 1 = 1/(1 - t)$.

We can now give the matrix for any desired target performance t:

	yes	no
Signal: 50%	-1	$t/(1-t)$
Noise: 50%	$1/(1-t)$	0,

(6)

or an appropriate multiple of this. The entries indicate the amount of change in the signal level. A positive sign corresponds to an increase in the signal level. Table I gives convenient matrices for several commonly used values of the target performance t.

TABLE I. Payoff matrices for different values of the desired target performance t. The entries indicate the number of steps to change the signal level for each event. An event consists of a combination stimulus (signal/noise) and response (yes/no). A positive sign stands for increasing the level. The signal quota is 50% for all matrices. The best threshold estimate is given by 50% target performance ($t = 0.5$).

	$t = 0.25$		$t = 0.33$		$t = 0.5$		$t = 0.67$		$t = 0.75$	
	yes	no	yes	no	yes	no	yes	no	yes	no
Signal: 50%	-3	1	-2	1	-1	1	-1	2	-1	3
Noise: 50%	4	0	3	0	2	0	3	0	4	0

Let us consider the most interesting case of $t = 0.5$. A portion of a possible track is given in Fig. 1. The upper row of the matrix corresponds to the von Békésy rule. The lower row implements the adjustments of level that are necessary to control the response criterion. As with the known up–down methods, midrun estimates may be used to analyze the data. It seems surprising that the signal level is changed even after noise trials, and a succession of noise–yes combinations (false alarms) increases the signal level without applying it until the next presentation of a signal. But these "silent changes" contribute to the convergence toward the desired target performance. This can be seen from the comparison of the SIAM procedure with the von Békésy method by means of the Monte-Carlo simulation (see below): The estimates obtained from the SIAM procedure are much less effected by variations in the form of the ROC than those obtained from the von Békésy method.

The restriction to a signal quota of 50% is arbitrary. Other quotas lead to other matrices for the same target performance. One could think of increasing the signal quota in order to avoid the nonmoving event "correct rejection." But soon unreasonably high adjustments for false alarms would be prescribed to achieve the same target performance. False alarms will become rare events, but when they happen they will disrupt the perceptual context of the adaptive task. This approach would ignore the contextual nature of detection tasks, while leading to mathematically correct results. It seems reasonable to keep the matrix in a range where the biggest step is not more than four times the smallest step. A signal quota of 50% leads to a procedure that is convenient to use and easy to understand with matrix entries of not more than four for target performances in the range of 0.25–0.75 (see Table I). A deviation from this quota is only appropriate at the beginning of a track where the region of interest should be reached more quickly: A signal quota of 75% is suitable before the first reversal.

III. MONTE-CARLO SIMULATION

The adaptive procedures presented in Sec. I represent a set of commonly used staircase techniques. To compare their efficiency to that of the method proposed here, a Monte-Carlo simulation was carried out. The 2IFC 2- and 3-step procedures as well as the von Békésy method were compared to the SIAM procedure.

A. The model and its implementation

A simulation model for adaptive psychophysics has to specify the psychometric function. The latter describes the performance of the subject as a function of the signal intensity. Its shape depends on the intensity scale used. For logarithmic intensity scales, the psychometric function is usually shaped like an arc tan, or a tanh function. It levels off at 1.0 for high intensities and at the chance performance θ for low intensities.

For forced-choice tasks, the performance of the subjects can easily be described by a single number: the probability p

to select the correct interval. The chance performance θ is one over the number of intervals presented (0.5 for 2IFC). The performance in yes–no tasks is usually described by two numbers: the hit rate p_n and the false alarm rate p_s. For very low intensities, both are equal to θ, as the subject has no possibility of distinguishing between signal trials and control trials. For high intensities, where the subject will perform perfectly, p_s will reach 1, whereas p_n will decrease to 0. In this simulation model the false alarm rate p_n was assumed to be related linearly with the miss rate: $p_n = (1 - p_s) \cdot \theta / (1 - \theta)$; the point (p_n, p_s) will move on a straight line from the chance performance point (θ,θ) at the diagonal to the upper left corner of perfect performance (0,1). Here, θ describes the asymmetry of the underlying receiver operating characteristics (ROC). The peak of a symmetric ROC lies always on the counterdiagonal ($\theta = 0.5$). Values smaller than 0.5 correspond to the commonly observed asymmetry of the ROC, which is directed toward the lower left-hand half of the probability unit square (see, e.g., Kaernbach, 1990).

