Rapid Evaluation of Perceptual Thresholds

The Best-Pest Calculator: A web-based application for non-expert users

Hans-Jörg Zuberbühler

Institute for Hygiene and Applied Physiology (IHA)
Swiss Federal Institute of Technology Zurich (ETHZ)
CH-8092 Zürich
zuberbuehler@iha.bepr.ethz.ch

Table of Contents

As an methodical outcome of the threshold experiments conducted at the IHA, we advanced the used best-PEST method to a fully independent, browser-based application. The idea was to provide experimenters with a tool for measuring thresholds, which can be used without spending any installation, compilation, or even programming effort (this is in contrast to other available software). The drawback of this premise lies in the missing interface. For safety reasons the program has no access to the client computer and therefore can not provide it with the estimated values directly. The experimenters have to insert the received threshold values in their testing environment by hand. This fact makes the Best-PEST Calculator useful especially for these threshold estimations, whose stimulus presentation can not be done with the aid of common computer-equipment, like e.g. smell and taste thresholds. The program can be downloaded from:

http://www.psychophysics.ethz.ch/tools/

Depending on the version used, the browser has to be updated with the Macromedia Director plug-in version 8.5. The software recognises automatically if an update is necessary, whereupon it will be done after three or four mouse-clicks.

In the following we describe the minimal background necessary to utilize the best-PEST-Calculator.

1   Psychophysical Theory

1.1   Testing paradigms

Psychophysical procedures dispose of various testing paradigms, of which we describe the *yes-no* and the *forced-choice* (nAFC: n-**a**lternative-**f**orced-**c**hoice) mode. With the yes-no mode subjects are given a series of trials, in which they must judge the presence or absence of a stimulus at each case. The ratio between the number of trials containing a stimulus and the total number of trials is usually 0.5, but can be any other value. Usually this ratio is told to the subject in advance. The rate of *yes*-responses for all tested stimulus intensities is defined as the dependent variable.

A basically different testing mode is represented by the forced-choice mode: Subjects are given a variety of n alternatives, from which they have to choose the one containing the stimulus. The alternatives are presented with either spatial or temporal coincidence, or without either coincidence. The subjects know that exactly one alternative contains the stimulus, and that the rest has a zero-stimulus. The differences between these two methods become obvious when the presented stimuli are faint. In the yes-no paradigm the proportion of yes-answers approaches zero, whereas in the forced-choice paradigm the proportion of correct answers approaches the value of equal probability for all alternatives, which is the reciprocal value of the number of alternatives. Likewise this means that e.g. in two-alternative forced-choice (2AFC) tasks the threshold is located where observers give 75% of correct responses, since they already give 50% of correct responses due to the 2AFC-inherent guessing. The basic advantage of 2AFC consists of its well-founded assumption that subjects will opt for the stimulus evoking the strongest perception, regardless their tendency to say "yes" or "no". This is in contrast to the yes-no paradigm, where decision making in the presence of uncertainty is according to the subject's psychological characteristics, like e.g. prudence. Unlike the *yes-no* mode, the dependent variable of nAFC is the rate of correct responses for all tested stimuli instead of the rate of *yes*-responses. In the following we subsume both kinds of dependent variables under the term *positive-response rate ψ*.

For most of psychophysical testing, be it in the clinic or in the research lab, efficiency is of great importance, i.e. the threshold should be estimated with satisfying accuracy after as few as possible trials. The requirement of minimal number of trials is given by the fact, that after a long run of trials experimental subjects tend to fatigue and to be bored, resulting in an apparently drift of their thresholds. For this reason, so called adaptive psychophysical procedures have been developed, whose prior purpose is to minimize the number of trials. We will recapitulate the adaptive procedure called *best-PEST* in chapter 2, for more details about adaptive procedures see the overview of Treutwein (Treutwein, 1995). In the next chapter we describe the theoretical background necessary to understand this procedure.

1.2   *Psychometric function* $\psi = f(\phi)$

The psychometric function assigns a positive-response rate ψ to the range of stimulus intensities. The particular properties of this function are described in the following:

The range of ψ is bounded as lower limit by the probability to give positive responses without perceiving the stimulus (false positive rate). This false positive rate consists of a methodical part (only in nAFC), and the "proper" false positive rate ε. The methodical part is equal to the reciprocal value of the alternatives n. The upper limit of ψ consists of $(1-\delta)$: Big stimulus intensities effect positive responses in virtually all the cases, only reduced by the false negative rate (i.e. misses) δ. The error terms δ and ε are caused by observers' inattention or fatigue for instance.

$$\psi_{-\infty} = p\left(\text{positive response}\middle|\phi \to -\infty\right) = \frac{1}{n} + \varepsilon \qquad\qquad \text{eq (1)}$$

$$\psi_{+\infty} = p\left(\text{positive response}\middle|\phi \to +\infty\right) = 1 - \delta \qquad\qquad \text{eq (2)}$$

ϕ	: stimulus intensity	$\{$	$\phi \in \mathbb{R}$ $\}$
n	: number of alternatives	$\{$	$n \in \mathbb{N}$ $\mid$ $2 \le n \le 100\}$
ε	: false positive	$\{$	$\varepsilon \in \mathbb{R}$ $\mid$ $0 \le \varepsilon \le 0.5\}$
δ	: false negative	$\{$	$\delta \in \mathbb{R}$ $\mid$ $0 \le \delta \le 0.5\}$

We define the threshold θ to be that value of stimulus intensity, that yields a specified positive-response rate. For practical reasons in testing, the threshold is located at the steepest slope of the psychometric function (derivation see chapter 2.1.). In the following we will exemplify the psychometric function by means of the logistic model, because this is the kernel function of the adaptive procedure *best-PEST*, which is the topic of chapter 2.1:

$$\psi^*(\phi) = \left(1 + e^{\beta^*(\theta-\phi)}\right)^{-1} \qquad\qquad \text{eq (3)}$$

$\psi^*(\phi)$: kernel function

β^* : steepness parameter

θ : threshold

Since the logistic function is rotationally symmetric in the inflection point, the threshold is in middle of the response range $[\psi_{-\infty}, \psi_{+\infty}]$. Therefore, the rate of positive responses at threshold is:

$$\psi_{\theta(nAFC)} = p\left(\text{positive response}\middle|\phi = \theta\right) = \frac{\psi_{+\infty} + \psi_{-\infty}}{2} = 0.5\left(1 - \delta + \frac{1}{n} + \varepsilon\right) \qquad \text{eq (4)}$$

