1993

A new reliability parameter for automated perimetry: inconsistent responses

Michael Sukmin Lee
Yale University

Recommended Citation
Lee, Michael Sukmin, "A new reliability parameter for automated perimetry: inconsistent responses" (1993). *Yale Medicine Thesis Digital Library*. 2843.
http://elischolar.library.yale.edu/ymtdl/2843

A NEW RELIABILITY PARAMETER FOR AUTOMATED PERIMETRY: INCONSISTENT RESPONSES

MICHAEL SUKMIN LEE

Yale University

1993

A NEW RELIABILITY PARAMETER FOR AUTOMATED PERIMETRY: INCONSISTENT RESPONSES

A Thesis Submitted to the Yale University School of Medicine
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Medicine

Advisor: Dr. Joseph Caprioli
Yale University Department of Ophthalmology and Visual Science

By

Michael Sukmin Lee

1993

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to pass on my sincere appreciation for all those individuals who volunteered for this study. Despite the extreme effort and time required for this research project, they all persevered. I am grateful to Doug Hoffman for providing me with his computer expertise. I would like to thank Dr. Caprioli for his constant guidance and encouragement. To Mario Zulauf, thank you for always being there to help me with my work, being patient when I was impatient and frustrated, and for being a friend. Finally, I would like to thank my family. Without their love, support and confidence in me, this thesis and my medical degree would not have been possible.

ABSTRACT

A proper interpretation of perimetric results first requires an adequate evaluation of patient reliability. The patient reliability parameters that are currently used require extra testing time and are based on a relatively small sample of the subject's responses. To develop a new, more robust reliability parameter, 160 visual fields of 20 eyes of 20 healthy volunteers were performed with a customized static threshold program of 44 test locations. The responses to the bracketing strategy algorithm were analyzed for inconsistencies which occurred when the subject claimed to see and not see the same intensity stimulus when re-presented at the same test location. These inconsistencies were summed over the entire visual field and termed "number of inconsistent responses". The number of inconsistent responses correlated significantly with the following parameters: the sum of false-positive and false-negative responses to catch trials ($r = 0.62$; $p = 0.003$), fixation losses ($r = 0.51$; $p = 0.020$), and total number of stimuli ($r = 0.61$; $p = 0.004$). In contrast to conventional reliability parameters, the number of inconsistent responses is based on all the subject's responses and requires no additional testing time.

INTRODUCTION

BACKGROUND

Visual field testing has undergone many extensive changes and developments. In the 5th century B.C., Hippocrates first reported the existence of hemianopsia in one of his patients and in 150 B.C. Ptolemy made the first recorded attempt to measure a visual field. Mariotte made the next major discovery in 1688, describing the existence of a scotoma, which is now known as the physiological blind spot. Thomas Young in 1801 made the first quantitative measurement of a visual field.

Visual field examinations and techniques continued to evolve and their value have become increasingly more apparent. In 1856 Van Graefe [1] introduced in a systematic way the campimeter into clinical practice. Through his examinations he identified amblyopias, central scotomas, and bitemporal and binasal hemianopias. In addition he was able to distinguish absolute from relative scotomas. In 1862 with the aid of Albert, Förster created the first arc perimeter that allowed the mapping of visual field abnormalities.[2] This allowed closer examinations of the peripheral visual field and led to the mapping of

major glaucomatous visual field defects. It was not until 1889, when Bjerrum described the arcuate scotoma using a tangent screen that the analysis of the central visual field became popular.[3] He understood the necessity of examining both the peripheral and central fields in glaucoma. Rönne developed the kinetic isopter perimeter in 1909 and described the nasal step in glaucoma.[4] Realizing its importance, Traquair and others became strong proponents of painstaking performance of kinetic perimetry on the tangent screen.[5, 6, 7]

Goldmann understood the need for a perimeter that would provide optimal conditions for visual field testing, and in 1945 he introduced his own projection perimeter.[8] Its spherical shape allowed easier, more accurate, and more complete testing of the visual field. The background luminance and the size, color, and brightness of the stimulus could now be precisely controlled and standardized. It also offered absolute freedom of test object movement. In addition he allowed the examiner to directly monitor fixation, so that he could better assess the cooperativity and reliability of the patient. Finally, it was flexible enough to allow perimetrists to use both static and kinetic techniques effectively. It was for these reasons that the Goldmann perimeter became the clinical standard throughout the world.

Louis Sloan, along with Ferree and Rand, were the first to recognize the true value of nonmoving, static threshold perimetry.[9] However, the

popularization of this testing method, and the appreciation of its clinical value was the result of Harms and Aulhorn's work.[10] They introduced the Tubinger perimeter, which was specifically designed for static perimetry. Although it was possible to achieve excellent measurements, the time and effort involved made it both impractical and unpopular.

STATIC VERSUS KINETIC PERIMETRY

Nevertheless, it was recognized that static perimetry has certain advantages over kinetic perimetry. Kinetic perimetry uses a stimulus of constant size and intensity that moves from an area of non-vision to an area of vision. It provides a good general indication of the shape of the visual field with isopters in the center and periphery of the visual field. However, accurate detection of the boundary between seeing and non-seeing requires a sloping area in the hill of vision. In areas where the hill of vision is flat the distance between isopters increases. Therefore, kinetic techniques are not very sensitive to changes in relatively flat portions of the visual field.[11] This testing method is susceptible to missing early localized depressions. On the other hand, static perimetry uses fixed stimuli that vary in intensity, until the threshold is determined. Therefore, it does not really depend upon the slope of the hill of vision, and is probably better suited to the flat portions of the hill of vision. With

static perimetry, small relative or absolute scotomas can be found that may be missed by kinetic techniques.[12-14] Static perimetry also delineates central scotomas well because it checks the macular area. In addition quantification of the threshold allows more reliable assessment of the progression of scotomas. Although kinetic perimetry is easier and faster to perform than static perimetry, it has the disadvantage of having much more complex algorithms, and therefore, it is much more difficult to program and computerize.

AUTOMATION OF PERIMETRY

With the development and advancement of computer technology, automated perimetry has developed rapidly over the past several decades. Dubois-Poulsen and Magis, in the early 1960's, recognized the value of automation and are credited with the first significant attempt at automating perimetry.[15] However, their efforts were limited by technology. The electomotors to shift the test stimulus location and change to filters to alter stimulus intensity were under the control of the patient. Realizing the importance of complete automation, Lynn and Tate followed up the work of Dubois-Poulsen and Magis by suggesting the possibility of using a microcomputer for automatic visual field examinations.[16] This idea was put into practical use by the work carried out by several different investigators, most

notably that of Fankhauser, Heijl, and Krakau.[17-20]

Because automation has made static perimetry much more efficient and because it has several advantages over kinetic perimetry, almost all of the automatic, computerized perimeters utilize static perimetry as a basis for quantitation and screening. Although there are a number of computerized perimeters that employ kinetic techniques, the reliability and reproducibility of these measurements need to be studied further.

Automated perimetry has proven to be invaluable for clinical research and for the care of glaucomatous patients. With the technology of computerized and automated perimetry, one could achieve better standardization of testing technique, more uniform control of testing variables, and reduction of technician bias.[17, 21] The visual field test can be performed much more rapidly and easily. Standardized quantitative results lend themselves well to comparative statistical techniques. Finally, the information could be easily retained in computer memory, making possible the display of data in a variety of forms.[21, 22]

DIFFERENTIAL LIGHT SENSITIVITY ESTIMATION

Static perimetry is a visual field testing method, which determines the retinal sensitivity by the detection of various intensities of a stimulus with a constant background luminance, stimulus size, and exposure duration. The sensitivity is known as the **differential light sensitivity**. A visual field is completed, when the differential light sensitivity has been determined for a number of test locations. The sensitivity is the inverse of the threshold, which is defined by the psychometric function of **frequency-of-seeing curve** of the subject.[23] The threshold is the light intensity, whereby the subject is able to see a stimulus of constant size and exposure duration, 50% of the time (Figure 1). Since scatter is a property of the threshold itself, thresholds clearly are not absolute values.[24] It should be noted that the frequency-of-seeing curve assumes a symmetry of barely seeing and readily seeing sides of the curve, such as in a Gaussian distribution, but this has recently been called into question.[25-27]

Various **bracketing** or **staircase methods** have been described to determine a threshold. Besky developed a method which was used to measure aural thresholds of audio frequencies.[28] Cornsweet adapted this method to determine threshold values for light sensitivity.[29] Although this method was

never used for perimetry, it is still employed for various psychophysical tests, such as the Nicoles CS-2000 contrast sensitivity.

Later, Spahr further developed a repetitive staircase method, based upon computer simulations and theoretical considerations.[30] His method worked by using bracketing steps of 4, 2, and 1 decibel. The principle behind this method is that a threshold can lie anywhere on the scale, so several stimulus presentations are need to provide a rough estimation, and then the bracketing steps are made smaller to make a more accurate determination.[23, 31] Initially, at a single test location, large steps are employed in either increasing or decreasing luminosity, until the "seen-not seen" or "not seen-seen" boundary is crossed. At this point, the bracketing step is halved and the stimulus luminosity changes in the opposite direction of the first bracketing steps, until the threshold is crossed once again. The first phase makes a rough search for the threshold and the subsequent phases increases the accuracy and reliability of the measurement (Figure 2). The bracketing steps can be made smaller and the threshold could be crossed more times for a more accurate measurement. However, the cost of increasing the accuracy and reliability of results is an increasing investment in time and effort, which is certainly a consideration for patient comfort in clinical settings.

The Octopus perimeter employs several methods that use this bracketing

process to different extents, which vary the examination time and the quality of information obtained. The most time consuming method is the **normal** or **full bracketing method**. For most perimeters, the threshold is crossed twice, independent of whether or not the test location is pathological or normal. Therefore, each location is measured without bias and produces the most accurate and reliable visual field measurements.

