■解説（VISION Vol.15, No.3, 189-195, 2003）

適応的心理物理学的測定法による閾値の推定

原澤賢充

東京大学大学院 人文社会系研究科 心理学研究室
〒113-0033 東京都文京区本郷 7-3-1

1．古典的心理物理学的測定法

知覚の心理物理学では，しばしば閾値を測定する場面に出くわす．その際に用いられる測定法のうちもっともなじみ深いものが Fechner (1860)[1] の考案したいわゆる古典的心理物理学的測定法と呼ばれる恒常法，極限法，調整法という３つの方法である．これらの方法は原理が単純であり実装しやすい反面，効率が悪いという欠点がある．そのため，Fechner以降，適応的測定法と呼ばれるより効率のよいさまざまな測定法が考案されてきた．ここでいう効率とは，ひとつの閾値の推定値がより少ない試行数で得られ，なおかつ得られた値のばらつきが小さいことを意味する[2]．

2．適応的測定法

適応的測定法と古典的測定法のおもな違いは，刺激強度の配置の仕方にある．適応的測定法では提示される刺激の強度がなるべく閾値に近くなるように配慮がなされている．これは，刺激強度が閾値から大きく離れていた場合，その試行で得られた反応は心理測定関数を決定するのにあまり意味を持たないからである．そのために，次の試行の刺激強度を決定するのにそれまでに得られた被験者の反応を参考にしているのである．

適応測定法は実験者が事前に持っている知識や知りたいと思っているものに応じていくつかのグループに分けることができる[3]．

１）心理測定関数は刺激強度に応じて単調に増加（もしくは減少）するが，その形はわからない．ある反応強度に対応する刺激強度が知りたい．
２）心理測定関数がいくつかの変数で記述できることがわかっている．心理測定関数の閾値と傾きを知りたい．
３）心理測定関数の形が完全にわかっていて，その位置だけがわからない．閾値だけが知りたい．

１)の場合は，ノンパラメトリックな方法が用いられ，２)と３)の場合にはパラメトリックな方法が用いられる．

3．ノンパラメトリックな測定法

測定法ごとに心理測定関数が満たしていなければならない条件は異なるが，ノンパラメトリックな測定法では心理測定関数が単調に増加（もしくは減少）することが保証できればよい．

3.1 階段法・上下法・変形上下法

階段法や上下法は極限法によく似た測定法だが，被験者の反応を刺激強度の決定に役立てている点が異なる．被験者の反応が(たとえば，「見えない」から「見える」へ，あるいは不正解から正解へ)変化すると，セッションがそこでうち切られたり，あるいは刺激強度の変化の方向が反転したりする．これらの方法では，最終的な推定値は被験者の反応の変化がある回数に達したときに，反応の変化が生じた刺激強度の平均値として求められる．

階段法や上下法では刺激強度の決定には直前の試行の反応のみが利用されていたが，変形上下法[4] ではそれまでの数試行の反応を利用している．たとえば，1-up / 2-down と呼ばれる形式

の場合，誤反応が1試行あると刺激強度が上がり，正反応が2試行連続すると刺激強度が下がる．刺激強度の変化の反転回数がある値に達したときに測定はうち切られ，そのときの刺激強度が最終的な推定値となる．条件を決定する数値に応じて，最終的に推定される強度の刺激によって生じる正反応の生起確率が異なる．たとえば，1-up/2-downでは正反応を0.707の確率で生じさせる刺激強度を求めることになる[4]．

3.2 PEST

PEST (Parameter Estimation by Sequential Testing)[5] は，PESTルールと呼ばれるヒューリスティックな規則を用いて刺激強度を決定する測定法である．

3.2.1 いつ刺激強度を変えるのか

ある確率 ϕ (標的確率) で正反応が生じるような刺激強度を閾値として求める場面を考える．ある刺激強度での正反応率が ϕ よりも大きいか小さいかを求めることによって次の試行の刺激強度を決定する．

ある強度での測定 (ステップという) が始まったら，その強度での正反応数 $N(C)$ を数えておく．そのステップでの試行数 n があらかじめ定められた値 m 以上になったら，ステップを継続するか, 刺激強度を変えて次のステップに移るかを判断する．もし現在の刺激強度が閾値 θ に等しいとすると, 正反応数の期待値は $E[N(C)] = n\phi$ となる． $N(C)$ がある範囲 $[n\phi - W, n\phi + W]$ に収まっていればこのステップを継続し，第 $n+1$ 試行を行う． $N(C)$ がより高い値なら強度を下げ，より低い値なら強度を上げて次のステップへ移行する． W は刺激強度の変化のしやすさを決定する変数なので， W が小さければ測定が早く終了し，大きければ測定値の信頼性が高まる．