The psychometric function was assumed to be a tanh function. The intensity scale was labeled in dB and the steps of the staircase procedures were 1 dB. The tanh function was chosen so that the tangent at the inflection point at 0 dB intersected the chance performance at -4 dB and the perfect performance at 4 dB. The total width of 8 dB corresponds to the experimental situation described below for subject CK. This psychometric function describes the percentage correct p in a forced-choice task or the hit rate p_s in a yes–no task.

The signal level was set to $+10$ dB at the beginning of each track. After two reversals the presentation counter was reset to zero and the reversals were discarded. After 24 further reversals the track was stopped. At every even number of reversals, the median intensity was calculated and the number of presentations up to that point was stored. Ten thousand tracks were simulated for each condition. For the yes–no tasks, θ took values between 0 and 0.3. For the SIAM procedure the matrix for $t = 0.5$ was applied (see Table I). The simulation was done on an IBM PC AT (12 MHz). Computation for each track lasted approximately 0.5 s.

B. Results and discussion

Figures 4 and 5 give the results of the simulation. The single data points correspond to stop criteria of 2,4,6,...,24 reversals. The statistical and systematic errors are shown as a function of the presentation number. The presentation number is twice the trial number for the 2IFC tasks. The statistical error (Fig. 4) is determined as the standard deviation of the median values. The systematic error (Fig. 5) is determined as the difference between the mean of the median values and the theoretical target values of each procedure. Squares show the values for the von Békésy method, circles for the SIAM procedure, upward triangles for the 2IFC 2-step and downward triangles for the 2IFC 3-step procedure (up to 18 reversals only). The data points for the yes–no task procedures were obtained for $\theta = 0$. The hatching shows the range that the yes–no procedures cover for θ up to 0.3.

The statistical error of the von Békésy method increases markedly with increasing θ, whereas the SIAM procedure suffers much less from the increase. The 2IFC 2-step procedure has slightly more statistical errors than the 3-step procedure, whereas it has slightly less systematic errors. The difference seems negligible in any case. Both yes–no task procedures are markedly faster than the 2IFC procedures with approximately the same statistical errors. The systematic error of the SIAM procedure shows nearly no dependence on θ. Small systematic errors are achieved markedly faster than with the 2IFC procedures. The systematic error of the von Békésy procedure (slanting hatching) depends largely on θ. The systematic errors of adaptive procedures are normally overestimations of the signal level at the threshold. This is due to the higher level of the starting point. With von Békésy, there exists the possibility of negative estimation bias: increasing values of θ will lower the measured threshold beneath the real threshold. It is these large systematic

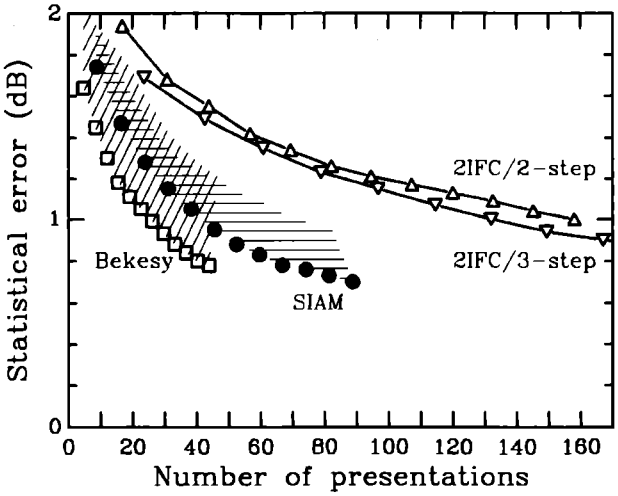

FIG. 4. Simulation: Statistical error as a function of the presentation number. Every data point represents a certain stop criterion (stop after 2,4,6,...,24 reversals). The hatched areas correspond to increasing θ from 0–0.3.