On the opposite extreme, there is a qualitative **screening method** of examining the visual field. It does not actually measure the threshold. Instead, it presents to the subject a suprathreshold stimuli of 4 dB greater luminance than the age-corrected normal values, and describes the test locations as either normal, relative defect, or absolute defect.[32] This method is extremely fast, and is good when then number of defects is small and when individual test locations vary dramatically from their neighbors.[33, 34] However, the cost for speed is much less information.

There is a **fast method** of visual field testing that incorporates aspects of the previous two methods mentioned above. It possesses both qualitative and quantitative features. Therefore, it is faster, but has less information than the full quantitation. On the other hand, it is slower, but more accurate and reliable than the screening method. It accomplishes this by only measuring the threshold of test locations that have been determined to be pathological. This

method first assesses whether or not a test location is normal or pathological by screening with a stimulus that is 4 dB brighter than the calculated age-corrected normal values. If the stimulus is not perceived by the subject, the perimeter performs a full bracketing procedure on that test location. The drawbacks of this method are that the perimeter will not perform a full bracketing procedure, unless the test location demonstrates a considerable defect. Hence, it may miss shallow scotomas. In addition it may underestimate a patient's actual normal threshold because screening is done with age-corrected normal values, which represents an average.[35]

PRESENTATION OF RESULTS

There are currently several ways to depict the results of visual field examinations. One such representation is the gray-scale, which has the advantage of simplifying the search for characteristic changes. However, gray-scales often assign values to test locations that have not been measured, do not make full use of the large amount of quantitative data collected, and fail to provide any information about the range of normal variation.[36] However, the graphical image makes it easier to understand and is good to show to patients.

Visual field examination results can be represented as a three

dimensional map. Although it is a good graphical image, the three dimensional depiction tends to mask certain areas, making it impossible to determine the depth of some depressions.[37, 38] In addition it may also unduly exaggerate or diminish visual field alterations.[39] Like gray scales, it has the advantage of being easier to explain visual field results to patients.

The depiction of visual fields that is currently the most popular is a numerical representation. This gives the ophthalmologist a clear picture of the fields, without the interference or distortions of the graphical representations mentioned earlier. However, the amount of numerical information can be overwhelming, making the visual field difficult to interpret.

Recently, "Bebie" or "rank" curves are becoming a much more popular representation of visual fields. This method first classifies the sensitivities of test locations from highest to lowest. It then plots this against the sensitivities themselves, producing a curve. This curve is then superimposed onto a normal age-corrected curve, so that one can determine whether the visual field is pathological and the type of damage in the field, either localized or diffuse. This has the advantage over the other representations of not distorting the data, but still providing a good graphical image that is easy to interpret.[40] The disadvantage is that spatial representation is lost, which is important in many clinical situations.

RELIABILITY

Now that ophthalmologists have a lot of information from visual fields at their disposal, they are left with the extremely complex problem of interpreting these perimetric results. An area of major difficulty is evaluating a subject's cooperativity. Since perimetry is a subjective psychophysical measurement of retinal sensitivity, it depends upon the reliability of the subject's responses. It has become an especially difficult predicament because the realities of clinical practice have forced ophthalmologists to relinquish the role of perimetric testing to technicians. If the ophthalmologist were performing the examination, they would be in an excellent position to evaluate the patient's reliability and attentiveness. However, this is not the case, and even if the technician is extremely skillful, an ophthalmologist is forced to rely upon second hand information. In addition, now that perimeters have become automated, the technician has become a more passive operator in visual field testing. Hence, it has become more difficult for the technician to comment on the patient's reliability. Therefore, there is a need for more objective measurements of a patient's cooperativity and attentiveness, so that the ophthalmologist can make an accurate assessment of the visual field.

Currently, the ophthalmologist has several parameters at his disposal to

evaluate his patient's reliability. Other than the technician's own evaluation, the most basic piece of information about the patient's cooperation and reliability is the **number of stimulus presentations** that are necessary for the patient to complete the visual field examination. The number of questions asked of the patient increases mainly due to the difficulty in accurately determining the specific thresholds. Therefore, an increased number of stimulus presentations tends to indicate an unreliable patient.[41]

Automated perimeters have the ability to actively monitor **patient fixation**. Some of them, such as the Octopus, have a system that is capable of detecting, through an IR camera, obvious losses in patient fixation and recognizing when the patient has their eye closed, as in a blink, or when the pupil is in a "wrong" place. If the perimeter registers a fixation loss during a stimulus presentation, it disregards the patient's response, whether it is positive or negative. Later in the visual field examination, it will randomly present the same stimulus again. The number of wandering fixations and eye closures is reported as number of repetitions.

Some other perimeters, such as the Humphrey and Digilab, do not monitor fixation as repetitions, but rather they employ the **Heijl-Krakau** method of monitoring fixation loss.[42] They are also equipped with a monitor and the operator should supervise the patient during visual field testing. However, with

this method, fixation is quantified using the blind spot. These machines initially determine the location of the physiologic blind spot. Later during the visual field examination, it presents a stimulus in the area of the blind spot with the lowest retinal sensitivity. If the fixation of the patient wanders, they will shift their blind spot and be able to perceive a stimulus that they should not be able to perceive. This information is reported as a ratio of the number of positive responses to the number of stimulus presentations to the blind spot, and is reported as fixation losses.

Fixation loss, whether it is reported as the number of repetitions or as a ratio of positive responses to blind spot presentations, gives the ophthalmologist information about possible problems the patient may be having with fixation or attentiveness. They could be experiencing difficulty with fatigue, increases in rhythmic blinking, or sensory deprivation phenomenon.

There has been some controversy over the fixation loss criteria that should be used to judge a patient's performance. Currently, with the Humphrey Visual Field Analyzer (30-2 program), fixation loss must be under 20% and the false-positive and false-negative catch trials must be less than 33% for a field to be considered reliable. Katz and Sommer have examined the reliability indices of automated perimetric tests and found that failure to meet fixation loss criteria was mainly responsible for considering visual field examination results

unreliable.[43] Other investigators have also come up with the same conclusion, and went on to suggest that the fixation loss criteria is much too stringent.[44, 45] They believe that increasing the cutoff to 33% would markedly increase the number of reliable fields with minimal effect on the sensitivity and specificity of the test. Whatever criteria is used, fixation monitoring certainly aids the ophthalmologist in determining the reliability and accuracy of the results.

False-positive answers can also be used as an indicator for assessing patient performance. Some perimeters create a sound before stimuli presentations and others do not. The Octopus generates a sound by moving filters and mirrors. The pitch and duration of the sound varies, depending upon the amount of movement necessary for the next stimulus presentation. In order to simulate a normal stimulus presentation, some perimeters will use an audible sound and others will use an artificial time period before a false-positive catch trial presentation. When no stimulus or an imperceptible stimulus is presented and the patient gives a response, it is counted as a false-positive answer. A large number of false-positive responses indicates that the patient is very anxious and over-eager. The term "trigger happy" has been coined for such patients.

Another method for evaluating the patient's cooperativity is to look at the **false-negative answers**. The perimeter first established a threshold value for

a test location. It has determined that the patient can detect a stimulus at that particular point. Later, during the visual field examination, the perimeter returns to that test location and presents a stimulus significantly brighter than the threshold. If the patient fails to respond to the stimulus, it tends to indicate the patient is not paying attention. In other words the false-negative answers are an indication of the attentiveness of the patient.

For both false-positive and false-negative catch trials, the perimeters spend approximately 5% of all questions asked for each of these catch trials. If either catch trial is greater than 20%, then the accuracy of the results and the patient's reliability should be questioned.

Fluctuation has been known to be present in manual perimetry.[46] In automated perimetry, ophthalmologists use the fluctuation rate or **short-term fluctuation** to provide some insight into the cooperativity of the patient. Perimeters make repeated measurements for the threshold at several test locations, typically ten within a visual field test. Although it is possible, perimeters generally do not make multiple determinations for every test location because it would greatly increase the amount of testing time. It has been shown that short-term fluctuation calculated on the basis of ten doubly tested points lies within 44% of the true short-term fluctuation at a 95% confidence level.[23] Most perimeters calculate short-term fluctuation or the root mean

square (RMS) fluctuation rate as the square root of the mean of the variances of all test locations that have been multiply tested. The Humphrey perimeter calculates a weighted short-term fluctuation, which gives more emphasis to double determinations performed centrally.[47]

As discussed earlier, the threshold is not an absolute value. The threshold of a test location is defined as the intensity of a stimulus that the subject will perceive 50% of the time. In other words threshold assessment is an inherently variable process, and all strategies of threshold determination only provide an approximation of the theoretical "true" threshold. Therefore, a small amount of fluctuation is expected, and many studies have defined the normal fluctuation averages between 1.1 to 1.9 dB within the central thirty degrees.[48-53] Manufacturer's upper limits of normal for short-term fluctuation is 2.0 dB for the Octopus and 2.3 dB for the Humphrey perimeter.

THE PROBLEM

Although the commercial perimeters use the number of stimuli presented, fixation losses, short-term fluctuation, false-positive and false-negative catch trials, these indices have significant shortcomings that impair their ability to quantify reliability. Each method with the exception of total number of stimuli,

make significant contributions to the total testing time. The amount of test time spent for these reliability indices vary according to the type of perimeter. With the Digilab 1500 perimeter, false-positive and false-negative catch trials can take as much as 10% of total testing time. Fixation loss determination with Heijl-Krakau method can add another 10% to the testing time. The Humphrey perimeters use less time to assess reliability, with 5% of all presentations are used for fixation losses and 3% are used for false-positive and false-negative catch trials. Fatigue from prolonged testing can have a significant effect on the outcome of visual field measurements.[54-59] Therefore, it would seem advantageous to reduce the testing time and perhaps enhance the reliability of the results.