3.2.2 つぎのステップの刺激強度

最初のステップの刺激強度は実験者が任意に決めることができる．第2ステップ以降の刺激強度は直前のステップの刺激強度にあるステップサイズを加減することによって決定される．このステップサイズは被験者の反応によって変化する．ステップサイズは次のように決定される (図1)．

1）刺激強度の移動の方向が反転するたびに半分に．
2）方向が変わらないときはそれ以前と同じ．
3）3回あるいはそれ以上連続して同じ方向に

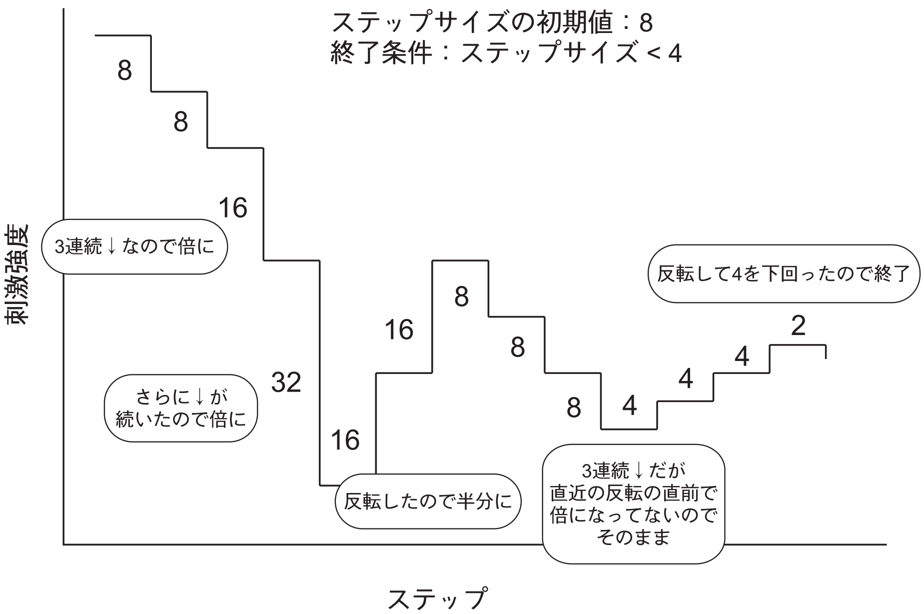

図1　PESTルールの適用例．横軸はステップの進行を，縦軸は各ステップでの刺激強度を示す．図中の数字はステップサイズを示す．刺激強度は $N(C) < n\phi - W$ で減少し， $N(C) > n\phi + W$ で増加する．

移動した場合はステップサイズを倍に．ステップサイズが大きくなりすぎないよう上限を決めておく．

4） 2回連続して同じ方向に移動した場合は，状況に応じてステップサイズを倍にしたりしなかったりする．直近の移動方向の反転の直前のステップサイズがその前の倍になっていたら，このときに倍にする．そうでなかったらそれまでのステップサイズを維持する．

3.2.3 終了の判断・最終推定

ステップサイズがあらかじめ決めておいた値よりも小さくなったら測定を終了する．そのときの刺激強度を閾値とする．

以上のように，PEST は実験者が事前に閾値についての予測をもっていなくても比較的柔軟に対応することができる．刺激強度の初期値が真の閾値から大きく外れていても，ほぼ安定して閾値に近い範囲で刺激を提示することが可能である．だが，階段法などの測定法と同様に，刺激強度の移動のしかたが被験者に予想されやすいという難点がある．その場合，複数のセッションを1試行ごとに交互に走らせるなどして予想のしやすさを軽減させることができる．PEST には RAT モード PEST [6] や More Virulent PEST [7] などいくつかの派生形が考案されている．

4．パラメトリックな測定法

パラメトリックな測定法では，心理測定関数の一般的な形について事前に決定しておかなければならない．つまり，累積正規分布，Weibull 関数，logistic 関数などといった関数のうちいずれかを心理測定関数のテンプレートとして選択し，1つか2つの自由変数すなわち閾値と傾きを測定することになる．