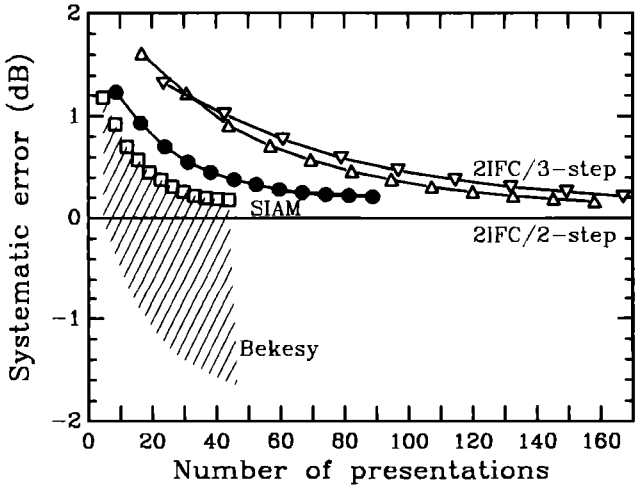

FIG. 5. Simulation: Systematic error as a function of the presentation number. The systematic error of the von Békésy method is indicated by the squares ($\theta = 0$) and the slanting hatching ($0 < \theta \leqslant 0.3$). It depends markedly on θ.

errors of the von Békésy method that rule out its use in objective psychophysics, as they are not under the control of the experimenter. They reduce the evidence of results obtained with this method to subjective measurements, strongly depending on the subjects response criterion, which generally may change markedly.[3]

The SIAM procedure proves to be the fastest of the unbiased procedures. For the 2IFC procedures, every two presentations form a trial. If the response time takes much longer than the presentation time, the number of trials is more important than the number of presentations. This would give comparable results for the 2IFC procedures and the SIAM procedure. In reality however, the average time for a 2IFC trial is certainly longer than that for a yes–no trial. In the experimental situation described below, the ratio of the 2IFC time to the yes–no time was 1.52. In addition, this experiment was an efficiency test with intentionally short presentations, whereas presentations used in ordinary research often last several seconds, yielding a ratio even closer to 2. This justifies the comparison of the number of presentations instead of the number of trials and makes the advantage of the SIAM procedure evident.

C. Optimum track length and discard

The statistical error can be reduced for all adaptive methods by carrying out more tracks. More short tracks instead of a few long tracks might lead to comparable statistical errors. However, this procedure would not be particularly efficient. The initial phase of the track, inclusive of discarded reversals, is repeated more often than necessary. Moreover, the systematic error is increased markedly by this procedure, as it will not level out with increasing track number. It seems to be reasonable to keep the track length at at least ten reversals. If the statistical error is decreased by carrying out more tracks, an appropriate lengthening of the tracks should be considered to decrease the systematic error.

Extensive tests on human subjects (Kollmeier *et al.*, 1988) show that the statistical error will not decrease as rapidly as predicted by the mathematical models, and that it will not approximate zero for increasing track lengths. This is presumedly due to the fact that the single trials may not be regarded as independent. Instead, Kollmeier *et al.* assumed that the threshold undergoes slow variations, staying nearly constant from trial to trial but changing from track to track. Therefore, they suggested using more shorter tracks instead of a few long ones. For n tracks this should lead to an additional decay of $1/\sqrt{n}$ of the statistical error. This is true only if the tracks are not undertaken one after the other, otherwise the results of the tracks themselves will not be independent. The same effect should also be seen for long tracks: interleaving of different tracks with a better distribution of the trials over time should make the results of the trials more independent. The distribution of the trials should be optimized in time, using either more shorter tracks or interleaving several longer tracks. The problem with short tracks is that the systematic error will not level out.

In the experimental test as well as in the simulation, two reversals were discarded per track as is generally employed for adaptive staircase procedures. A further test investigated

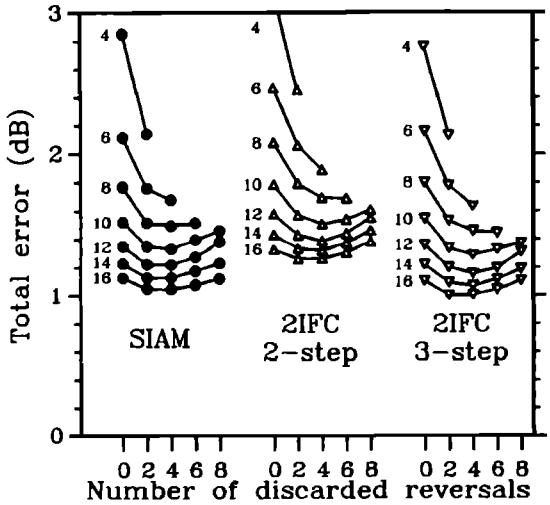

FIG. 6. Simulation: Total (systematic plus statistical) error as a function of the number of discarded reversals. The single curves are labeled with the total reversal number. Two to four reversals should be discarded, with four being better for shorter tracks.