In order to create a formal link between the two testing paradigms, the *yes-no* situation can be considered as *forced-choice* situation with infinitive number of alternatives. In this case the threshold converges to the value where the positive-response rate is:

$$\psi_{\theta(Yes/No)} = \lim_{n \to \infty} 0.5\left(1 - \delta + \frac{1}{n} + \varepsilon\right) = 0.5\left(1 - \delta + \varepsilon\right) \qquad \text{eq (5)}$$

The psychometric function $\psi^*(\phi)$ has to be adjusted due to the observers false positive and false negative rates. For these purposes the kernel function is shifted to $n^{-1} + \varepsilon$ and scaled to the response range $[\psi_{-\infty}, \psi_{+\infty}]$, which distance is – according to eq (1) and eq (2) – equal to $\left|1 - \delta - n^{-1} - \varepsilon\right|$:

$$\psi(\phi) = n^{-1} + \varepsilon + \left(1 - \delta - n^{-1} - \varepsilon\right)\psi^* \qquad \text{eq (6)}$$

$\psi(\phi)$: adjusted psychometric function

In order to deal with a well known constant, which is comparable between different magnitudes of stimuli, we let β be the slope of the inflection point of the normalized psychometric function. We define the threshold to be at stimulus intensity of 0.5, thus we normalize the stimulus intensity to two threshold units, with the result of obtaining the "real" slope in a equal-scaled plot (i.e. the slope is equivalent to the tangent of the gradient angle):

$$\beta = \left.\frac{d\psi}{d\phi}\right|_{\phi=\theta} = \frac{\beta^*\left(1 - \delta - n^{-1} - \varepsilon\right)}{4} \qquad \text{that is} \qquad \beta^* = \frac{4\beta}{1 - \delta - n^{-1} - \varepsilon} \qquad \text{eq (7)}$$

$\beta \qquad$: slope of the psychometric function at threshold (inflection point)

eq (7) inserted in eq (6) leads to:

$$\psi(\phi) = n^{-1} + \varepsilon + \left(1 - \delta - n^{-1} - \varepsilon\right)\left(1 + e^{\frac{4\beta(\theta-\phi)}{\left(1-\delta-n^{-1}-\varepsilon\right)}}\right)^{-1} \qquad \text{eq (8)}$$

Equation eq (8) is the underlying, generic formula for the threshold estimation by the best-PEST calculator. In Figure 1 the mapping of eq (8) is shown with different parameter settings:

$n \in \{2,\ 4,\ \infty\}$

$\beta \in \{1.5,\ 3,\ 7\}$
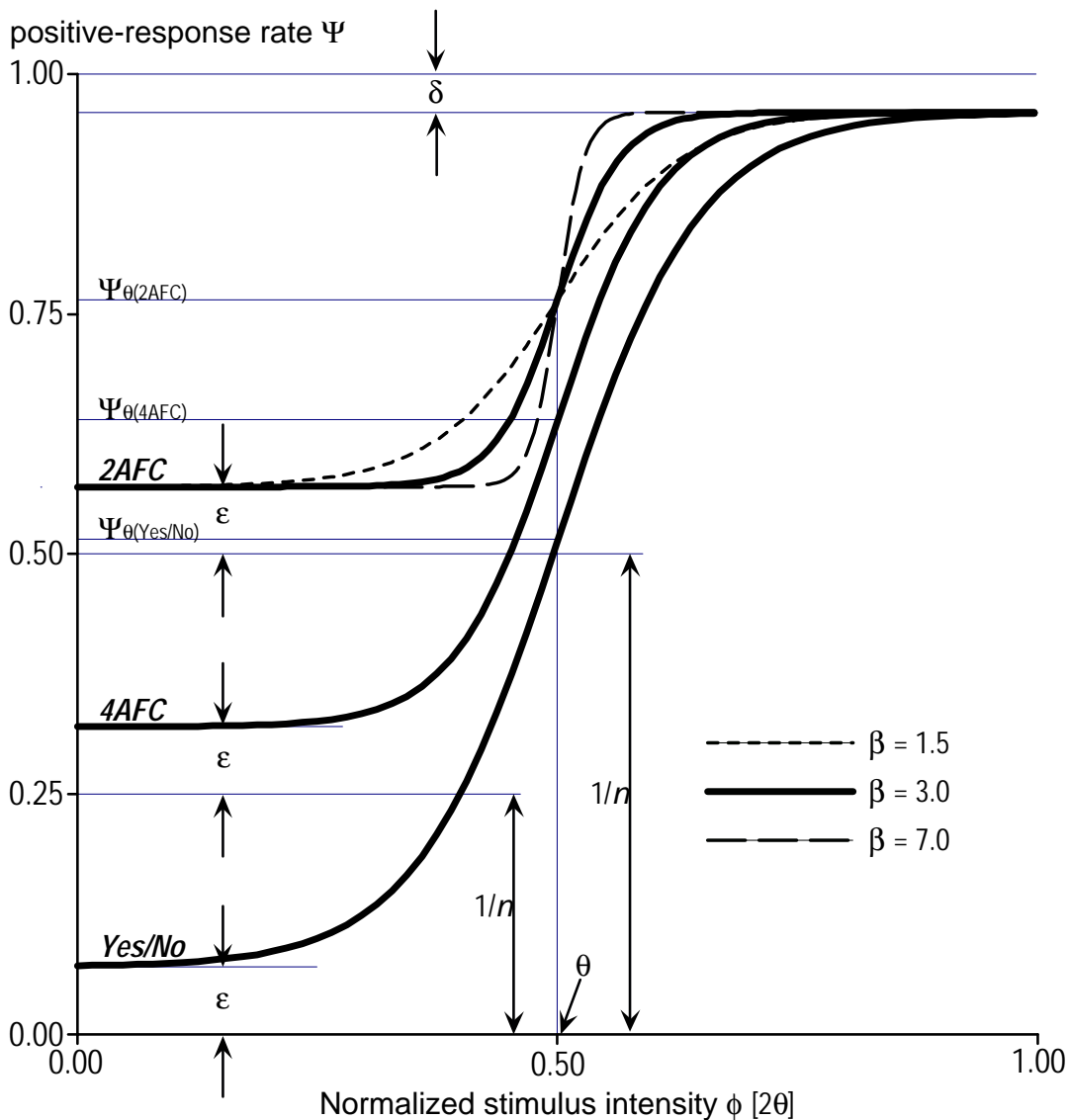
$\varepsilon = 0.07$, $\delta = 0.04$

Figure 1   Logistic psychometric graphs depicting yes-no and forced-choice situations (nAFC). The asymptotes are at $(1/n + \varepsilon)$, and at $(1-\delta)$. The slope β is 3 (straight lines), and 7 and 1.5 (dashed lines) respectively. The stimulus intensity is normalized to 2 threshold units.

