# Procedural and statistical methods in the use of the two-flash threshold

BRUCE W. MAASER and FRANK H. FARLEY
*University of Wisconsin, Madison, Wisconsin 53706*

The major recent procedures in the assessment of the two-flash threshold are reviewed, including the method of limits, a modified method of limits, method of constant stimuli, and methods based on signal detection theory. Statistical procedures for these methods are critically analyzed. Recommendations are made for the choice of method.

A number of statistical methods are possible in the testing of sensory capabilities. Each is based on one of several hypotheses about perceptual mechanisms, and each implies a specific procedure for gathering data. One such test of sensory capabilities that can employ these various methods is the two-flash threshold (TFT), which measures the ability to distinguish two consecutive flashes of light. Typically, the range of individual differences in humans is from 10 to 150 msec of interflash interval (IFI).

One commonly employed statistical and procedural method in the use of the TFT is the method of limits (Herrick, 1967). This method assumes the "threshold" hypothesis: Between IFIs A and B, the longer the interval, the greater the probability of the subject perceiving two flashes. Procedurally, in the ascending method of limits, IFI is increased in increments from a very short interval until the subject reports two flashes. Similarly, in the descending method of limits, IFI is decreased in increments from a long interval until the subject reports one flash. Often, the obtained intervals from the ascending and descending methods are averaged, and this average is defined as the two-flash threshold.

A variation of the method of limits is the modified method of limits (Maaser & Farley, Note 1; Severson & Farley, Note 2). The subject is first presented with a long IFI, and if he reports two flashes, he is then presented with a short IFI. If he reports one flash, the procedure is repeated with decreasing range in 10-msec steps initially, until the point is reached at which two IFIs occur 10 msec apart for which the subject reports two flashes for the longer and one for the shorter. The procedure is then repeated with 2-msec steps until two steps are found that are 2 msec apart and for which the subject reports two flashes for the longer and one for the shorter. These pairs of flashes are then repeated twice, and if the subject maintains the same response to

them, the last IFI at which he reports one flash is taken as his threshold.

Another variation of TFT procedure and analysis is the method of constant stimuli (Herrick, 1967). Here, a certain number of IFIs are chosen ranging from very short to very long. An equal number of presentations for each interval are given to the subject in random order. Probabilities for each interval of distinguishing two flashes are calculated, and a threshold can be determined.

A procedure combining the method of limits and the method of constant stimuli is the random method of limits (Herrick, 1973). The data can be analyzed using either method of limits or method of constant stimuli statistical procedure. The random method of limits is based on the "phi-gamma" hypothesis, which states that the probability values obtained for each IFI are described by a cumulative normal curve.

The methods described thus far are subject to the major drawback that response mode or attitude of the subject can influence the thresholds of sensory tests. For example, schizophrenics are more likely to report two flashes when only one occurred, thus artificially lowering the two-flash threshold (Gruzelier & Venables, 1975). A threshold in the traditional psychophysical sense is affected by the criterion the subject adopts for his response and by his neural sensitivity. Threshold differences between groups may reflect either or both of these factors, as a study of neurotics has demonstrated (Boissonneault, Dorosh, & Tong, 1970). Study of the effect of arousal level and hormone cycle in women has also shown that the response criterion a person uses in distinguishing one flash from two is an important factor along with neural sensitivity (Tong & DeMarchi, 1972).

Signal detection theory is capable of separating and quantitatively assessing the effects of response criterion and neural sensitivity (Tong & DeMarchi, 1972). It assumes that the central effect of a stimulus varies because of physical and neural noise; consequently, the detection of a signal requires a central statistical decision procedure (Treisman & Watts, 1966). This decision policy is based on the relative worths of being either

0090-5054/80/030188-03$00.55/0

correct or incorrect (Green & Swets, 1966). Signal detection theory is based on the hypothesis that for any given IFI, in a particular presentation to a particular subject, only one of two responses can occur, rather than assigning a particular probability to each stimulus level.