the optimum number of reversals to discard. To this end, 3000 simulations of SIAM, 21FC 2-step and 2IFC 3-step procedures were carried out with fixed total reversal numbers in the range of 4–16. The total error (the orthogonal sum of systematic and statistical error) was evaluated as a function of the number of discarded reversals. For instance, for a total of ten reversals, four discarded reversals mean that only six reversals contributed to the calculation of the threshold estimates. Figure 6 gives the results. The single curves are labeled with the total reversal number. It is evident that there should be some discard, but the optimum number does not increase with increasing total number of reversals. Two or four discarded reversals are optimal, with a slight preference of four for the shorter tracks. Even if the total number of reversals is only six, it is better to discard four of them instead of two. Up to 14 total reversals are best treated by discarding 4. This is the result of a computer simulation. With human subjects a discard of four reversals should be even more preferable, as it will compensate better for a possible temporal concentration loss in the initial phase.

The best strategy seems to be to interleave tracks of at least 12 reversals, 4 of which should be discarded. If a threshold is to be estimated as a function of some parameter (e.g., frequency), several tracks of different kinds (i.e., frequency) should be interleaved to distribute the trials as homogeneously as possible in time. However, to improve the context of the tasks, the sections should not be too short. Groups of ten trials of a specific kind could be introduced by short demonstrations.

IV. EXPERIMENTAL TEST

The SIAM procedure has been used in psychoacoustic work since the summer of 1988. It is easy to learn and convenient to use, and the task as well as the adjustment rules seem clear and evident. With feedback, especially silent changes, the subjects learn quickly to adopt a critical attitude.

TABLE II. Average threshold signal level relative to the spectral power in the third octave band around 1 kHz, evaluated from all valid tracks for each subject and method. The discarded tracks are shown in relation to the total track number.

Subject		Yes–No		2IFC	
		von Békésy	SIAM	2-step	3-step
CK	level (dB)	0.5 ± 0.3	1.7 ± 0.3	0.3 ± 0.3	2.0 ± 0.15
	discards	0/14	1/14	0/14	2/14
LD	level (dB)	-2.9 ± 0.5	-0.4 ± 0.35	-0.6 ± 0.45	0.4 ± 0.3
	discards	1/14	3/14	1/14	2/14
ME	level (dB)	-4.0 ± 0.7	-0.3 ± 0.4	-1.7 ± 0.35	0.5 ± 0.25
	discards	1/15	2/15	2/15	0/15
TG	level (dB)	0.7 ± 0.65	2.0 ± 0.45	1.4 ± 0.3	2.8 ± 0.35
	discards	0/14	1/14	2/14	3/14

A. Method

The following presents a study measuring the efficiency for a typical psychoacoustic experiment performed with the adaptive methods simulated in Sec. III. The threshold for a sine tone in white noise was measured with each procedure until 18 reversals had occurred. The first two reversals were discarded. Sixteen or 17 tracks were carried out with each method; the first two tracks were discarded. Four subjects (three males and one female) participated in this study. Two of them (CK and LD) had much experience with psychoacoustic procedures, whereas for the other two (ME and TG) this was their first psychophysical experiment.

The experiment was controlled by a Masscomp Unix system. The subjects were seated in a soundproof booth and the signals were presented to them via Sennheiser 2002 headphones. The signals were digitally generated and converted by 16-bit converters at a rate of 40 kHz. The masker was a 600-ms burst of white noise with 20-ms cosinusoidal onset and offset ramps. The total noise power of the noise in the third octave band around 1 kHz was 36 dB SPL. The signal was a 1-kHz sine tone of 200 ms, with the same 20-ms ramps. It was centrally placed in the masker.