In addition to adding to the testing time, each of these indices have been shown to be poor estimators of reliability. The number of stimuli presented increases with the difficulty in determining the threshold. Although this could be secondary to unreliability, it has also been demonstrated to occur with abnormal visual fields.[60] In addition there has also been a learning effect demonstrated with the number of stimuli presented.[61]

False-positive and false-negative catch trials have many shortcomings. Firstly, they are determined by only a small sample of all the responses given during the examination, and hence, may not be robust estimators of reliability.

As with the number of stimuli presented, both false-positive and false-negative catch trials tend to decrease with experience and learning.[44, 62, 63] The number of false-positive replies has been found to be positively correlated with fixation losses.[44, 63, 64] False-positive catch trials are also associated with an increase in the differential light sensitivity and an increase in short-term fluctuation.[48, 64, 65] False-negative catch trials increases with a decrease in differential light sensitivity, abnormal visual fields, and an increase in short-term fluctuation.[43, 45, 49, 64, 65] Therefore, an increase in the rate of false-negative answers may be an indicator of abnormality and not necessarily of patient reliability.

Fixation loss determination can also be very problematic. In the Heijl-Krakau method, the blind spot may not be found, making it impossible to assess fixation loss. Anything that would cause the patient not to respond to a blind spot check, such as an eyelid closure, would be recorded as inadequate fixation. Studies have shown that because this method relies on accurate mapping of the blind spot before the test, a head tilt of as little as five degrees may affect fixation loss.[66] As with the other reliability parameters the number of fixation losses decreases with experience of the patient, but is independent of age.[44] Poor fixation also correlates positively with fluctuation, which is a poor reliability indicator.[67] Finally, for monitoring used by the Octopus perimeter, some literature suggests that this method may not be sufficient for recording fixation losses.[68]

Although it is used as a measure of reliability, short-term fluctuation is generally not considered a good reflection of patient cooperativity for several reasons. Short-term fluctuation increases with glaucoma and with defect depth.[12, 48, 49, 69, 70] It reaches a maximum of approximately 5 dB at a sensitivity of 10 dB and decreases as the sensitivity approaches 0 dB.[48] Presumably this is due to the fact that the possible range of sensitivity decreases with such minimal sensitivities.[71] Short-term fluctuation is larger for glaucomatous eyes than for normal eyes and glaucoma suspects fall in between these two groups.[49] Short-term fluctuation is particularly high around the border of scotomas, which has been attributed to slight eye movements.[72, 73] In addition to abnormal fields short-term fluctuation also increases toward the periphery.[48, 50, 51, 74, 75] However, there has been some controversy over the importance of eccentricity, with some studies reporting significant increases of short-term fluctuation and others reporting only minimal changes.[47, 74, 76] There is also controversy surrounding the effect of age. Some claim that age causes a significant increase in short-term fluctuation,[77, 78] but others have noted no such effect.[45, 48, 51, 54] Another factor that must be considered is a learning effect, whereby short-term fluctuation decreases with patient experience.[61, 62, 79, 80]

THE GOAL

Numerous studies have clearly demonstrated that all the methods currently used to assess reliability, the number of stimuli presented, fixation losses, short-term fluctuation, false-positive and false-negative catch trials, have significant shortcomings. They are time consuming and may contribute to fatigue during visual field examination. In addition they are influenced by many factors that prevent them from being true measures of patient reliability. Therefore, the purpose of this study was to determine whether a new parameter, inconsistent responses, could be used as an alternative reliability index. This hypothesis was tested by correlating the results of all responses during visual field testing with the standard reliability parameters.

MATERIALS AND METHODS

Normal volunteers who have never performed visual field tests were entered into the study. Each subject underwent a eye examination which consisted of visual acuity measurement, intraocular pressure measurement, slit lamp examination, and indirect ophthalmoscopy. The subjects met the following inclusion criteria to qualify for the study: best corrected visual acuity of at least 20/20, intraocular pressure less than 21 mmHg, refractive error less than 5 diopters spherical equivalent, normal slit lamp examination and no abnormalities of the optic disc or retina. If both eyes qualified, one was randomly chosen for study.

A Digilab 1500 perimeter (software version 1.9) was used to perform a customized visual field of 44 test locations in a central thirty degree field (Figure 3). It uses a 2mm stimulus of various durations with random time intervals, and has a background illumination that is maintained at 4 apostilbs by photoelectric feedback. Stimulus thresholds are determined by first calculating the precise thresholds at four primary points in the central 10 degree circle. Based on these values and the knowledge of the normal hill of vision, the starting values are determined for neighboring points. The program performs a full threshold

examination at each test location. A 1.5-1.5-1.5 dB bracketing strategy is used and is summarized as follows: a stimulus is presented at each location based upon an expected threshold level. If the subject sees the stimulus, the perimeter will decrease its intensity by 1.5 dB steps, until the subject is no longer able to perceive it. The program then increases the stimulus intensity by 1.5 dB steps, until the stimulus is seen again. Finally, it decreases the stimulus intensity by 1.5 db steps, until it is no longer seen. The average brightness (rounded to the nearest 1 dB) of the stimuli of the seen- and not-seen crossings is considered the threshold of differential light sensitivity for that location. If the stimulus is not seen initially, the program will increase its intensity by 1.5 dB intervals until it is seen. The stimulus intensity is then decreased by 1.5 dB steps until it is no longer seen. Finally, it is increased by 1.5 dB steps until it is seen, and the threshold is calculated.

With each visual field examination the patient's reliability was assessed in the traditional manner with fixation losses, false-positive and false-negative catch trials. A false-positive catch trial for the Digilab 1500 perimeter is accomplished by presenting a random, artificial time intervals, during which no stimulus is presented. A false-negative response was measured by retesting a stimulus, which had been determined to have a threshold value of 3 dB or greater. A failure to respond at the 0 dB level was considered a false-negative response. Fixation losses were measured with the Heijl-Krakau method. The

initial segment locates the point of lowest retinal sensitivity in the area of the physiological blind spot. Once a target of lowest sensitivity has been selected that target was presented randomly during the test.

During each test session, each subject performed two visual field tests with a five to ten minute break between the tests. Each subject underwent four such test sessions within thirty days, with at least a twenty-four hour interval between each session.

The visual fields were analyzed for inconsistencies in the subject's responses. If the subject claimed to have seen and not seen the same stimulus during the testing algorithm, it was counted as an inconsistent response. The number of inconsistent responses was summed over the entire visual field. After the testing was completed, false-positive catch trials, false-negative catch trials, fixation losses, and inconsistent responses were averaged from the eight visual fields of each subject. The average of inconsistent responses were compared to the averages of currently used visual field reliability parameters with Pearson's correlation coefficient. The sum of the false-negative and false-positive responses to catch trials for each visual field was termed "false responses". Short-term fluctuation was calculated from all test locations between the two examinations given during each test session.

The visual field examinations, data collection, data analysis were completed by the primary investigator. The eye examinations were performed by Dr. Mario Zulauf.

RESULTS

Twenty eyes of twenty volunteers underwent visual field examination. Thirteen subjects were males and seven were females. The age range was twenty to twenty-five years. Thirteen were Caucasians, six were Asians, and one was Hispanic. The median (range) refractive error was -1.37 (0 to -4.25). The median (range) time for visual field performance was 7.0 minutes (5.7 to 9.3 minutes).

The median (range) of the standard reliability parameters for the 160 visual fields were: false-positive replies 0% (0% to 43.8%), false-negative replies 0% (0% to 23.8%), fixation losses 2.7% (0% to 63.6%), and number of stimuli 378 (325 to 475). The median (range) number of inconsistent responses was 57 (29 to 91) per visual field.

Eight visual field examinations were performed on each subject during the four test sessions. The average threshold values were significantly lower ($p \leq 0.014$) in the first test (23.2 ± 1.4 dB) than in the subsequent tests (range: 23.8 - 24.4 dB). No statistically significant differences were found between test sessions for any of the reliability parameters (Table 1).

Inconsistent responses correlated significantly with the standard reliability parameters (Figures 4, 5, 6). The correlation (Pearson r, n=20) values for inconsistent responses with these indices were: false responses to catch trials 0.62 (p = 0.003), fixation losses 0.51 (p = 0.020), and number of stimuli 0.61 (p = 0.004). No significant correlation was detected between inconsistent responses and short-term fluctuation.

DISCUSSION

Because automated perimetry is a psychophysical test, one must have a means to assess patient reliability, in order to adequately interpret visual fields. Several indices are used to quantify reliability for commercial automated perimeters: false-positive and false-negative answers to catch trials, number of fixation losses. short-term fluctuation, and number of stimulus presentations required to complete a test. However, each of these parameters are influenced by many factors, such as abnormal visual fields, patient experience and fatigue. This hampers the ability of an ophthalmologist to assess the reliability of patient's responses.