Typically psychometric functions are – as depicted in Figure 1 – of statistical value (unless they represent a heaviside step function with its "step" at the threshold value). I.e. when an observer is presented on several occasions with the same stimulus, he or she is likely to respond yes on some trials and no on other trials. Thus, the threshold cannot be defined as the stimulus value below which detection never occurs and above which detection always occurs, but rather as the stimulus value which is perceptible in a predefined percentage of the trials (usually 50%). Experimenters are confronted with the question, how to determine the psychometric function of an experimental subject or of a study cohort. For that purpose classical psychophysics offers several methods, which we will not explain here in detail. Readers interested in this topic may consult the standard work of Gescheider (Gescheider, 1997). Recapitulating, we hold that with these methods we determine the detectability of several stimulus intensities, and fit an appropriate sigmoid shaped

curve to these data to obtain the psychometric function. From this function the 50% threshold can be read out.

In order to measure the empirical threshold, the experimenter must decide what stimulus intensities should be used in the experiment. It should be clear that choosing intensities that are all greatly above or below the threshold will provide little information leading to an accurate estimation of the threshold. In addition to the problem of requiring a large quantity of trials to obtain the threshold, waste trials are likely to occur with these methods, unless the testing range is known in advance. An approach with these characteristics is far from optimally efficient and consequently the adaptive methods for measuring threshold have evolved.

2 Adaptive Psychophysical Procedures

In all adaptive procedures, the intensity of a stimulus presented on a particular trial is determined by the observer's performance in detecting stimuli presented on prior trials. Except for one class of procedures called *maximum-likelihood methods* all other methods described in (Gescheider, 1997) suggest more or less heuristic rules after how many trials and how much the presented stimulus intensity has to be adjusted. Even though it is a characteristic of all adaptive procedures to recall information from the past history of an experimental run, only the maximum-likelihood procedures determine the next stimulus presentation based on a statistical estimation of the observer's threshold, which is made from all of the results obtained from the beginning of the run. The statistical technique of maximum-likelihood estimation assumes that the underlying psychometric function has a specific form. For example it could be a Gaussian (the cumulative normal distribution), logistic, Weibull, or some other sigmoid-shaped function. Because these functions have similar forms, the estimated thresholds are not greatly different, and the choice may only be of importance if e.g. a particular perception model is under test. In the following we describe the best-PEST method suggested by (Pentland, 1980). PEST is the acronym for *Parameter Estimation of Sequential Trials*.

2.1 Maximum-Likelihood: best-PEST

In best-PEST the approach taken to the problem of determining a threshold is to maximise the information gained with each measurement. In so doing the smallest possible number of measurements will be required. First we derive the choice of the sampling point on the psychometric function:

For any value ϕ of the stimulus range $[0,k]$, there is a probability Ψ of a positive answer. Given n samples taken at ϕ, of which p were positive, our estimate of Ψ is:

$$\widehat{\psi} = \frac{p}{N}$$

eq (9)

$\widehat{\psi}$: estimate of the probability of a positive response

p : number of positive responses

N : number of samples

the variance is

$$\sigma = \frac{\widehat{\psi}(1-\widehat{\psi})}{N}$$

eq (10)

σ : variance of estimation

and the confidence intervals are

$$CI_{\widehat{\psi}} = w\sqrt{\sigma}$$

eq (11)

$CI_{\widehat{\psi}}$: width of the confidence interval about $\widehat{\psi}$

w : level of desired confidence (e.g. 0.95)

Equations eq (9) and eq (10) inserted in eq (11) leads to

$$CI_{\widehat{\psi}} = w\sqrt{\frac{p(N-p)}{N^3}}$$

eq (12)

To get the stimulus range ϕ corresponding to the confidence interval of the dependent variable, it has to be divided by the slope of the psychometric curve:

$$CI_{\widehat{\phi}} = \frac{CI_{\widehat{\psi}} \cdot d\phi}{d\psi}$$

eq (13)

$CI_{\widehat{\phi}}$: width of the confidence interval about $\hat{\phi}$

Thus, in order to minimise the estimated confidence interval about the stimulus ϕ for a given number of trials we have to maximise the slope of the psychometric function. For all sigmoid-shaped functions, the steepest slope is located at the inflection point. In the rotationally symmetric logistic function used in best-PEST this point is at the "center" of the curve. In the yes-no mode this is at 50% if E=0 and S=1; in the 2AFC mode this is at 75% if E=0.5 and S=0.5.

In order to explain the best-PEST procedure we reformulate Equation eq (8) and obtain the probability of getting a positive (if r=1) or negative (if r=-1) response at the i-th trial:

$$\widehat{\psi_i^{+-}}\left(\phi, r_i, \widehat{\theta_i}\right) = E + S\left(1 + e^{r_i\left(\widehat{\theta_i}-\phi\right)4\beta S^{-1}}\right)^{-1}$$

eq (14)

r_i : response of the observer at i-th trial. $r_i \in \{1, -1\}$

$\widehat{\theta_i}$: i-th estimate of the threshold

E : elevation of the psychometric function according to eq (1)

S : scaling of the psychometric function to the response range according to eq (1) and eq (2)

The strategy in best-PEST is to calculate the likelihood of the sampling point's being at each point within the testing range and taking as new estimate the stimulus value that is assigned to the highest probability. After n-1 trials, we find the n-th point of measurement by solving:

$$\widehat{\theta_N} = \max_{\phi \in (0,k)} P\left[x \text{ is at maximum slope}|(\theta_1, r_1),...(\theta_{N-1}, r_{N-1})\right],$$

eq (15)

where (0, k) is the test range of the stimulus ϕ and (θ_i, r_i) denotes the results of the i-th measurement that was taken at value θ_i.