The subjects performed the different procedures in the order von Békésy, SIAM, 2IFC 2-step, 2IFC 3-step or inverse. A block of four tracks, one with each procedure, took approximately 12 min. The track started with 62-dB signal SPL (26 dB above noise level). The initial step size was 4 dB. It was halved after each of the first two reversals, so that the resulting step size was 1 dB. The 2- or 3-step rule for the 2IFC tasks was applied after the first reversal, whereas before, each correct response led to a decrease in signal level. The signal quota for the SIAM procedure was initially set to 75% and set to 50% after the first reversal. The matrix for $t = 0.5$ was applied (see Table I).

Each trial started 250 ms after the preceding response. The stimuli were presented, separated for the 2IFC tasks with a 250-ms pause. The possibility to answer was given from the beginning of the last presentation. The subject had to give the answer by pressing one of two buttons (yes/no or first/second). The response time was not limited. A feedback for mistakes was given (additional 400 ms), but not for correct responses. The average time for a trial was 2.1 s for the single-interval tasks and 3.2 s for the two-interval tasks.

B. Results and discussion

If the slope of a straight line, fit to the midpoints of every second run, exceeded 10 dB per track, this track was discarded. This procedure follows a suggestion of Hall (1983). Equally, if the first reversal of a track occurred below 22-dB signal SPL (that is 14 dB below noise level), it was discarded. The latter exclusion criterion became necessary as the starting phase of the tracks turned out to be too fast: at the threshold, four additional hits by chance led to a level 16 dB beneath the threshold. It would have been a better design to reduce the stepsize to 1 dB, slightly above the expected threshold, without waiting for a reversal. A total of 21 tracks out of 228 were discarded. Table II gives the average threshold signal level relative to the spectral power of the noise in the third octave band around 1 kHz. Under optimal conditions for hearing a sine tone in white noise this value should be -5 ± 2 dB. The obtained values of 1 ± 2 dB are compatible with the short signal duration.

Both yes–no task procedures claim to converge to the midpoint of the psychometric function. Assuming a nearly symmetric psychometric function of the 2IFC performance between 70% and 80%, this should lie halfway between 2IFC 2-step and 3-step convergence points or higher.[4] Let C_2, C_3, and C_{yn} be the 2IFC 2-step, 3-step, and the yes–no procedure convergence point. The relative position $R_{yn} = (C_{yn} - C_2)/(C_3 - C_2)$ should be compatible with or greater than 0.5. Table III lists the relative positions R_{yn} with their error bars for both yes–no task procedures. The

TABLE III. The relative position of the results of the yes–no task procedures between the results of the 2IFC 2-step and 3-step procedures. Only the SIAM results are compatible with 0.5, whereas the results of the von Békésy method lie systematically at too low a signal level and show a higher variance.

Subject	von Békésy	SIAM
CK	0.1 ± 0.2	0.8 ± 0.2
LD	-2.3 ± 1.7	0.2 ± 0.5
ME	-1.0 ± 0.5	0.6 ± 0.2
TG	-0.5 ± 0.6	0.4 ± 0.4
Average	-0.9 ± 0.9	0.5 ± 0.2

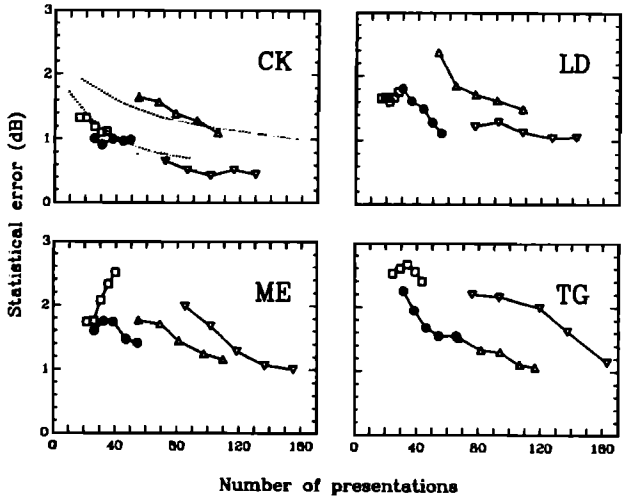

FIG. 7. Statistical error of the experimental threshold estimate for four subjects as a function of the presentation number in the track. Comparison between von Békésy (squares), SIAM (circles), 2IFC 2-step (upward triangles) and 3-step (downward triangles). The dotted lines in the panel of subject CK indicate the predictions of the Monte-Carlo simulation for SIAM and 2IFC.

errors of R_{yn} result from the errors of C_{yn}, C_2 and C_3. The results of the SIAM procedure are compatible with 0.5, indicating that the subjects successfully maintained their neutral response criterion—optimally operating at the peak of the ROC—and produced results compatible with the 2IFC results. However, the results of the von Békésy method lie systematically at too low a signal level and differ markedly from subject to subject. Hence, this method will not adjust comparable threshold levels for different subjects, and the sensitivity it claims to measure is not well defined.