In order to find an alternative to current indices, Lynn et al originally pursued the idea of examining the results of the bracketing strategy and used "boo-boo" counts as an indicator of reliability.[81] He only compared this index with short-term fluctuation and did not find a significant correlation. However, this idea of using the bracketing strategy could be pursued further, since Lynn et al used short-term fluctuation as a standard of comparison, which is presently considered a poor measure of reliability.[48-51,62,69,70,73,74]

This study compared the number of inconsistent responses with number of stimuli, fixation losses, and false responses to catch trials, which are the sum of false-positive and false-negative responses for each visual field. The inconsistent responses index used here does not make a distinction between an inconsistency that occurs with stimulus intensities less or greater than threshold, which is implied with false-positive and false-negative catch trials. The number of inconsistent responses showed a statistically significant correlation with the current reliability indices: $r = 0.61$ for number of stimuli presented, $r=0.51$ for fixation losses, and $r = 0.62$ for false-responses to catch trials. Although short-term fluctuation was not calculated in the conventional manner, no correlation was found with inconsistent responses, which agrees with the results of Lynn et al.[81]

The number of inconsistent responses as a reliability index offers several advantages over currently used reliability parameters. The distribution of fixation losses, false-positive and false-negative catch trials are all markedly skewed toward zero (Figures 7, 8, 9, 10). In this study 51.9% of the visual fields had no false-positive responses and 83.1% had no false-negative responses. Therefore, it may be difficult to assess whether these indices are truly providing reliability information or are offering a lack of it. Inconsistent responses to catch trials is more normally distributed, and may give more information about reliability because it provides some information about every

visual field. In addition false-positive and false-negative catch trials may provide only a rough estimation of patient reliability because they rely upon responses to extreme catch trials, i.e. no stimulus and a very bright stimulus. Inconsistent responses may give a more subtle and therefore more realistic and clinically relevant assessment of reliability. This index uses data from all patient responses, whereas current reliability parameters may only use anywhere from 3%-10% of stimuli presented, for catch trials and fixation losses. Hence, inconsistent responses may be a more robust reliability parameter. Finally, this index only increases computer processing time and does not require additional testing time; thus inconsistent responses may help reduce fatigue effects.

Although inconsistent responses has demonstrated much promise as a reliability index, it still needs to be explored much further. The major point that needs to be addressed is the selection of subjects. They are all healthy young subjects with normal visual fields, so this provides no information on whether or not age has an effect upon inconsistent responses. In addition this study demonstrates no statistically significant differences between test sessions for any of the currently used reliability parameters and for inconsistent responses. The literature suggests than a learning effect does occur with fixation losses,[44] catch trials.[44, 62, 63] and the number of stimuli presented.[61] The lack of improvement with patient experience, in this study, may be due to the choice of

very reliable subjects. Therefore, the presence or lack of a learning effect with inconsistent responses certainly has not been conclusively demonstrated and will need to be evaluated further.

To be clinically useful inconsistent responses must be able to assess reliability with abnormal visual fields. Inconsistent responses are inexorably linked to the frequency-of-seeing curve. Because abnormal visual fields will cause broadening of the frequency-of-seeing curve, the number of inconsistent responses will presumably increase with the degree of abnormality of the visual field. In addition, one may expect that the number of inconsistent responses will be particularly high around the border of scotomas. This has been shown to be true for short-term fluctuation, and this has been presumably attributed to slight eye movements.[72, 73] This imperfect fixation will probably cause inconsistent responses to increase independent of reliability. It may be possible to maintain the integrity of this index and get around this problem by excluding those test locations around the border of scotomas.

There are several aspects of the relationship between variability and inconsistent responses that need to be studied further. The literature has demonstrated that short-term fluctuation increases with eccentricity.[48, 50, 51, 74, 75] Because there is more threshold variability with eccentricity, inconsistent responses may also increase with eccentricity. Therefore, as is done with

short-term fluctuation, the use of inconsistent responses occurring in more central test locations may improve its ability to assess reliability. Not only is there variability with eccentricity, but also normal variability. As discussed earlier, inconsistent responses depends upon the frequency-of-seeing curve. Since the calculated threshold is theoretically considered the stimulus that is seen 50% of the time, those inconsistent responses close to threshold may not truly be "errors," but rather are normal physiological responses. It may be necessary to exclude those normal physiological inconsistent responses to improve the ability of this index to measure reliability.

Finally, other issues also need to be addressed with inconsistent responses. Currently, catch trial responses are used to assess reliability in different ways. Many false-positive answers are generally considered an indicator of an anxious or "trigger happy" patient and many false-negative responses tends to indicate an inattentive patient. Inconsistent responses may be able to convey this same information. Dividing the inconsistent responses that occur at intensities greater and less than the calculated threshold may be analogous to false-negative and false-positive catch trials, respectively. Inconsistent responses must be studied further with different testing algorithms. This study used a 1.5-1.5-1.5 dB strategy. However, the most commonly used automated perimeters, the Octopus and Humphrey, both use a 4-2-1 dB bracketing strategy. It is unknown what impact a different testing algorithm will

have on inconsistent responses.

With normal visual fields of healthy young subjects, inconsistent responses has been shown to be a good possibility as an alternative or at least a supplement to currently used reliability parameters. However, as discussed earlier, it must be studied further with abnormal visual fields to see if it will be able to avoid the problems that plague short-term fluctuation, fixation losses, number of stimuli presented, and false-positive and false-negative catch trials. When this work is completed, we will better know the true value of inconsistent responses as a measure of reliability.

REFERENCES

1. Von Grafe A: Uber dieubtersuchung des gesichtsfeldes bei amblyopischen affectionen. Arch Ophthal 2: 258-298, 1855.

2. Förster: Vorzeigung des Perimeter. Klin Monatsbl Augenheilkd 7: 411-422, 1869.

3. Bjerrum JP: Om en tilfojelse til saeddvanlige synsfeltundersogelse samt om synfeltet ved glaukom. Nord Ophth Tidsskr Kjobenh 2: 144-185, 1889.

4. Rönne H: Uber des Gesichtfeld beim Glaukom. Klin Monatsbl Augenheilk 47: 12-33, 1909.

5. Traquair HM: Introduction to clinical perimetry. London: Klimpton, 1927.

6. Walker CB: Some new instruments for measuring visual field defects. Arch Ophthalmol 42: 577-591, 1913.

7. Peter L: The Principles and Practice of Perimetry, 3rd ed. Philadelphia: Lea & Febinger, 1931.

8. Goldman H: Demonstration underes neuen Projektionskugel perimeters samt theoretischen und kinkschen Bermerkungen uber Perimetre. Ophthalmologica 111: 187-192, 1946.

9. Feree CE, Rand G, Sloan LL: Selected cases showing the advantages of a combined tangent screen and perimeter. Arch Ophthalmol 10: 166-184, 1933.

10. Harms H, Aulhorn E: Vergleichende Untersuchungen uber den Wert der quantitativen Perimetrie, Skiaskotometrie und Verschmelzungsfrequenz fur die Erkennung beginnender Gesichtsfeldstorungen beim Glaukom. Documenta Ophthalmologica 13: 303-332, 1959.

11. Aulhorn E, Harms H: Early visual field defects in glaucoma, in Glaucoma. Ed. W. Leydhecker. Basel. S. Karge Ltd, p.151-155, 1969.

12. Werner EB, Drance SM: Early visual field disturbances in glaucoma. Arch Ophthalmol 95: 1173-1175, 1977.

13. Drance SM: The early field defects in glaucoma. Invest Ophthalmol 8: 84-86, 1969.

14. Li SG, Spaeth GL, Scimeca HA, Schatz NJ, Savino NJ: Clinical experiences with the use of an automated perimeter (Octopus) in the diagnosis and management of patients with glaucoma and neurological diseases. Ophthalmology 86: 1302-1312, 1979.

15. Dubois-Poulson A, Magis C: Premiered applications a l'ophthalmologie des techniques moderned d'automation et d'analyse de l'information. Bull Mem Soc Franc Opthalmol 79: 576-583, 1966.

16. Lynn JR, Tate GW: Computer controlled apparatus for automatic visual field examination. US patent 3,833,234, issued May 1975.

17. Fankhauser F, Koch P, Roulier A: On automation of perimetry. Albr v. Graefes Arch Ophthalmol 184: 126-150, 1972.

18. Koch P, Roulier A, Fankhauser F: Perimetry-the information theoretical basis for its automation. Vis Res 12: 1619-1630, 1972.

19. Spahr J, Fankhauser F: On automation of perimetry - problems and solutions. In L'Anne theapeutique et clinique en ophthalmologie, Tome XXV. Marseille: Librairie Fueri Lami, p.337-345, 1974.

20. Heijl A, Krakau CET: An automatic perimeter for glaucoma visual field screening and control. Albr v Graefes Arch Ophthalmol 197: 13-23, 1975.

21. Fankhauser F, Spahr J, Bebie H: Some aspects of the automation of perimetry. Surv Opthalmol 22: 131-134, 1977.

22. Heijl A: Studies on computerized perimetry. Acta Ophthalmol (suppl) 132: 1-32, 1977.

23. Bebie H, Fankhauser F, Spahr J: Static perimetry: strategies. Acta Ophthalmol 54: 325-338, 1976.

24. Fankhauser F, Bebie H: Threshold fluctuations, interpolations, and spatial resolutions in perimetry. Docum Ophthalmol Ser 19: 295-309, 1979.

25. Flammer J, Zulauf M: The frequency distribution of the deviations in static perimetry. Doc Ophthalmol Proc Ser 42: 17-24, 1985.

26. Heijl A: Computerized perimetry in glaucoma management. Acta Ophthalmol (Copenh) 67: 1-12, 1989.

27. Langerhorst CT: Automated Perimetry in Glaucoma, Fluctuation Behavior and General and Local Resolution of Sensitivity. Amsterdam, Milano, Kugler & Ghedini, 1988, pp.67-69, 107-145.

28. Besky G von: A new audiometer. Acta Oto-laryng (Stokholm) 35: 411-422, 1974.

29. Cornsweet TN: The staircase method in psychphysics. Amer J Psychol 75: 485-491, 1962.

30. Spahr J: Optimisation of the presentation pattern in automated static perimetry. Vis Res 15: 1275-1281, 1975.

31. Heijl A: Computer test logics for automatic perimetry. Acta Ophthalmologica 55: 837-853, 1977.

32. Johnson CA, Keltner JL: Automated suprathreshold static perimetry. Am J Ophthalmol 89: 731-41, 1980.

33. Keltner JL, Johnson CA, Balestrery FG: Suprathreshold static perimetry: initial clinical trials with the Fieldmaster automated perimeter. Arch Ophthalmol 97: 260-272, 1979.