The maximum likelihood estimator is known to be the most efficient unbiased estimator. One problem arises: the product of all the probability distributions approaches zero for large numbers of trials. To overcome this problem, we apply a logarithmic transformation to the likelihood function with the result of obtaining the sum instead of the product of all likelihood functions. That way, the log-likelihood functions do not need to be standardised to the overall probability of 1. Since the logarithmic function is strictly monotonic increasing, the locations of maxima are preserved:

$$\max_{x \in (a,b)} \prod_{i=1}^{N} f(x) = \max_{x \in (a,b)} \sum_{i=1}^{N} \log f(x) \qquad \text{eq (16)}$$

For the case of the used function eq (14), the n-th threshold estimation is calculated according to eq (15) and eq (16):

$$\widehat{\theta_N} = \max_{\phi \in (0,k)} \sum_{i=1}^{N-1} \log\left(E + S\left(1 + e^{r_i\left(\widehat{\theta_i}-\phi\right)4\beta S^{-1}} \right)^{-1} \right) \qquad \text{eq (17)}$$

Figure 2 depicts the expansion of the log-likelihood functions according to eq (17). The following parameter settings are used:

Parame-	Used Paradigm	
	A = 2AFC	B = yes-no
N	10	10
E	0.5	0
S	0.5	1
β	2	2
r	{1, 1, -1, 1, 1, -1, 1, 1, 1, -1}	{1, -1, 1, -1, 1, 1, -1, -1, 1, -1}

Table 1 Parameter settings of the curves depicted in Figure 2.

Hans-Jörg Zuberbühler, ETH Zürich   zuberbuehler@iha.bepr.ethz.ch
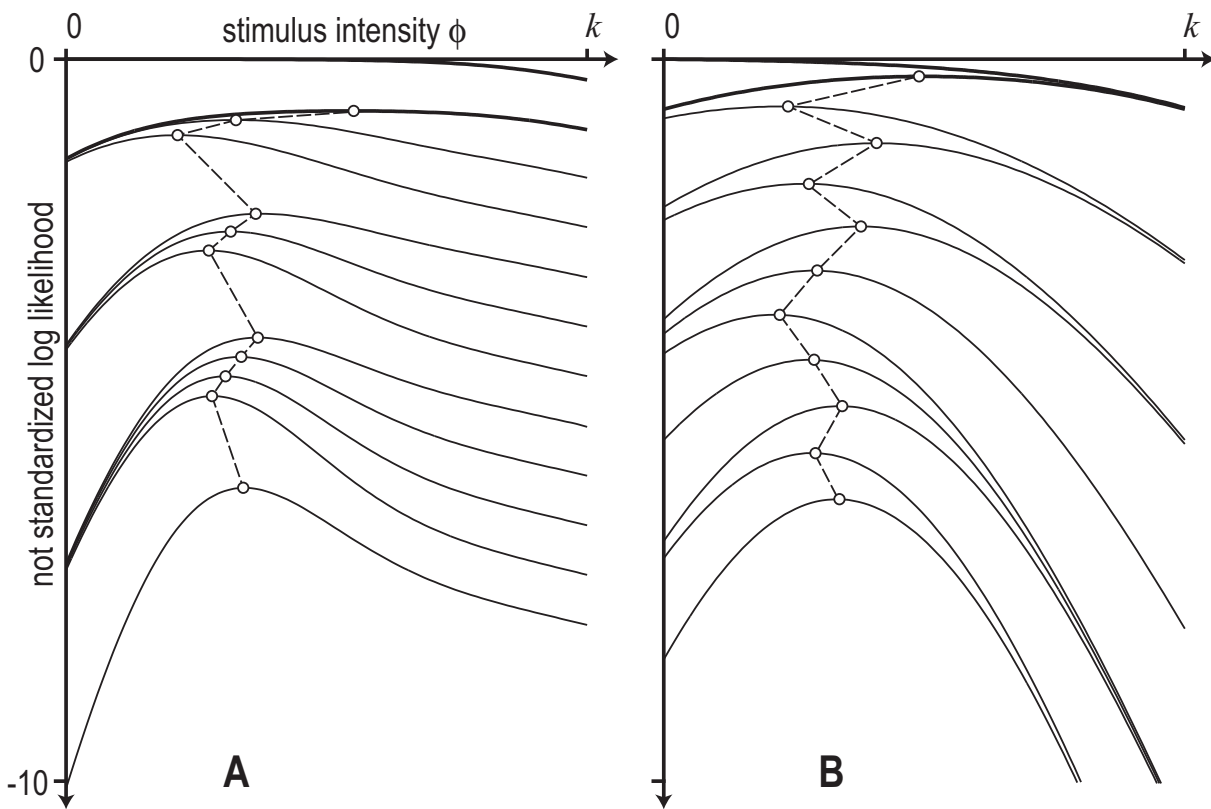
Figure 2   Expansion of the log-likelihood functions in the stimulus interval [0, k] of the adaptive procedure *best-PEST*. Circles indicate the relative maxima, dashed lines show the progression of the threshold convergence. Bold lines represent the predefined initialisations, thin lines are calculated according to the responses *r*.   **A**: 2-alternativ forced-choice (2AFC) paradigm. **B**: yes-no paradigm.

3   Description of the best-PEST Calculator

In the following the best-PEST Calculator is described. Screenshots of the three masks of the program are shown and the input and output fields are explained where they are not self-explanatory (indicated by numbers).
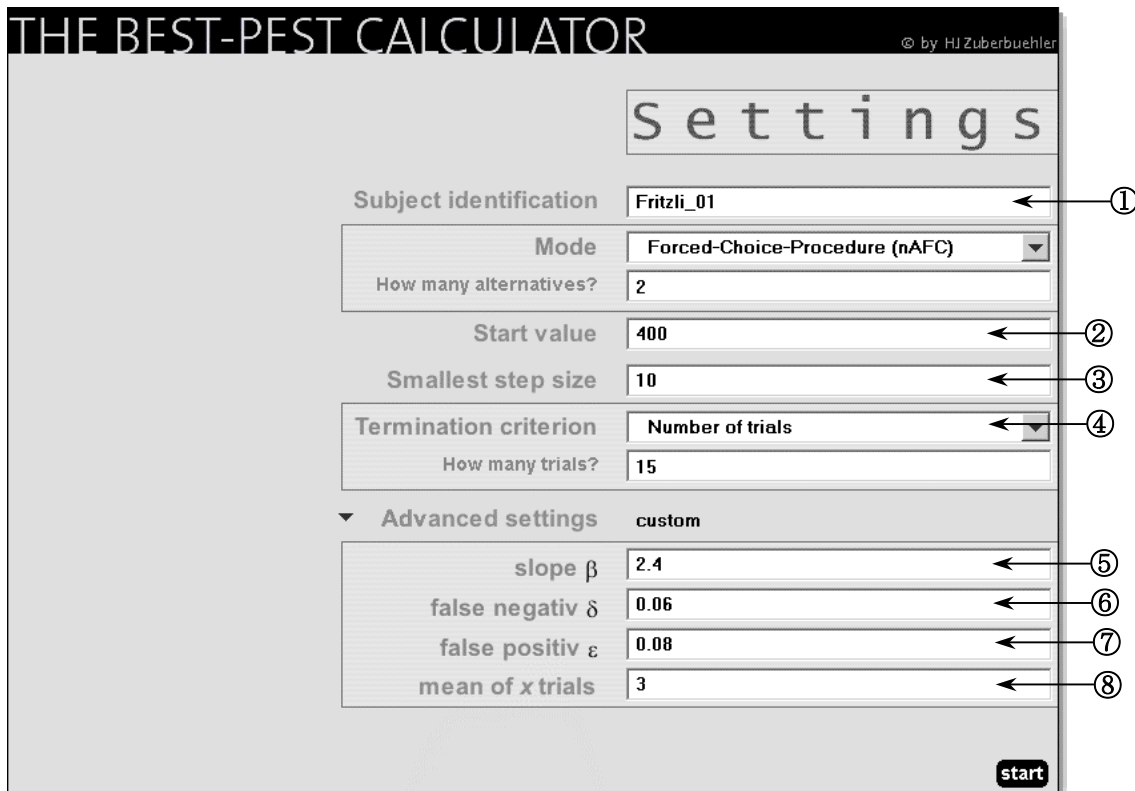