Figure 7 shows the statistical error of the threshold estimate as a function of the presentation number. The latter is twice the trial number for 2IFC tasks. The five points that form a curve correspond to breakoff conditions of 8, 10, 12, 14, and 16 reversals, respectively. Breakoff conditions for lower reversal numbers were not included. To obtain reliable data on the efficiency of tracks with such few reversal, more than 16 tracks should have been obtained for each procedure and subject. The upper two panels of Fig. 7 correspond to the two experienced subjects, whereas the lower two panels give the results for the two unexperienced subjects. There is an interesting difference between these two groups. The experienced subjects reflect the theoretically predicted superiority of the 3-step rule (Kollmeier *et al.*, 1988), whereas for the unexperienced subjects the 2-step rule works better. For both groups the SIAM procedure is slightly better than the most efficient 2IFC method, and much better than the less efficient 2IFC method. The SIAM procedure works markedly faster without much loss of precision. It can be judged as an efficient procedure for all four subjects, whereas the efficiency of the 2IFC procedures is smaller and depends distinctly in this study on the rule.

The von Békésy rule is for all subjects a fast method with little standard error (cf. Hesse, 1986). However, this method is not objective, as the guessing behavior of the subjects will strongly influence the threshold estimates (see the discussion of the threshold estimates above; compare Kollmeier, 1988). The observed small standard error seems to be the result of a learned behavior that will vary from person to person. Variations in the starting point as well as the step size could increase the variability of the estimates markedly. In addition, it should be doubted that this behavior can be kept constant over several years. Thus the estimate should also vary in time. It does not correspond to any well-defined detectability measure. The learning process for this rule can be seen from the peculiar results obtained from the unexperienced subjects. The standard error is much higher with these results, and increasing presentation number may even increase the error. This could indicate that the behavioral response pattern was learned only up to a certain trial number (e.g., 25 trials for subject ME). Afterward, the subjects had no fixed response pattern that could have enabled them to manage the arbitrary nature inherent in the von Békésy method.

The dotted lines in the panel of subject CK indicate the predictions of the Monte-Carlo simulation (Fig. 4) for SIAM and 2IFC. The coincidence is quite good for SIAM and 2IFC 2-step, whereas 3-step is surprisingly better than the simulation. In comparing the yes–no task performance to the 2IFC performance at the same signal level (compare Fig. 3), the data of all tracks of one subject were pooled into signal level bins of 3 dB. The results are given in Fig. 8. The experiment was not designed to distinguish between the Gaussian and the threshold model of signal detection. However, the trend of the data seems to confirm the threshold model: the points are equally distributed above and below the diagonal, but only two points are above the Gaussian curve. This would justify the reasoning leading to a single-interval procedure: If the subject makes binary decisions, also with the forced-choice task, it is superflous to present more than one interval per trial.

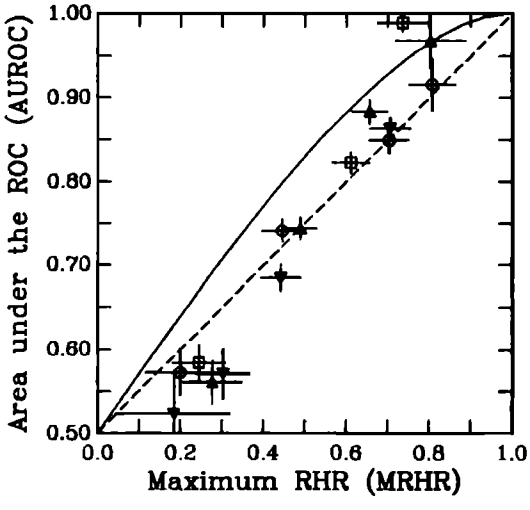

FIG. 8. Comparison between yes–no task outcome (MRHR) and 2IFC performance (AUROC) for the same signal level (bins of 3 dB SPL) and subject. Squares: CK (35–41 dB), circles: ME (32–41 dB), upward triangles: TG (35–44 dB), downward triangles: LD (29–38 dB). The diagonal corresponds to a low threshold model of signal detection, whereas the solid curve reflects a Gaussian observer.