34. Johnson CA, Keltner JL, Balestrery FG: Suprathreshold static perimetry in glaucoma and other optic nerve disease. Ophthalmol 86: 1278-1286, 1979.

35. Mills RP: Evaluation of diagnostic capabilities of interactive test strategies in automated perimetry. Ophthalmol 92: 1181-86, 1985.

36. Weber B, Spahr J: Zur automatisierung der perimetrie. Acta Ophthalmol 54: 349-362, 1976.

37. Flammer J, Nagel G, Glowazki A, Moser HR, Fankhauser F: Detection and definition of scotomata of the central visual field by computer methods. Docum Ophthal Proc Ser 26: 33-41, 1981.

38. Mizoguchi F, Owada Y, Shirato S, Kitazawa Y: Three-dimensional display of static perimetry. Docum Ophthal Proc Ser 35: 493, 1983.

39. Hart WH, Harty RK: Computer-generated display for three-dimensional static perimetry. Arch Ophthalmol 100: 312-318, 1982.

40. Bebie H, Flammer J, Bebie Th: The cumulative defect curve, separation of local and diffuse components of visual field damage. Graefes Arch Clin Exp Ophthalmol 227: 9-12, 1989.

41. Zulauf M, Caprioli J, Boeglin B, Lee M: Number of stimuli as a reliability parameter in perimetry. German J Ophthalmol 1: 86-90, 1992.

42. Heijl A, Krakau CET: An automatic static perimeter, design and pilot study. Acta Ophthalmol 53: 293-310, 1975.

43. Katz J, Sommer A: Reliability indexes of automated perimetric tests. Arch Ophthalmol 106: 1252-1254, 1988.

44. Blicker-Bluth M, Trick GL, Kolker AE, CooperDG: Assessing the utility of reliability indicies for automated visual fields. Opthalmol 96: 616-619, 1989.

45. Nelson-Quigg JM, Twelker JD, Johnson CA: Test responses properties of normal observers and patients during automated perimetry. Arch Ophthalmol 107: 1612-1615, 1989.

46. Aulhorn E, Harms H: Visual perimetry. In Handbook of Sensory Physiology. Eds Jameson D, Hurvich LM. Berlin: Springer-Verlag 1972, pp.102-145.

47. Heijl A, Lindgren G, Olsson J: A package for the statistical analysis of computerized fields. Doc Ophthalmol Proc Ser 49: 153-168, 1987.

48. Flammer J, Drance SM, Fankhauser F, et al: Differential light threshold in automated static perimetry. Factors influencing short-term fluctuation. Arch Ophthalmol 102: 876-879, 1984.

49. Flammer J, Drance SM, Zulauf M: Differential light threshold. Short- and long-term fluctuation in patients with glaucoma, normal controls, and patients with suspected glaucoma. Arch Ophthalmol 102: 704-705, 1984.

50. Rutishauser C, Flammer J, Haas A: The distribution of normal values in automated perimetry. Graefes Arch Clin Exp Ophthalmol 227: 513-517, 1989.

51. Heijl A, Lindgren G, Olsson J: Normal variability of static perimetric threshold values across the central visual field. Arch Ophthalmol 105: 1544-1549, 1987.

52. Katz J, Sommer A: A longitudinal study of the age-adjusted variability of automated visual fields. Arch Ophthalmol 105: 1083-1086, 1987.

53. Young WO, Stewart WC, Hunt H, et al: Static threshold variability in the peripheral visual field in normal subjects. Graefes Arch Clin Exp Ophthalmol 228: 454-457, 1990.

54. Caprioli J. Automated Perimetry in Glaucoma: The Interpretation of Visual Fields. In: Walsh, TJ ed. Rochester, MN: American Academy of Ophthalmology, 1990, pp.71-106.

55. Heijl A: Time contrast thresholds during automated perimetry. Acta Ophthalmol 55: 696-708, 1977.

56. Heijl A, Drance SM: Changes in the differential threshold in patients with glaucoma during prolonged perimetry. Br J Ophthalmol 67: 512-516, 1983.

57. Marra G, Flammer J: The learning and fatigue effects in automated perimetry. Graefes Arch Clin Exp Ophthalmol 229: 212-217, 1991.

58. Johnson CA, Adams CW, Lewis RA: Fatigue effects in automated perimetry. Appl Optics 27: 1030-1037, 1988.

59. Heijl A, Drance SM: Deterioration of threshold in glaucoma patients during perimetry. Doc Ophthalmol Proc Ser 35: 129-136, 1983.

60. Fankhauser F, Koch P: On automation of perimetry. Albrecht v Graefes Arch Ophthalmol 184: 126-150, 1972.

61. Wild JM, Dengler-Harles M, Searle AE, et al: The influence of learning effect on automated perimetry in patients with suspected glaucoma. Acta Ophthalmol 67: 537-545, 1989.

62. Heijl A, LIndgren G, Olsson J: The effect of perimetric experience in normal subjects. Arch Ophthalmol 107: 81-86, 1989.

63. Katz J, Sommer A, Witt K: Reliability of visual field results over repeat testing. Opthalmol 98: 70-75, 1991.

64. Reynolds M, Stewart WC, Sutherland S: Factors that influence the prevalence of positive catch trials in glaucoma patients. Graefes Arch Clin Exp Ophthalmol 228: 338-341, 1990.

65. Katz J, Sommer A: Screening for glaucomatous visual field loss. The effect of patient reliability. Ophthalmol 97: 1032-1037, 1990.

66. Newman S, Wall M: The effect of head tilt on fixation monitoring of the Humphrey perimeter. In Mills RP, Heijl A: Perimetry Update, 1990/91, Proceedings of the IXth International Perimetric Society Meeting. Amsterdam, Kugler & Ghedini, 1991, p. 111.

67. Henson DB, Bryson H: Is the variability in glaucomatous field loss due to poor fixation control? In Mills RP Heijl A: Perimetry Update, 1990/91, Proceedings of the IXth International Perimetric Society Meeting. Amsterdam, Kugler & Ghedini, 1991, pp.217-220.

68. Newman SA: Insensitivity of the automated monitoring system of the Octopus 2000R to small eye movements. Invest Ophthalmol Vis Sci Suppl 29: 270, 1987.

69. Werner EB, Petrig B, Krupin T, Bishop KI: Variability of automated visual fields in clinically stable glaucoma patients. Invest Ophthalmol Vis Sci 30: 1083-1089, 1989.

70. Heijl A, Lindgren A, Lindgren G: Test-retest variability in glaucomatous visual fields. Am J Ophthalmol 108: 130-135, 1989.

71. Heijl A, Lindgren A, Lindgren G: Inter-test threshold variability in glaucoma: Importance of censored observations and general field status. In Mills RP, Heijl A: Perimetry Updat, 1990/91, Proceedings of the IXth International Perimetric Society Meeting, Kugler & Ghedini, 1991, pp. 189-192.

72. Stark H: Fluctuation and stable elements in the visual field in glaucoma. Klin Monatsbl Augenheilkd 191: 16-21, 1987.

73. Haefliger IO, Flammer J: Increase of the short-term fluctuation of the differential light threshold around a physiological scotoma. Am J Ophthalmol 107: 417-420, 1989.

74. Heijl A, Asman P: A clinical study of perimetric probability maps. Arch Ophthalmol 107: 199-203, 1989.

75. Rabineau PA, Gloor B, Tobler HJ: Fluctuations in threshold amd effect of fatigue in automated static perimetry (with Octopus 201). Doc Ophthalmol Proc Ser 42: 25-33, 1985.

76. Rutishauser C, Flammer J: Retests in static perimetry. Graefes Arch Clin Exp Ophthalmol 226: 75-77, 1988.

77. Autzen T, Work K: The effect of learning and age on short-term fluctuation and mean sensitivity of automated static perimetry. Acta Ophthalmol 68: 327-330, 1990.

78. Haas S, Flammer J, Schneider U: Influence of age on the visual fields of normal subjects. Am J Ophthalmol 101: 199-203, 1986.

79. Werner EB, Adelman A, Krupin T: Effect of patient experience on the results of automated perimetry in clinically stable glaucoma patients. Ophthalmol 95: 764-767, 1988.

80. Werner EB, Krupin T, Adelman A, Feitel ME: Effect of patient experience on the results of automated perimetry in glaucoma suspect patients. Ophthalmol 97: 44-48, 1990.

81. Lynn JR, Batson EP, Fellman RL: Internal inconsistencies versus root mean square as measure of threshold variability. Doc Ophthal Proc Ser 39: 7-15, 1985.

FIGURE LEGENDS

Table 1. Current reliability indices and inconsistent responses for all the visual field examinations. The values are given as the mean ± standard deviation.

Figure 1. Frequency-of-seeing curve. As the stimulus intensity increases (abscicca) increases, the probability of seeing the stimulus increases (ordinate). The threshold is defined as the light intensity whereby the subject sees the stimulus 50% of the time. The probability of perceiving a stimulus never reaches either 0% or 100% because of the possibility of false-negative and false-positive responses.

Figure 2. Quantitative threshold determination. These are four examples of threshold determination, using a 4-2 dB bracketing method. In the first phase (A), a 4 dB stimulus is used until the threshold is crossed. In the second phase (B), the stimulus size step is decreased to 2 dB, and the luminosity is changed in the opposite direction until the threshold is crossed once again.

Figure 3. The visual field test pattern used in this study consisted of 44 test locations within 30 degrees.

Figure 4. The number of inconsistent responses versus false response rate to catch trials. Inconsistent responses correlated with false responses to catch trials, which represents the sum of false-positive and false-negative replies ($r = 0.62$, $p = 0.003$, $n = 20$).

Figure 5. The number of inconsistent responses versus fixation losses. Inconsistent responses correlated with fixation losses ($r = 0.51$, $p = 0.020$, $n = 20$).