Figure 3   Screenshot of the first mask (input), where the settings for the experiment are entered. If all the fields are filled out in the requested format, pressing the "start" button will lead to the second input mask. If not, a dialogue window pops up, indicating the missing or false input. Clicking the arrow opens the "advanced settings" fields. By default these settings are: "slope β" = 2, "false negative δ" = 0," false positive ε" = 0, "mean of x trials" = 3.

① Mode
In the drop-down menu *mode*, the users have the choice between the *yes-no* and the *forced-choice* (*n*AFC) paradigm. If they choose *n*AFC, an additional input field appears, where the number of alternatives *n* is to insert. If *n* > 100 is entered, the program switches automatically to the *yes-no* calculation mode. It is to state that experimental subjects most likely will be overstrained if they have to make repeated de-

cisions about the presence of a stimulus from more than hundred alternatives. Anyhow, if such experiments are planned, one can expect the error caused by the slightly inadequate calculation being much smaller than the error caused by any other interference – for instance the subject's lapses.

② Start value k

Setting of the test interval [0, k], where k determines the highest stimulus value that can be obtained during the run. The upper limit k should be at least twice as large as the expected threshold value. Note that the start value will not be presented to the subject, assuming this value so high that subjects will perceive it in all the cases. In order to deal with comparable slope values, the algorithm uses the normalized range [0, 1] of the stimulus intensity. The stimulus intensity ϕ denote therefore:

$$\phi = \frac{\phi^*}{k} \qquad\qquad \text{eq. (18)}$$

ϕ^*: stimulus intensity in desired unit $\qquad \left\{ \phi^* \in \mathbb{R} \;\middle|\; \phi^* \geq 0 \right\}$

k : stimulus maximum $\qquad\qquad\qquad \left\{ k \in \mathbb{R} \;\middle|\; k > 0 \right\}$

③ Smallest Step Size

Determines the size of the smallest stimulus change that can be obtained. Ideally this is the difference threshold of the particular stimulus. If this value is not known – in the case where we just want to determine it – we have to estimate a suitable step size. Experimenters have to be aware of too small or too big step sizes, since both result in large measurement bias of the thresholds. If the ratio between "start value" and "smallest step size" is larger than 1000, the program will prompt a warning and ask for either bigger step size or smaller start value. This is a precaution in order to prevent the users of too long computing time.

④ Termination Criterion

Users have the choice between "Number of Trials" and "Number of Reversals". A reversal R is defined as a change from increasing to decreasing (or the other way around) of the presented stimulus intensities M.

$$M = \left\{ m \in \mathbb{R} \;\middle|\; m \text{ is presented at trial } i \right\} \qquad \text{eq. (19)}$$

$$R = \left\{ m_i \in M \;\middle|\; \left(m_{i-1} > m_i < m_{i+1} \right) \vee \left(m_{i-1} < m_i > m_{i+1} \right) \right\} \qquad \text{eq. (20)}$$

M : set of presented stimulus intensities

R : set of reversals

⑤ Advanced Setting: Slope β

As an advanced setting, the users have the opportunity to enter the estimated or known slope of the particular psychometric function. For the definition of the slope see Figure 1 and eq (7). The slope value is calculated according to equal-scaled axes. Entering β implies knowledge about the tested cohort or subject, usually gained through pre-testing. If the slope is not known, β will be set by default to two.

⑥ Advanced Setting: false negative δ

δ specifies the false negative rate (or miss rate). This rate is constituted by the observers negative answers even tough the stimulus intensity is at maximum. Entering δ implies knowledge about the tested cohort or subject, usually gained through pre-testing. By default this value is zero.

⑦ Advanced Setting: false positive ε

ε specifies the false positive rate (or false alarm rate). This rate is constituted by the observers positive answers even tough the stimulus intensity is zero. In forced-choice experiments, ε does not comprise the methodical false alarm rate, which is the reciprocal value of the number of alternatives. Entering ε implies knowledge about the tested cohort or subject, usually gained through pre-testing. By default this value is zero.