4.1 最尤法

パラメトリックな適応的測定法では，刺激強度や最終的な推定値の決定にしばしば最尤法が用いられる．

求めるべき心理測定関数が $\psi(x \mid \Theta)$ であったとする． x は刺激強度の物理軸，$\Theta = (\theta_1, \theta_2, \ldots, \theta_k)$ は心理測定関数の位置と形状を決める変数（傾きや閾値など）の集合とする．簡便のために，閾値 θ 以外の変数は既知であるとして， $\psi(x \mid \theta)$ を考える．このとき n 回の測定を様々な強度 $(x_1, \ldots, x_n) = \mathbf{x}$ で行ったとすると，それぞれについての被験者の反応 $(z_1, \ldots, z_n) = \mathbf{z}$ が得られる．すると，それぞれの刺激強度につき1度ずつ θ に関する推定すなわち確率密度関数 $L(\theta \mid x_i)$ を得ることができる．この関数は，θ が取りうる値それぞれがどれだけ閾値として確からしいかを示している．これを $\mathbf{x}$ 全体について考えると， θ についての同時確率密度関数が得られ，

$$L(\theta \mid \mathbf{x}) = \prod_{i=1}^{n} L(\theta \mid x_i)$$

と表すことができる．これは θ の尤度関数に等しく，この関数が最大の値を取るときの θ が実験の結果から得られた心理測定関数の変数となる．これを θ の最尤推定量という．

4.2 ベイズ推定

いっぽう，ベイズ推定も適応的測定法に頻繁に用いられる手法のひとつである．ベイズ推定では，事前推定とデータを組み合わせることによって事後推定を導く．適応的測定法における事前推定とは，実験者が事前にもっている心理測定関数の閾値や傾きに関する知識や，その試行までに得られた被験者の反応を意味する．これに新たなデータを加えることによって事後推定を導くのである．この場合の事後推定とは，最終的な閾値や傾き，あるいは次の試行の刺激強度を意味する．

4.3 QUEST

QUEST [8] はベイズ推定と最尤法を取り入れた適応的測定法である．ベイズ推定を用いて試行ごとの刺激強度を決定し，最尤法を用いて最終的な閾値を求めるのである．

4.3.1 理論

QUEST では，心理測定関数の形状はあらかじめわかっているものとして，その位置つまり閾値のみを推定する．他の適応的測定法と同様に QUEST においても刺激強度がなるべく真の閾値

の近くに配置されるよう考慮されている．そこで用いられているのがベイズ推定である．4.2 で述べたとおり，ベイズ推定では事前推定とデータから推定を行う．第１試行の刺激強度は，事前推定つまり実験者があらかじめもっている閾値に関する情報に基づいて決定される．すると第２試行では，事前推定とデータ（第１試行の被験者の反応）から事後推定が得られ，ここから刺激強度が決定される．そして第２試行の事後推定が第３試行の事前推定となり，これに第２試行の反応がデータとして加えられて，次の刺激強度が決定される．以降，この繰り返しで動的に刺激強度が配置されていくのである．

しかし，最終推定にはベイズ推定を用いない．実験者があらかじめ持っていた知識を用いることによってより効率的に刺激強度を決定することが可能になるが，これは言うなれば実験者の偏見であるので，最終的な閾値の推定に用いるのは好ましくない．そこで，最終推定ではこれを除いて，第１試行から最終試行までのデータのみを用いて閾値を求めるのである．

4.3.2 実践

QUEST による閾値推定の方法を具体的に見てみる．心理測定関数の形が $\psi(x)$ で，閾値が θ であったとすると，求める心理測定関数は

$$p_\theta(x) = \psi(x - \theta)$$

として表せる（図２a）．刺激強度の決定には QUEST 関数と呼ばれる対数確率密度関数 $Q(\theta)$ を用いる．この関数が最大の値を取る点が刺激の強度として採用される．実験の開始時点での QUEST 関数は

$$Q_0(\theta) = \log h(\theta)$$

として表される． $h(\theta)$ は事前推定を表し，どのような関数形をとっていてもかまわないが，閾値がありそうな箇所を頂点とする正規分布関数が用いられることが多い（図２c．いちばん上の曲線）．この正規分布関数の分散や高さを操作することによって，刺激強度の配置に対する実験者の事前知識の影響の強さを操作することができる．