Figure 6. The number of inconsistent responses versus the number of stimuli presented. Inconsistent responses correlated with the number of stimuli used during the visual field examination ($r = 0.61$, $p = 0.004$, $n = 20$).

Figure 7. Distribution of inconsistent responses. This index is somewhat normally distributed.

Figure 8. Distribution of the number of stimuli presented. Like inconsistent responses, the number of stimuli demonstrates a fairly normal distribution.

Figure 9. Distribution of false responses to catch trials. In contrast to inconsistent responses, this shows a marked skew towards zero.

Figure 10. Distribution of fixation losses. This index is conspicuously skewed toward zero. In many cases this makes it difficult to determine whether this index is offering reliability information or a lack of it.

Table 1. - Current Reliability Indices and Inconsistent Responses

Visual Field	Inconsistent Responses	False-Positive Responses(%)	False-Negative Responses (%)	Fixation Losses(%)	Number of Stimuli
1	59.7 ± 11.5	2.2 ± 3.1	1.9 ± 5.1	5.7 ± 4.8	386.7 ± 32.4
2	55.8 ± 9.2	5.0 ± 7.7	1.3 ± 3.4	6.3 ± 4.8	381.1 ± 20.1
3	59.3 ± 10.2	7.7 ± 8.8	1.5 ± 4.0	4.2 ± 5.0	383.6 ± 30.2
4	60.3 ± 12.1	7.8 ± 8.8	2.0 ± 4.4	5.3 ± 4.2	384.3 ± 18.9
5	56.1 ± 6.9	7.3 ± 11.2	1.0 ± 2.4	5.8 ± 5.2	379.1 ± 19.5
6	58.1 ± 13.5	3.7 ± 9.8	2.1 ± 5.6	4.9 ± 4.3	377.9 ± 23.6
7	54.4 ± 10.3	4.9 ± 8.9	0.7 ± 2.2	5.7 ± 4.9	379.8 ± 24.2
8	56.0 ± 11.2	6.7± 6.7	1.7± 3.0	6.1 ± 5.1	379.3 ± 23.6

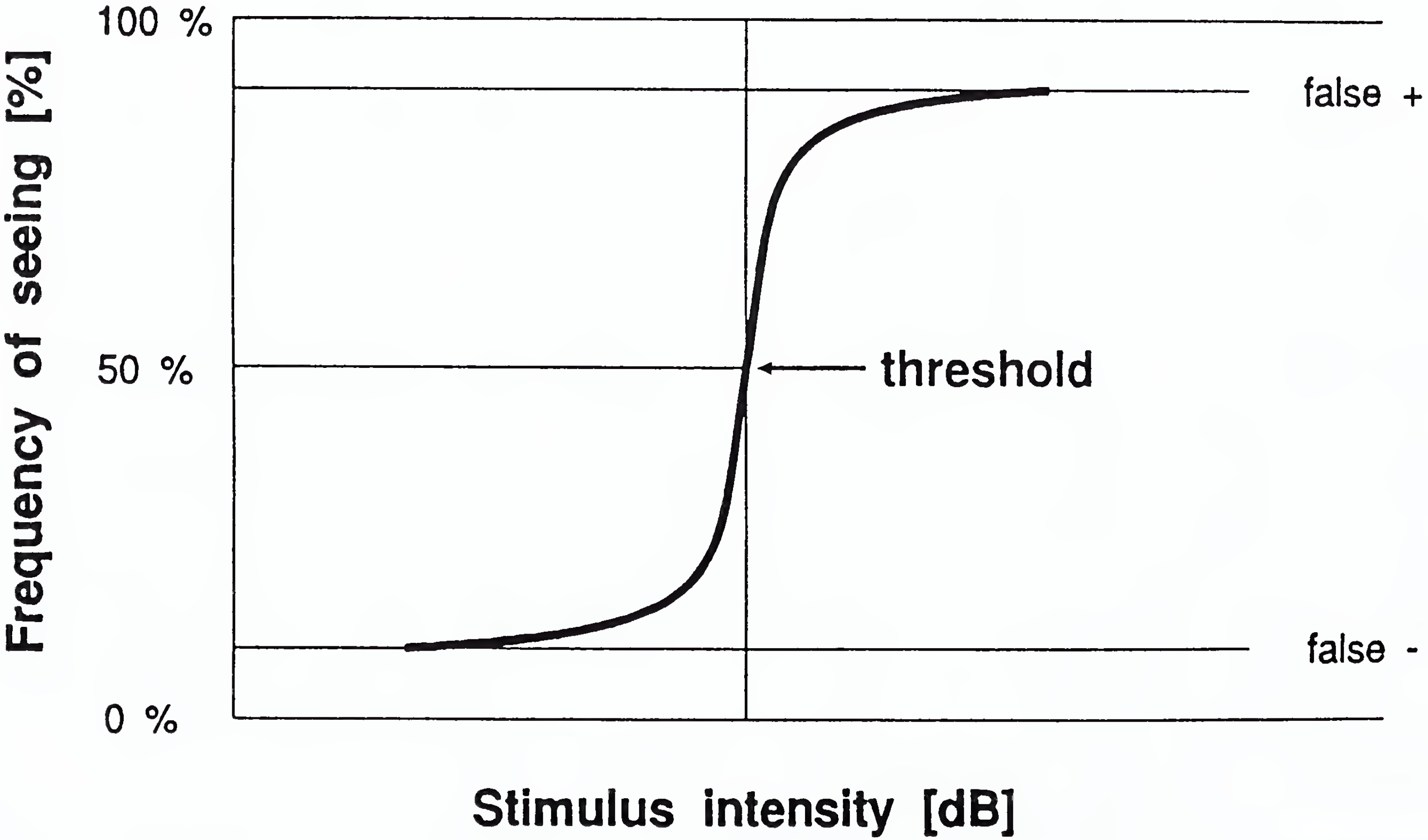
Frequency of seeing [%]
100 %
50 %
0 %
false +
false -
threshold
Stimulus intensity [dB]

Figure 1.

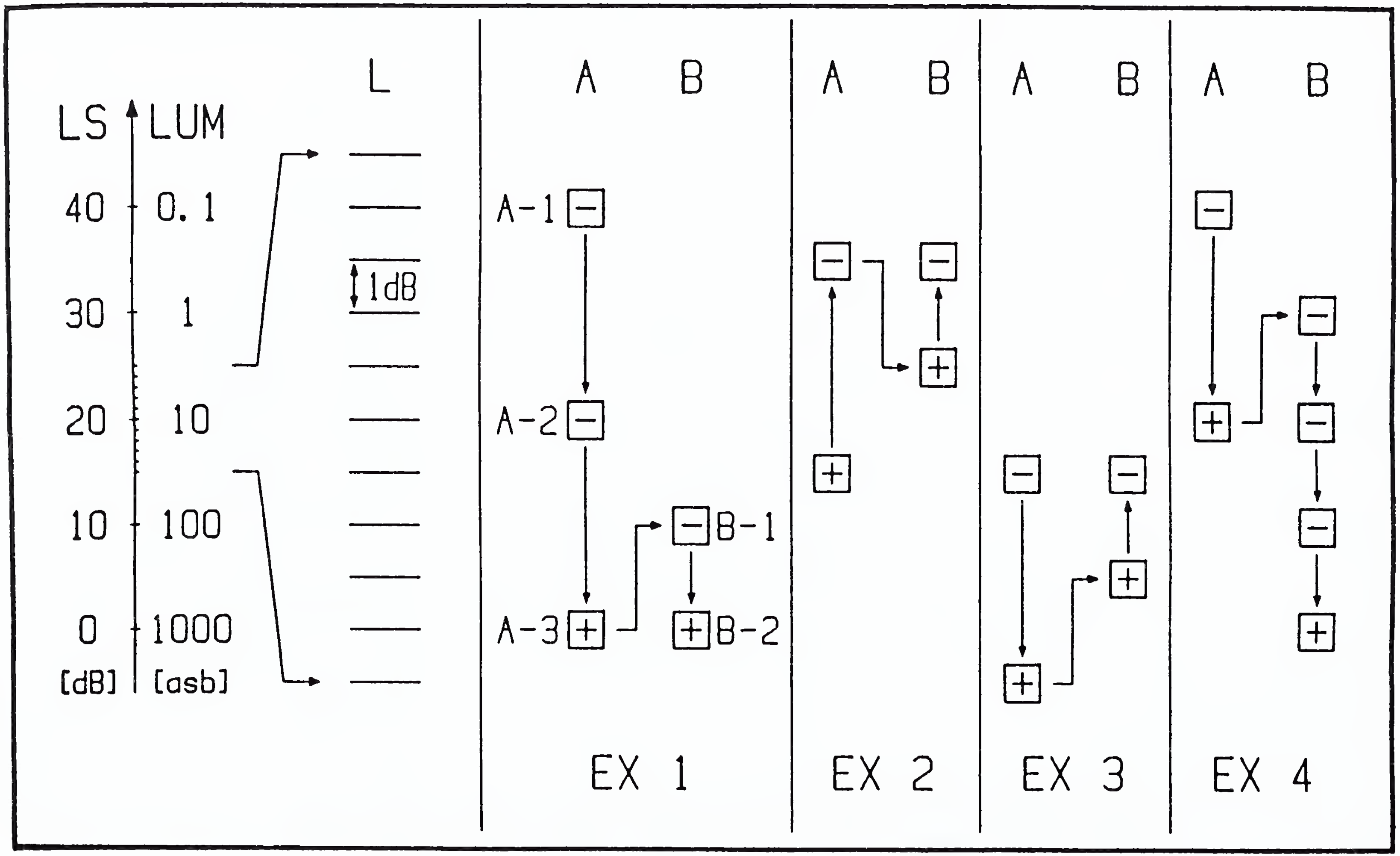
LS
LUM
40
30
20
10
0
[dB]
0.1
1
10
100
1000
[asb]
L
1dB
A
B
A-1
A-2
A-3
B-1
B-2
EX 1
EX 2
EX 3
EX 4

Figure 2.