⑧ Advanced Setting: mean of x trials

x specifies the number of trials to take at the end of an experimental run for calculating the mean threshold value. As a rule-of-dump, larger numbers of trials permit larger numbers of x. By default this value is three.
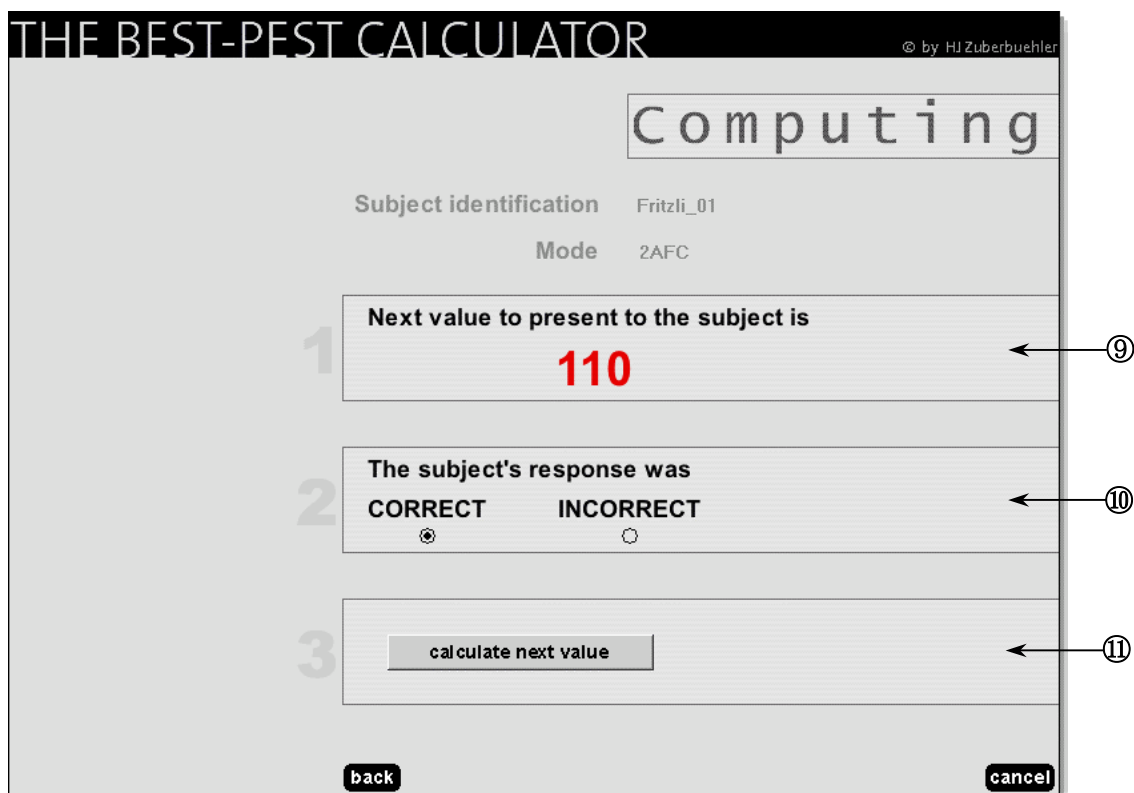

Figure 4   Screenshot of the second mask (input/output), where the computation of the actual maximum likelihood threshold is done. Pressing the button "back" will abort the computation and returns to the first mask to modify the settings. Pressing the "cancel" button will abort the computation and goes to the results mask displaying the recent status of the experiment, without having reached the termination criterion.

⑨ **Step 1: output from the best-PEST algorithm**

The output value m_i is to present to the subject. It is the maximum likelihood estimation of the threshold, obtained from all available information. Since there is no information available from the subjects in the very first trial, the initialisation is conducted assuming that the subjects will perceive the stimulus for sure at the start intensity k, and that at zero intensity they will not perceive the stimulus for sure. Therefore the first output will be somewhere in the middle of the test interval.

⑩ **Step 2: response of the subject**

After the subjects were presented with the stimulus intensity obtained from step 1, the radio button is to select corresponding to the subject's response. In the nAFC mode the buttons are labelled with "CORRECT" and "INCORRECT", and in the *yes-no* mode they are labelled with "YES" and "NO".

⑪ **Step 3: next value**

Pressing the button "calculate next value" will trigger the next calculation, whereupon a new value will appear in the output field. Step 1 to 3 have to be repeated until the termination criterion is reached. Pressing then this button will bring to program to the "results" mask.
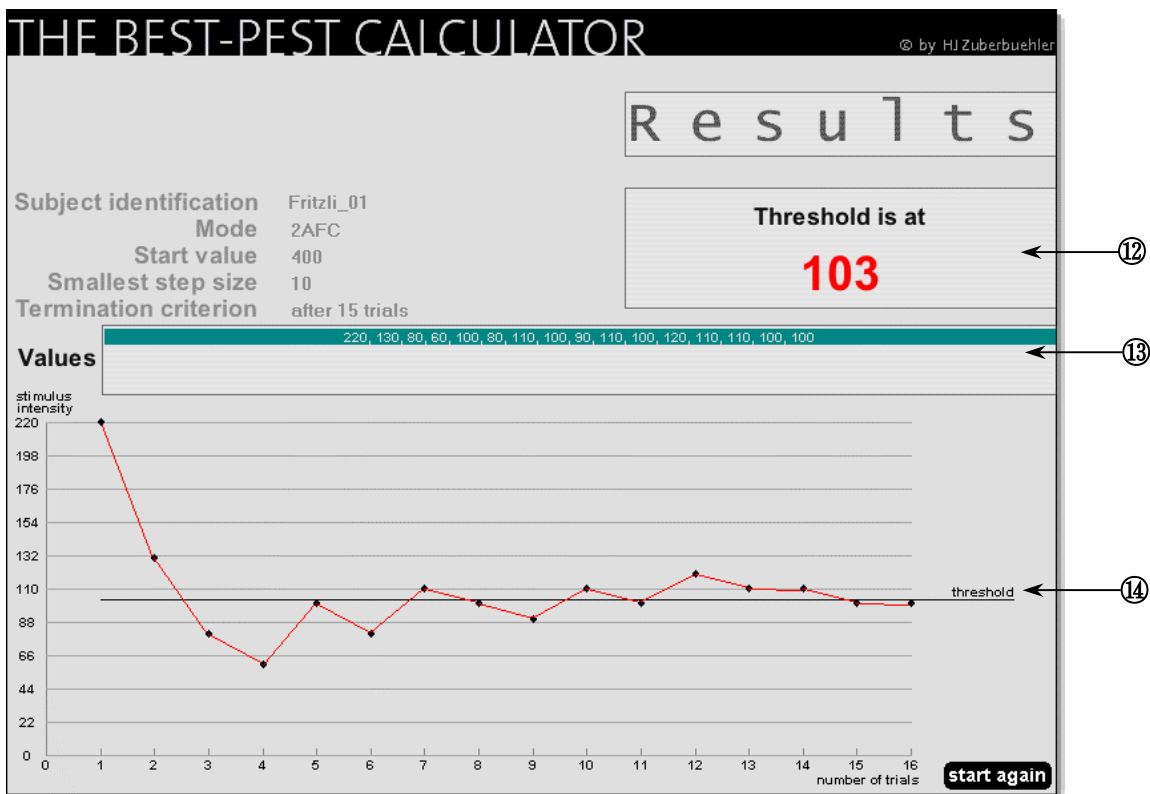

Figure 5   Screenshot of the third mask (output), where the results of the entire experimental run are displayed. Pressing the "start again" button will return to the first mask, and leave the settings as they are.

⑫ **Threshold value**

Output of the final threshold estimation, which is the mean value of the x last trials.