In signal detection theory, the concept of threshold as a single summarizing measure is discredited and is instead replaced with the concepts of sensitivity and criterion (although a threshold measure can still be defined and measured). Sensitivity, or signal-to-noise ratio, symbolized by $d'$, is defined as $(M_{sn} - M_n)/\sigma n$, that is, the difference between the means of the signal-plus-noise distribution and the noise distribution divided by the standard deviation of the noise distribution. Criterion is the neural effect level or location on the neural effect axis, or, in a more general definition, the likelihood ratio axis that the subject chooses to divide one-flash responses from two-flash responses (Treisman & Watts, 1966).

While statistical and procedural methods specifically based on signal detection theory exist, the methods described earlier based on threshold theory can be modified to be compatible with signal detection theory (Boissonneault et al., 1970). The method of limits and the method of constant stimuli are adequate if trials containing only one flash are presented in sufficient numbers to allow the reliable determination of the proportion of "false-two" responses, so that an estimate of the observer's response criterion can be made.

Two main statistical and procedural methods exist based on signal detection theory. The first (Treisman & Watts, 1966) gives the subject 50% of the flash presentations with 0 IFI (i.e., only one flash), and the other 50% of the presentations vary from 10- to 100-msec IFI. The various IFI presentations including 0 IFI are presented in a random order. The probability p(T) of the subject reporting two flashes at each IFI is then calculated. Each p(T) is then expressed as a "probit score" (Finney, 1947). After eliminating all p(T)s of 0 or 1, straight lines representing the least-squares regressions of probits on IFIs are then calculated for each subject. Threshold is defined as the IFI value where the probit score equals 5 (Tong & DeMarchi, 1972). Sensitivity ($d'$) is defined as $[1/(I_2 - I_1)]$ (Treisman & Watts, 1966); The response criterion is defined as $[I_1/(I_2 - I_1)]$, where $I_1$ is the IFI threshold and $I_2$ has the probit score of 6 or the p(T) = .84. The criterion may vary in the context of the immediately preceding stimuli. Therefore, the criterion as here defined is a value averaged over the total set of presentations given to the subject.

The probit transformation is a modification of the normal curve, or another way of converting percentage values to values on a curve. The relation is: probit value = $5 + (x - \mu)/\sigma$, where x is the log value of the parameter, $\mu$ is the mean of the distribution, and $\sigma$ is its standard deviation (Finney, 1947). The probit transformation is easier to use arithmetically than the normal because it has no negative values.

The second statistical and procedural method based on signal detection theory (Green & Swets, 1966) uses one of three types of tasks. The first is the "one-two" task: The subject on each trial responds as to whether he detected one or two flashes. Green and Swets (1966) refer to this as the "yes-no" task, a name more applicable to absolute sensory detection procedures. The second is the rating task, in which the subject rates on a scale (such as 0-10) whether he saw one or two flashes. The third is the forced-choice task, in which, for example, two flashes may be presented in one of two or more temporal intervals, while one flash is presented in the other interval(s). The subject is instructed to select the one interval he/she believes contained, in this example, the two-flash presentation (Green & Swets, 1966).

In the one-two task, a conditional probability table is set up, with entries for the case of two flashes presented/subject responds "two," and for the other three possibilities. A receiver operating characteristic (ROC) curve is then plotted, using two flashes presented/subject responds "two" against one flash presented/subject responds "two." By changing the reward for a correct "two" response without changing IFI, a curve can be generated. The slope of this curve at any point is the likelihood-ratio criterion that generates that point, where the likelihood ratio is the probability that a correct two-flash response occurred to the probability that an incorrect two-flash response occurred. An ROC curve based on likelihood-ratio criterion must have a correct two-flash response probability that is a monotonically increasing function of the incorrect two-flash response probability and a slope that is monotonically decreasing. The greater the area under the curve for a given likelihood ratio, the greater the sensitivity ($d'$) (Green & Swets, 1966).