V. CONCLUSIONS

The present study introduces a new adaptive staircase method, based on yes–no tasks instead of forced-choice tasks. Noise and signal presentations are randomly interleaved, and an appropriate adjustment matrix is used to control the subject's response criterion and to adjust the signal level to the desired target performance. A Monte-Carlo study and a test on four subjects compared the SIAM procedure to the von Békésy and to the 2-step and 3-step 2IFC methods. Whereas the von Békésy method needs about 30 presentations for 16 reversals and the SIAM procedure needs 50 presentations, the 2IFC procedures need 100 or more. Thus when applying 70% more presentations than von Békésy, one avoids response criterion effects, whereas 2IFC methods need 200% to 300% more presentations than the von Békésy method. In this way, the SIAM procedure provides a useful gain of efficiency. This is true especially for tasks where the presentation time can not be neglected in relation to the response time. Some applications might even favor the SIAM procedure only because of its special design: A yes–no task might be more appropriate for certain psychophysical studies than a forced-choice task. Furthermore, yes–no tasks give more information than forced-choice tasks, as they allow for differentiating the responses to signal and noise.

The experimental test was not exhaustive. The number of tracks per subject and procedure should have been higher so as to give detailed information about the efficiency of the procedures under study. However, the aim of this paper was not to compare well-known adaptive procedures, but to introduce a new adaptive procedure and to display its advantages. The Monte-Carlo study as well as the experimental test allow for judging the SIAM procedure as a fast, robust, and unbiased adaptive method.

ACKNOWLEDGMENTS

Egbert de Boer accompanied the study from its beginning and was always ready to give advice. Jean-Marie Aran and Yves Cazals welcomed the author in their lab and helped concretely with various scientific and nonscientific problems. Laurent Demany helped with the experimental design and gave helpful comments to the manuscript. The author wishes to thank Birger Kollmeier for helpful discussions. The author was supported by a poste vert from INSERM.

[1] The terms "two-alternative forced-choice" (2AFC) and "two-interval forced-choice" (2IFC) are sometimes confused. Whereas "interval" refers to the stimulus presentation, "alternative" refers to the response possibilities. A "two-interval three-alternative" (2I3A) task could contain the additional response possibility "do not know." The combination of the term "alternative" with "forced choice" is unfortunate, as the subject is always forced to choose among the response alternatives. A yes–no task has two possible responses, it could thus come under the term 2AFC. Furthermore, an "alternative" should be the other possible choice of a set of two. In contrast, it seems reasonable to speak of an "n-interval forced-choice" with respect to an n-interval task with n possible responses, namely, marking one of the n intervals. Therefore, the term 2IFC will be preferred.

[2] This restriction is arbitrary. Its purpose is to reduce the number of parameters of the adjustment matrix. The only reason for the choice $M_{cr} = 0$, $M_h = -1$ is the similarity to the von Békésy method for $t = 0.5$. The choice $M_{cr} = -1$, $M_h = 0$ works as well (the resulting matrix is turned 180°), and so does the choice $M_{cr} = M_h = -1$, which gives a symmetric matrix: $M_m = M_{fa} = (1 + t)/(1 - t)$. For the halfway target performance $t = 0.5$, the adjustment for correct responses (hit, correct rejection) would be a decrease of one step in signal level, whereas the adjustment for incorrect responses (miss, false alarm) would be an increase of three steps. The symmetric matrix has its advantages for small values of t, where it is more balanced. One remarkable difference is that it will lead more easily to reversals as it does not contain any zero entry. A comparative study could reveal whether this influences the efficiency.

[3] The systematic error of the von Békésy method depends strongly on θ. This parameter describes actually not a change in the response criterion (the latter was assumed to be neutral, producing the maximal RHR), but a change in the asymmetry of the underlying ROC curves. This can be the result of a change of attention, which affects the internal noise. It is remarkable that the von Békésy method fails to operate correctly even for a neutral response criterion. A Gaussian observer with a neutral response criterion would produce hit rates of more than 50% for any signal level, which would prevent the von Békésy rule from converging against a point of the psychometric function. The Gaussian observer will therefore be forced to operate with a non-neutral response criterion to make the von Békésy method produce some result.