⑬ **All values**

The presented stimulus intensities of the entire experimental run are displayed and marked in the field "values" in order to copy them to the clipboard (Ctrl + C).

⑭ **Graph**

The values of the entire experimental run as well as the final threshold are depict in a diagram with stimulus intensity as ordinate and number of trials as abscissa.

4   Monte-Carlo Simulations

The following Monte-Carlo-Simulations were made to evaluate the convergence behaviour of the best-PEST algorithm. All simulations were made in the *yes/no* mode with equal start values. A built-in random process simulated the response behaviour of an assumed experimental subject which we call *stochastic observer*. For that purpose we assumed that the stochastic observer answers in a logistic manner with a stable threshold – an assumption which is in fact made by best-PEST:

According to eq (17) on page 11, $\widehat{\theta_N}$ is the n-th threshold estimate done by best-PEST. For this estimate there is – according to eq (8) on page 5 – a probability $\psi\left(\widehat{\theta_N}\right)$ for a positive response. We obtain the particular answer of the stochastic observer by applying the following procedure: If $\psi\left(\widehat{\theta_N}\right)$ is greater than a jointly distributed random number between 0 and 1, the stochastic observer answers *no*, if $\psi\left(\widehat{\theta_N}\right)$ is equal or smaller than the random number, the stochastic observer answers *yes*. That way, after a sufficient number of runs we map the whole assumed psychometric function of the stochastic observer onto the outcome of the best-PEST procedure, and we are possibly able to establish an empirical law of the algorithm's behaviour.

In the following we show the results of three simulation runs. 0 lists the corresponding parameter settings for the conducted simulations, whose results are displayed in Figure 6, Figure 7, and Figure 8.

Parameter	Value		
	Figure 6	Figure 7	Figure 8
Mode	yes/no	yes/no	yes/no
Start value k	1.7391	1.7391	1.7391
Threshold θ of the stochastic observer	1.0000	1.0000	1.0000
Start value k / smallest step size	40	40	40
Termination criterion: Number of Trials	15	5 to 50	50
Slopes of best-PEST's model	1.0 to 3.5	0.1 to 5.0	0.1 to 5.0
Slopes of the stochastic observer's psychometric function	same steps	same steps	0.1 to 5.0
False negative δ	0	0	0
False positive ε	0	0	0
Mean of x trials	3	3	3
Number of threshold determinations per measuring point	3	1000	1000
Number of measuring points	2500	2500	2500

Table 2 Parameter settings used for the Monte-Carlo-Simulations separated for three conditions. For an explanation of the parameters see the previous chapter.

In order to gain an idea of what accuracy the best-PEST algorithm provides, we ran a simulation with realistic parameter settings: As a trade-off between accuracy and practicability, the subject would have to accomplish three threshold determinations consisting of 15 stimulus presentations with corresponding decision-making. In such a way the whole procedure would last the feasible time of about 30 minutes, which is of course depending on the duration of each stimulus presentation. Anyway, with such a scenario experimenters can be sure, that the subjects' fatigue will play an negligible role. For the simulations, we ran the above-mentioned scenario with slopes from 1.0 to 3.5, resulting in a total amount of 2500 threshold means. The histogram of this distribution is depicted in Figure 6.
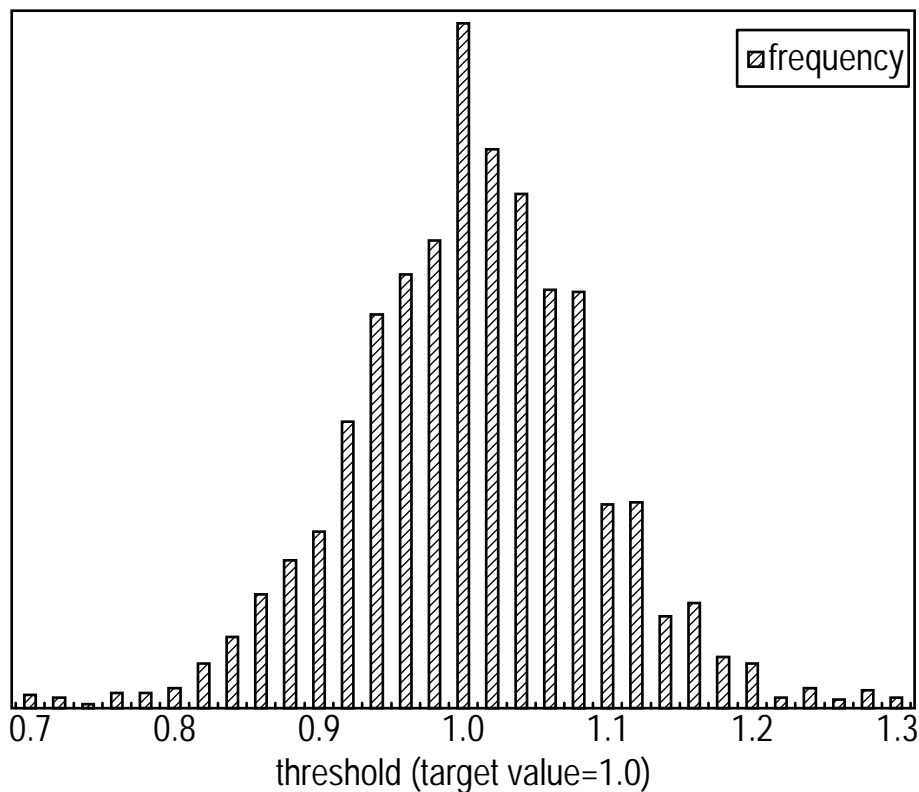

Figure 6   Distribution of the obtained threshold values with the best-PEST algorithm. The stochastic observer's threshold is 1.0 (target value). Basis for the distribution are 2500 threshold determinations, each representing the mean of 3 runs.

The distribution is approximately Gaussian with a mean of 0.99755, and a variance of 0.00764.