第１試行の刺激強度は以上のように $Q_0(\theta)$ の最大値として決定されるが，第２試行の刺激強度はこれに第１試行の実験結果を加味することによって導かれる．第１試行が成功（「見えた」「聞こえた」あるいは強制選択における正解などの正反応）だった場合には成功関数 $S(x)$ を，失敗（「見えない」「聞こえない」あるいは不正解などの負反応）だった場合には失敗関数 $F(x)$ を加える．強度 x の刺激に対する成功の確率は $p_{S|\theta}(x) = p_\theta(x)$ として，失敗の確率は $p_{F|\theta}(x) = 1 - p_\theta(x)$ として記述することができ，これらはそれぞれ

$$p_\theta(x) = \psi(x - \theta)$$
$$1 - p_\theta(x) = 1 - \psi(x - \theta)$$

と書くことができる．ここから成功関数，失敗関数はそれぞれ

$$S(x) = \log \psi(-x)$$
$$F(x) = \log[1 - \psi(-x)]$$

と書ける（図２b）．右辺のカッコの中が $-x$ になっていることに注意したい．これは，心理測定関数を左右入れ替えたことを意味し，たとえばある刺激強度で成功したとすると閾値はそれより小さいであろうし，失敗したとするとより大きいであろうことを表現している．よって第２試行の刺激強度は

$$Q_1(\theta) = Q_0(\theta) + \begin{cases} S(\theta - x_1) & \text{成功のとき} \\ F(\theta - x_1) & \text{失敗のとき} \end{cases}$$

の最大値として導くことができる（x_n は第 n 試行の刺激強度）．つまり，図２b，c にあるように，成功／失敗関数をそのとき提示した刺激強度に応じて水平方向にずらしてものをそれまでの QUEST 関数に加えるのである．よって QUEST 関数を一般化すると，

$$Q_n(\theta) = \log h(\theta) + \sum_{i=1}^{n} R_i(\theta - x_i)$$

と表現できる．$R_n(x)$ は第 n 試行の反応によって $S(x)$ もしくは $F(x)$ の形を取る．

刺激強度はどの試行においてもその時点でのQUEST関数の最大値として導かれ，そのQUEST関数は実験開始時点での事前推定にその試行までの被験者の反応を積み重ねたものとして表されるのである（図2c）.

以上のようにして，刺激強度を決定し被験者の反応を集めることが可能になったので，最終的な閾値の推定について見てみる．セッションを終了する条件は2つの方法のいずれかから選択することができる.