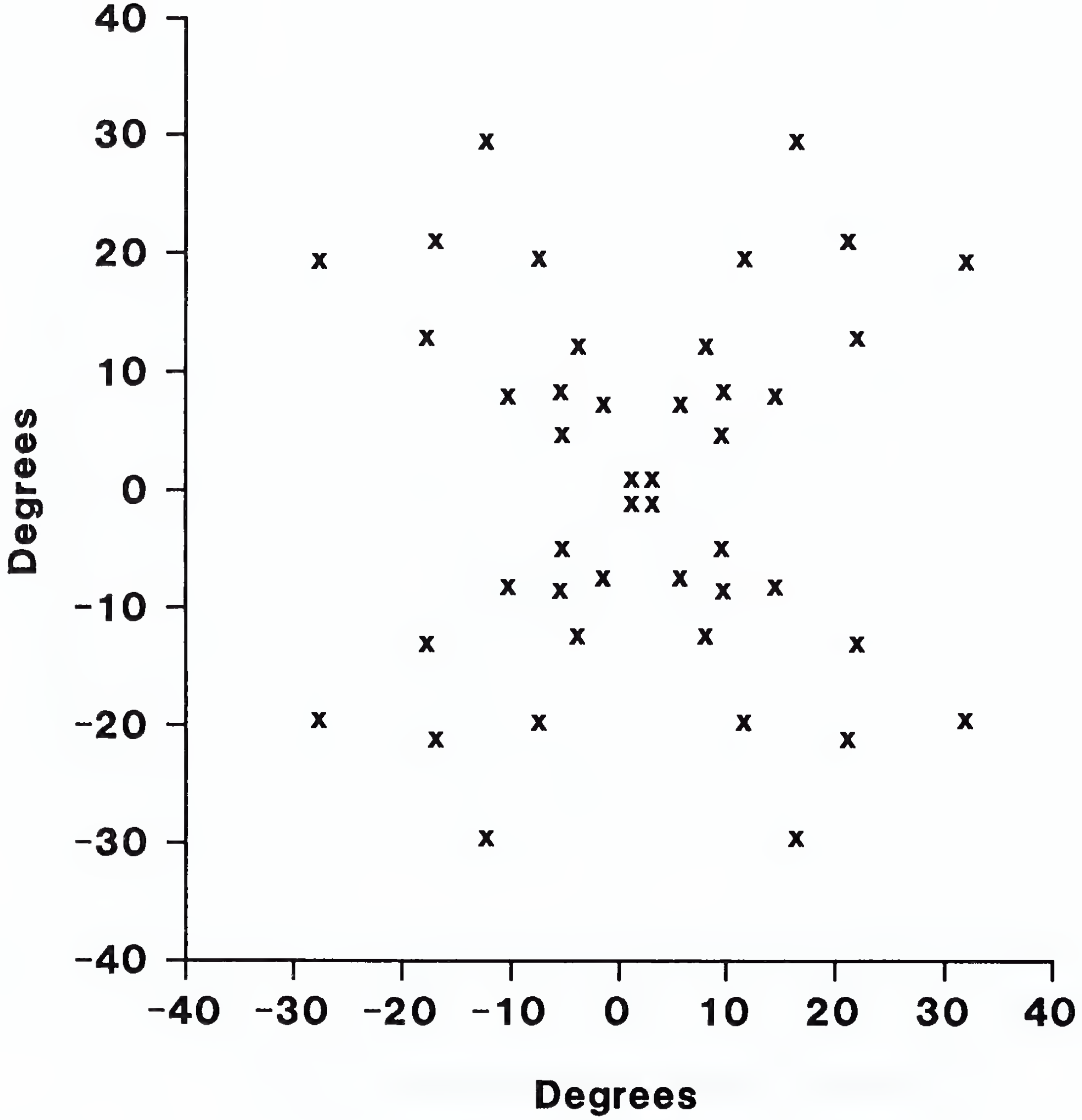

Figure 3.

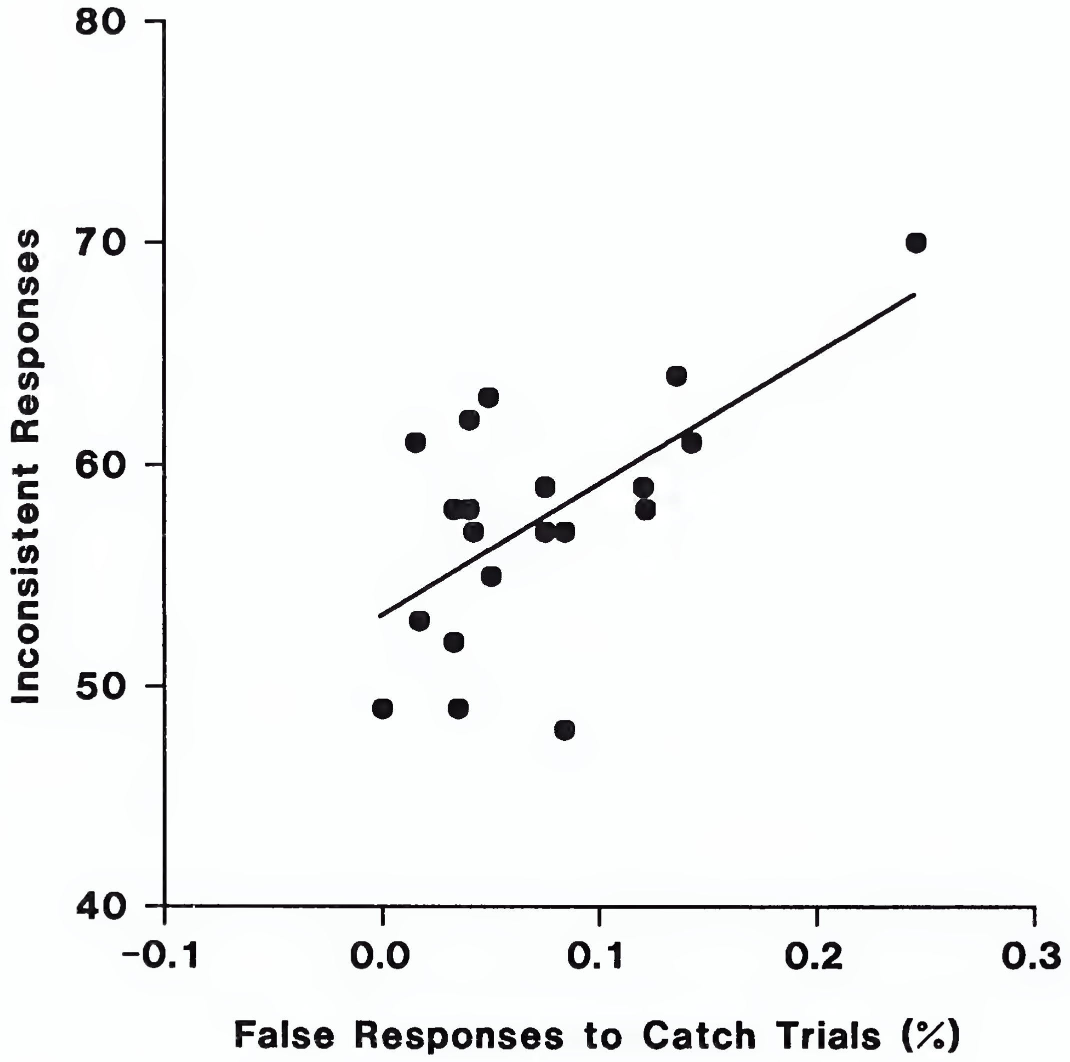
Inconsistent Responses
80
70
60
50
40
-0.1
0.0
0.1
0.2
0.3
False Responses to Catch Trials (%)

Figure 4.

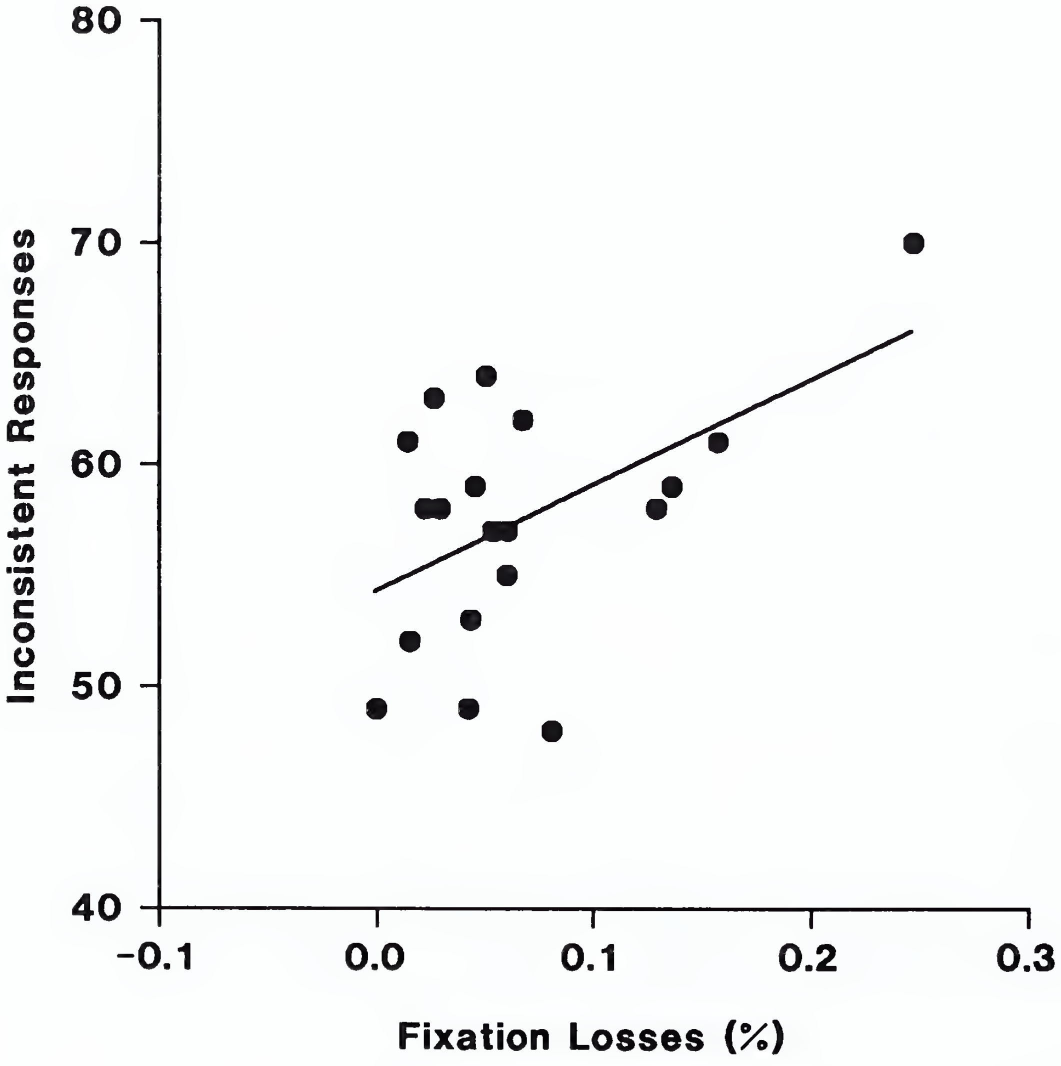
80
70
60
50
40
Inconsistent Responses
-0.1
0.0
0.1
0.2
0.3
Fixation Losses (%)