The aim of the second simulation was to gain insight in the convergence behaviour of best-PEST for different numbers of trials until termination, and for different slope values of both stochastic observer and best-PEST model. For that purpose we calculated the variance of the mean threshold after 1000 runs as a function of the mentioned variables. The contour lines of equal variance in the range [0, 0.05] can be seen in Figure 7.
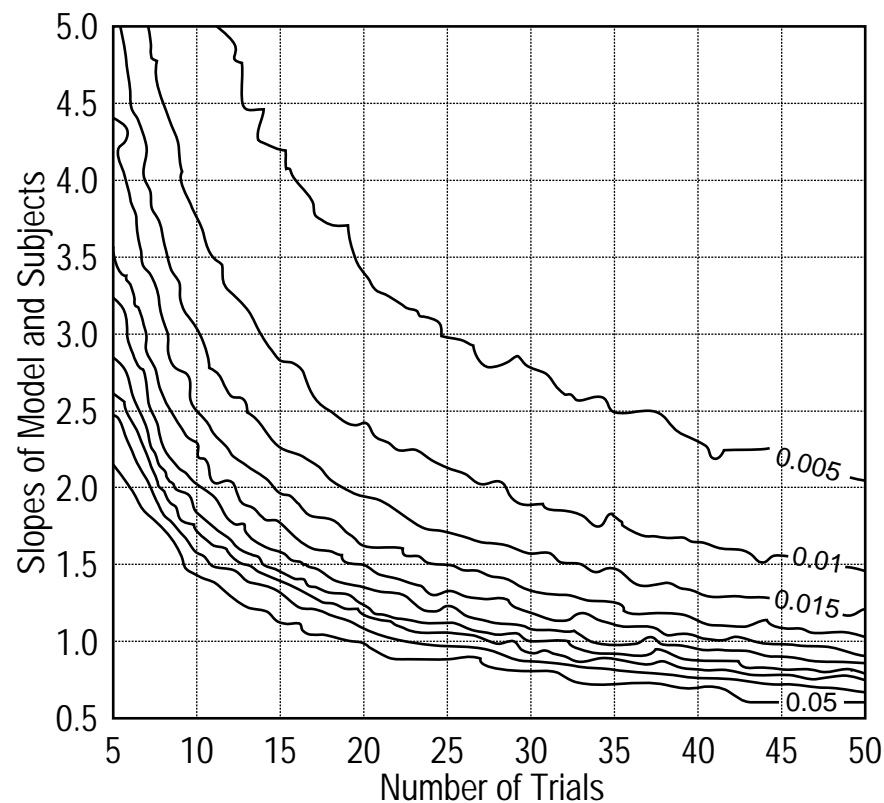
Figure 7 Simulation of threshold determination with the best-PEST algorithm. The curves show contour lines of threshold variances up to 0.05. The number of trials until a threshold determination stops is on the abscissa, the slopes of the psychometric functions of both stochastic observer and model are on the ordinate. The variance is calculated on the basis of 1000 threshold determinations for each measuring point. The slope's increment is 0.1, the number of trials' increment is 1.

The equal variances of the mean threshold describe approximately exponential curves, which is coherent with the interpretation that increasing number of trials diminish the marginal utility. This interpretation is obvious when we consider the nature of the best-PEST procedure: the information increase relative to the existing information is decreasing with every additional trial, and therefore changes of the estimated thresholds become smaller. A further prediction that can be made from these data is that the number of trials play an important role only for big slopes of the psychometric functions.

The third simulation was made in order to analyse the convergence behaviour of best-PEST for different, interdependent slope values of the observer's and of the model's psychometric function. For that purpose we calculated – as in the second simulation – the variance of the mean threshold after 1000 runs as a function of the two slope variables. The contour lines of equal variance in the range [0, 0.05] can be seen in Figure 8.
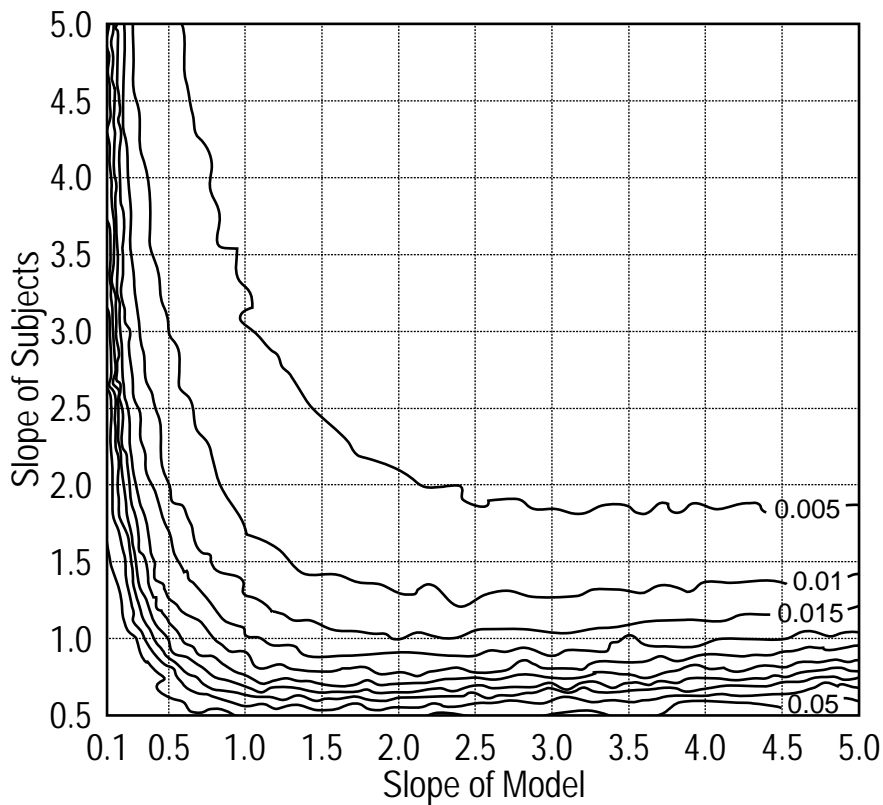
Figure 8   Simulation of threshold determination with the best-PEST algorithm. The curves show contour lines of threshold variances up to 0.05. The slope of the model is on the abscissa, the slope of the stochastic observer is on the ordinate. The variance is calculated on the basis of 1000 threshold determinations for each measuring point. The increment is 0.1 for both variables.

On first sight the curves of equal variance indicate no reasonable and explainable model of the interdependent behaviour of the two slope parameters. It can be read out that there is no reason to chose much bigger model than observer slopes, since they increase the variance for a given observer slope, especially in its lower range. As a rule of thumb we can say that a model slope twice as big as the observer slope will provide best results, since it seems, that there are relative minima of the contour lines at these points.

5   References

Gescheider, G. A. (1997), *Psychophysics: The Fundamentals*, Lawrence
Erlbaum Associates, Mahwah, NJ.

Pentland, A. (1980), *Maximum likelihood estimation: The best PEST*,
Perception & Psychophysics, 28 (4), 377-379.

Treutwein, B. (1995), *Adaptive Psychophysical Procedures*, Vision Research,
35 (17), 2503-2522.