ひとつは信頼区間の大きさによるものである．QUEST関数から実験開始時点での事前推定を取り除くと θ の対数尤度関数 $L(\theta)$ が得られ，

$$L_n(\theta) = Q_n(\theta) - Q_0(\theta)$$

と書ける．実験を進めていくとこの関数はだん

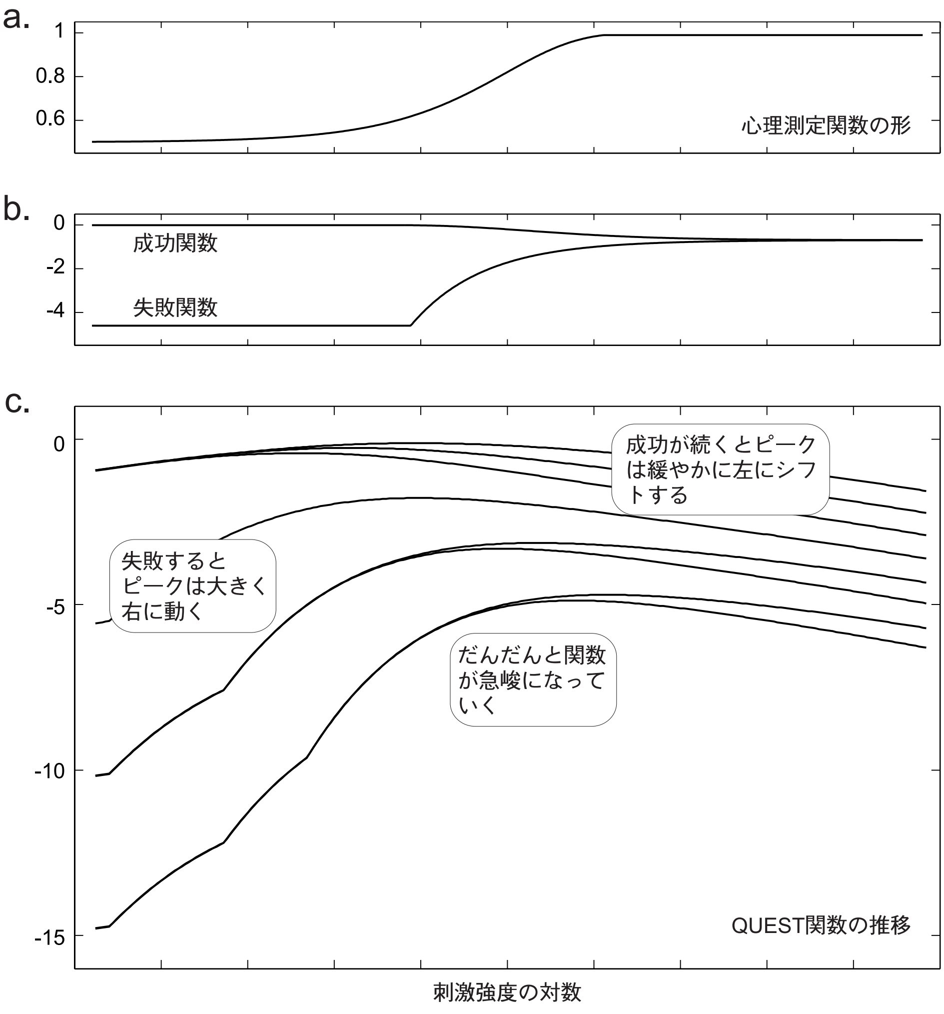

図2　QUEST関数の算出の仕方．a）二肢強制選択課題での心理測定関数の形を示してある．実験場面では，この関数がどの程度水平方向に移動するかを求めることになる．b）成功関数と失敗関数．a）の心理測定関数を左右入れ替えて対数にしたのが成功関数．1から心理測定関数を引いて左右入れ替えてから対数にしたのが失敗関数．二肢強制選択課題の場合，失敗関数のほうが成功関数よりも振幅が大きい．これは失敗したときのほうがより大きな情報量を持つことを意味する．実験では，これらの関数を提示した刺激強度に合わせて水平方向にずらしたものをQUEST関数に加えることになる．c)QUEST関数の推移．いちばん上の曲線は事前推定 $Q_0(\theta)$ を意味する．試行ごとに成功関数もしくは失敗関数を加えていく．成功関数が加えられた場合は，ピークはゆるやかに左にシフトする．失敗関数が加えられた場合は，ピークは大きく右にシフトする．これらを繰り返す内に，しだい急峻な曲線が描かれるようになっていく.

だんと $L_n(\theta)$ の最大値 $\theta_{\rm like}$ を中心として図２c と同様に急峻なピークを描くようになっていく．この過程で，θ 軸上のある値 θ_j がある信頼水準において $\theta_{\rm like}$ と等しいとされる信頼区間を求めることができる．この区間がある程度以上小さくなったときに測定を終了するのである．

もうひとつの方法はあらかじめ終了までの試行数を決めておく方法である．

いずれの場合においても，測定を終了した時点での $L_n(\theta)$ の最大値 $\theta_{\rm like}$ を最終的な閾値の推定値とする．最終推定や信頼区間には事前情報が関与していないという点が重要である．

4.3.3 注意など

ここまでの説明では，実験がyes/ no 課題であるのか強制選択課題であるのかについて言及していないが，いずれの場合においてもQUESTを用いることができる．ただし，yes/ no 課題の場合には，被験者の反応バイアス，特にfalse alarmの値を $\psi(\theta)$ に加味する必要がある．また，どちらの場合においても，被験者がキーの押し間違えなどをする確率も $\psi(\theta)$ に加味するとよい．

最終的に推定される閾値は必ずしも古典的な正反応率（二肢強制選択課題なら 0.75，yes/ no 課題なら 0.5）を示さない．もしこういった値が必要ならあらかじめ $\psi(\theta)$ を水平方向にいくらかずらしておけばよい．ただしその場合，効率はいくらか低下することになる．