[4] The MRHR at the convergence level of the SIAM procedure used here is 50%. This corresponds to a 2IFC performance (AUROC) between 75% and 87.5% (compare Fig. 3). The convergence point of the SIAM procedure will thus lie just halfway (75%) between 2IFC 2-step and 3-step performance or higher.

de Boer, E., and van Breugel, H. (1984). "Distribution of judgements in adaptive testing," Biol. Cybern. **50**, 343–355.

Emerson, P. L. (1984). "Observations on a maximum likelihood method of sequential threshold estimation and a simplified approximation," Percept. Psychophys. **36**, 199–203.

Findlay, J. M. (1978). "Estimates on probability functions: A more virulent PEST," Percept. Psychophys. **23**, 181–185.

Green, D. M., and Swets, J. A. (1974). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Green, D. M., Richards, V. M., and Forrest, T. G. (1989). "Stimulus step size and heterogeneous stimulus conditions in adaptive psychophysics," J. Acoust. Soc. Am. **86**, 629–636.

Hall, J. L. (1968). "Maximum-likelihood sequential procedure for estimation of psychometric functions," J. Acoust. Soc. Am. **44**, 370.

Hall, J. L. (1974). "PEST: Note on the reduction of variance threshold estimates," J. Acoust. Soc. Am. **55**, 1090–1091.

Hall, J. L. (1981). "Hybrid adaptive procedure for estimation of psychometric functions," J. Acoust. Soc. Am. **69**, 1763–1769.

Hall, J. L. (1983). "A procedure for detecting variability of psychophysical thresholds," J. Acoust. Soc. Am. **73**, 663–667.

Hesse, A. (1986). "Comparison of several psychophysical procedures with respect to threshold estimates, reproducibility, and efficiency," Acustica **59**, 263–273.

Kaernbach, C. (1990). "Poisson signal detection theory," Percept. Psychophys. (submitted).

Kollmeier, B. (1988). "Adaptive AFC-methods in audiology," Audiol. Akustik **6**, 168–179.

Kollmeier, B., and Gilkey, R. H. (1983). "Adaptive staircaise methods—A comparison using Markov theory," J. Acoust. Soc. Am. Suppl. 1 **73**, S104.

Kollmeier, B., Gilkey, R. H., and Sieben, U. K. (1988). "Adaptive staircase techniques in psychoacoustics: A comparison of human data and a mathematical model," J. Acoust. Soc. Am. **83**, 1852–1862.

Levitt, H. (1971). "Transformed up–down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Luce, R. D. (1963a). "A threshold theory for simple detection experiments," Psychol. Rev. **70**, 61–79.

Luce, R. D. (1963b). "Detection and recognition," in *Handbook of Mathematical Psychology*, edited by R. D. Luce, R. R. Bush, and E. Galanter (Wiley, New York), Vol. 1, pp. 103–189.

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (1986). "Thresholds for hearing mistuned partials as separate tones in harmonic complexes," J. Acoust. Soc. Am. **80**, 479–483.

Pentland, A. (1980). "Maximum likelihood estimation: The best PEST," Percept. Psychophys. **28**, 377–379.

Schlauch, R. S., and Rose, R. M. (1986). "A comparison of two-, three-, and four-alternative forced-choice staircase procedures," J. Acoust. Soc. Am. Suppl. 1 80, S123.
Shelton, B. R., and Scarrow, I. (1984). "Two-alternative versus three-alternative procedures for threshold estimation," Percept. Psychophys. 35, 385–392.
Shelton, B. R., Picardi, M. C., and Green, D. M. (1982). "Comparison of three adaptive psychophysical procedures," J. Acoust. Soc. Am. 71, 1527–1533.
Taylor, M. M., and Creelman, C. D. (1967). "PEST: Efficient estimates on probability functions," J. Acoust. Soc. Am. 41, 782–787.
Taylor, M. M., Forbes, S. M., and Creelman, C. D. (1983). "PEST reduces bias in forced choice psychophysics," J. Acoust. Soc. Am. 74, 1367–1374.