The rating task has the advantage of not requiring numerous trials to determine each point on the ROC curve. Each point on the curve can be determined by using a given category of response by the subject (such as each digit on a 0-10 scale) and all those categories of response less likely than it (Green & Swets, 1966).

To avoid the problem of determining a criterion for decision, a forced-choice task is used in which the subject is presented a number (usually two or four) presentations, only one of which contains the two-flash presentation. For the present analyses to apply, the presentations would occur in consecutive time intervals. Symmetry may be assumed in that an equal number of signals will occur in all intervals, and the subject's bias on which intervals are most chosen can also be checked. The percentage correct in a two-alternative forced-choice task is equal to the area under the one-two or rating ROC curve, and a relation exists for comparing area and percentage correct in a more-than-two-alternative forced-choice task (Green & Swets, 1966).

Calculations for sensitivity (d') and criterion can be made with the help of tables for d' in Swets (1964). For example, in the one-two task, sensitivity is determined by taking the percentage of correct two-flash responses as the y-coordinate on the ROC curve. The percentage of incorrect "two" responses is taken as the x-coordinate on the same graph. The sensitivity scores associated with these two coordinates are obtained from the table for d'. Criterion scores are determined by estimating the slope of the ROC curve at the associated coordinates. The next pair of coordinates to the right of the obtained coordinates is taken from the d' tables. The actual criterion scores are obtained from the equation: slope = $(Y_2 - Y_1)/(X_2 - X_1)$, where $X_1$ and $Y_1$ are the experimentally obtained coordinates and $X_2$ and $Y_2$ are from the d' tables. Criterion scores obtained in this way, although more accurate than visually determined slopes, are slightly lower than the actual slope (Boissonneault et al., 1970). The concept of threshold is not defined in the Green and Swets (1966) method.

## RECOMMENDATIONS

For TFT experimentation, the methods based on signal detection theory seem more desirable than methods based on threshold theory, for reasons previously described. The one-two, rating, and forced-choice tasks could probably all be analyzed by Green and Swets (1966) or Treisman and Watts (1966) statistical procedures. The rating task seems more complicated to operate from the subject's point of view than the other methods. The forced-choice task, if conducted using more than one consecutive temporal interval, seems to have no advantage over the one-two task, and it is possibly more complicated to explain to the subject. The one-two task therefore seems to be the best alternative for experimentation. It would involve an equal number of one-flash and two-flash presentations. The two-flash presentations would comprise a number of intervals from 15 to 150 msec (for example), with an equal number of presentations in each interval. All presentations would be arranged in a random order.

## REFERENCES

Boissonneault, D. R., Dorosh, M. E., & Tong, J. E. The effect of induced heart rate change and neuroticism on the resolution of temporally paired flashes. *Psychonomic Science,* 1970, **20**, 98-99.

Finney, D. J. *Probit analysis.* Cambridge, England: Cambridge University Press, 1947.

Green, D. M., & Swets, J. A. *Signal detection theory and psychophysics.* New York: Wiley, 1966.

Gruzelier, J. H., & Venables, P. H. Two-flash threshold, sensitivity and criterion in normal subjects and schizophrenics. *Journal of Psychiatric Research,* 1975, **12**, 73-85.

Herrick, R. M. Psychophysical methodology: Comparison of thresholds of the method of limits and of the method of constant stimuli. *Perceptual and Motor Skills,* 1967, **24**, 915-922.

Herrick, R. M. Psychophysical methodology: VI. Random method of limits. *Perception & Psychophysics,* 1973, **13**, 548-554.

Swets, J. A. (Ed.). *Signal detection and recognition by human observers: Contemporary readings.* New York: Wiley, 1964.

Tong, J. E., & DeMarchi, G. W. Menstrual, diurnal and activation effects on the resolution of temporally paired flashes. *Psychophysiology,* 1972, **9**, 362-367.

Treisman, M., & Watts, T. R. Relation between signal detectability theory and the traditional procedures for measuring sensory threshold. *Psychological Bulletin,* 1966, **66**, 438.