よりよい効率のためには測定はなるべく閾値近傍で行われることが望ましいが，それによって被験者の心理的負荷が高まる場合がある．その場合，QUEST 関数から導かれる刺激強度にある程度の範囲の乱数を加える方法もある．効率は低下するが被験者の負担は低下する．

ここで扱う心理測定関数や尤度関数などはすべて刺激の物理強度に対する関数となっているが，この場合の強度は実際には対数強度である．たとえば，これは輝度コントラストの対数であったり，音圧レベルであったりする．

心理測定関数の傾きはあらかじめ決定しておくことが求められるが，もし値がわからなかった場合は真の値よりも緩やかな傾きを与えたほうがよい．この場合，閾値の収束に時間がかかるが，その分おかしな値に収束してしまう可能性を低くすることができる[3]．

MathWorks 社製ソフトウェアMatlab 上で使用できる Psychophysics Toolbox[9, 10] を使用すると，QUEST を実験プログラムに容易に実装することができる．他の環境で実験を行う場合も，ソースコードが参考になるだろう．

4.4 そのほかの測定法

QUEST では，最終推定や刺激強度の設定において尤度関数や QUEST 関数の最頻値（= 最大値）を用いていたが，ほかの推定量を用いる方法も考案されている．ZEST（Zippy Estimation by Sequential Testing）[11] では，最頻値の代わりに平均を用いる（よって，ZEST はしばしばmean-QUEST とも呼ばれる）．つまり，

$$\theta_{\rm mean} = \int \theta Q(\theta) d\theta \Big/ \int Q(\theta) d\theta$$

のようにして刺激強度を推定するのである．最終推定では，$Q(\theta)$ の代わりに $L(\theta)$ を用いればよい．平均値を推定量に用いることによって，中央値や最頻値を用いた場合よりも効率がよくなることがコンピュータシミュレーションから示されている[11]．ZEST も QUEST と同様に Psychophysics Toolbox を用いて実装することが可能である．

５．おわりに

Fechner[1] の古典的測定法以降さまざまな測定法が考案されてきた．その原動力のひとつは計算機性能の向上であろう．近年，心理物理学の実験にも計算機が多用されており，刺激の生成のみならず刺激強度の選定，閾値の推定にも頻繁に用いられている．たとえば，QUEST[8] が 1980 年前後に発表されているが，アルゴリズムの複雑さや計算量は当時の計算機の発達の程度と一致していたと考えられる．たとえばそれは彼らが確率密度関数の積ではなく対数の和を用いていたことや，成功関数や失敗関数をあらかじめ数表にしていたことなどからうかがい知ることができる．

近年はさらに計算機の性能が高くなっているた

め，測定法のアルゴリズムにより複雑な計算を含めることができるようになり，事実そういった方法が多く考案されるようになってきた[11-15]．今後もますます効率のよい測定法が考案されると思われるが，それを使う実験者が理解しておかねばならないことも同時に増えていくことを忘れてはならないだろう．

文　献

1）G. T. Fechner: Elemente der Psychophysik. 1860.（English translation: D. H. Howes and E. C. Boring（eds）and H. E. Adler（tansl.）, Holt（Rinehart & Winston）, New York, 1966.）

2）M. M. Taylor: On the efficiency of psychophysical measurement. *Journal of the Acoustic Society of America*, **49**, 505-508, 1971.

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

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

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

6）H. L. Kaplan: The five distractors experiment: Exploring the critical band with contaminated white noise. *Journal of the Acoustic Society of America*, **58**, 504-511, 1975.

7）J. M. Findlay: Estimates on probability functions: A more virulent PEST. *Perception and Psychophysics*, **23**, 181-185, 1978.

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

9）D. H. Brainard: The Psychophysics Toolbox. *Spatial Vision*, **10**, 433-436, 1997.

10）D. G. Pelli: The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, **10**, 437-442, 1997.

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

12）L. L. Kontsevich and C. W. Tyler: Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, **39**, 2729-37, 1999.

13）J. Miller and R. Ulrich: On the analysis of psychometric functions: the Spearman-Karber method. *Perception and Psychophysics*, **63**, 1399-1420, 2001.

14）F. A. Wichmann and N. J. Hill: The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception and Psychophysics*, **63**, 1293-1313, 2001.

15）F. A. Wichmann and N. J. Hill: The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception and Psychophysics*, **63**, 1314-1329, 2001.