Figure 5.

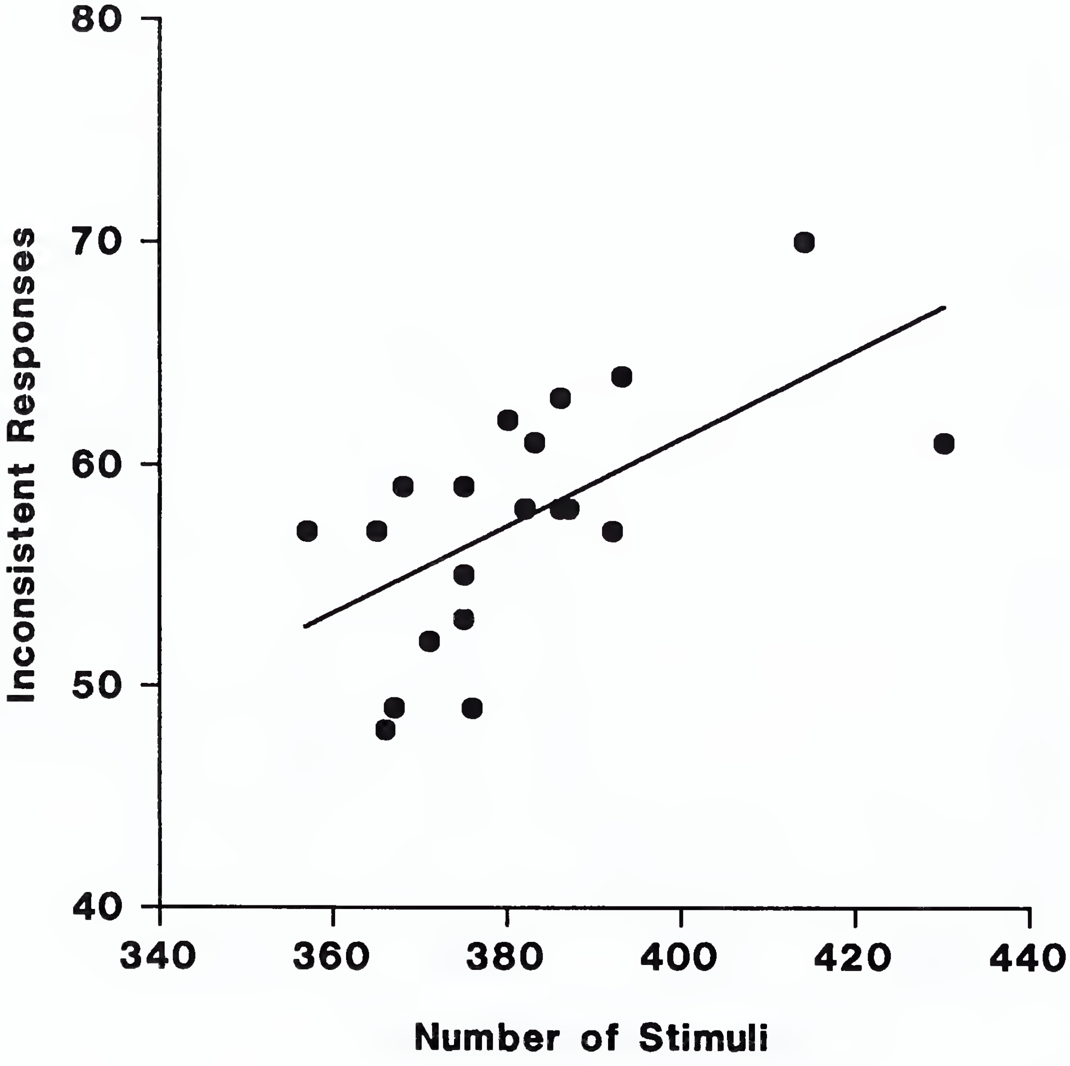
80
70
60
50
40
Inconsistent Responses
340
360
380
400
420
440
Number of Stimuli

Figure 6.

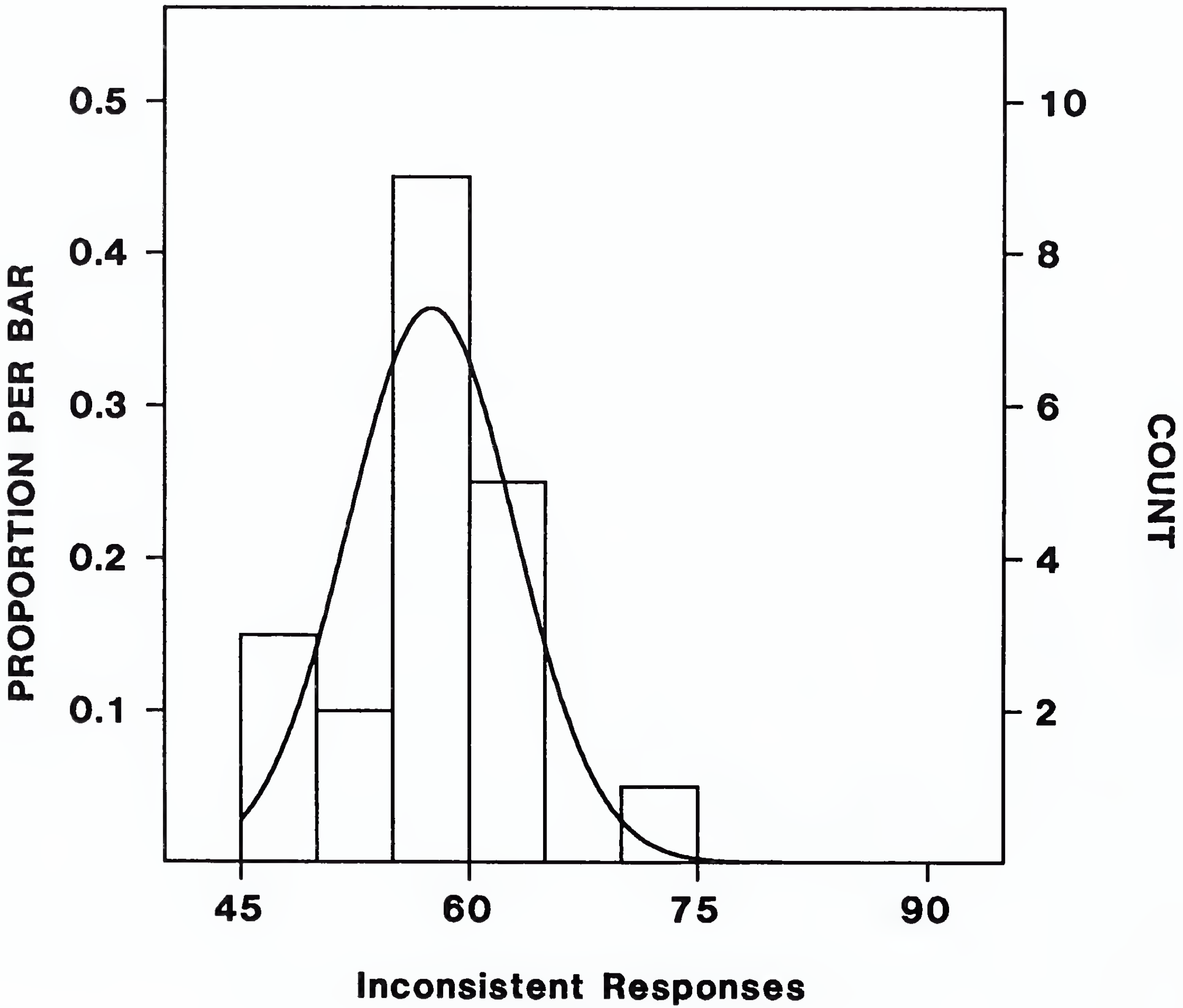
0.5
0.4
0.3
0.2
0.1
PROPORTION PER BAR
10
8
6
4
2
COUNT
45
60
75
90
Inconsistent Responses

Figure 7.

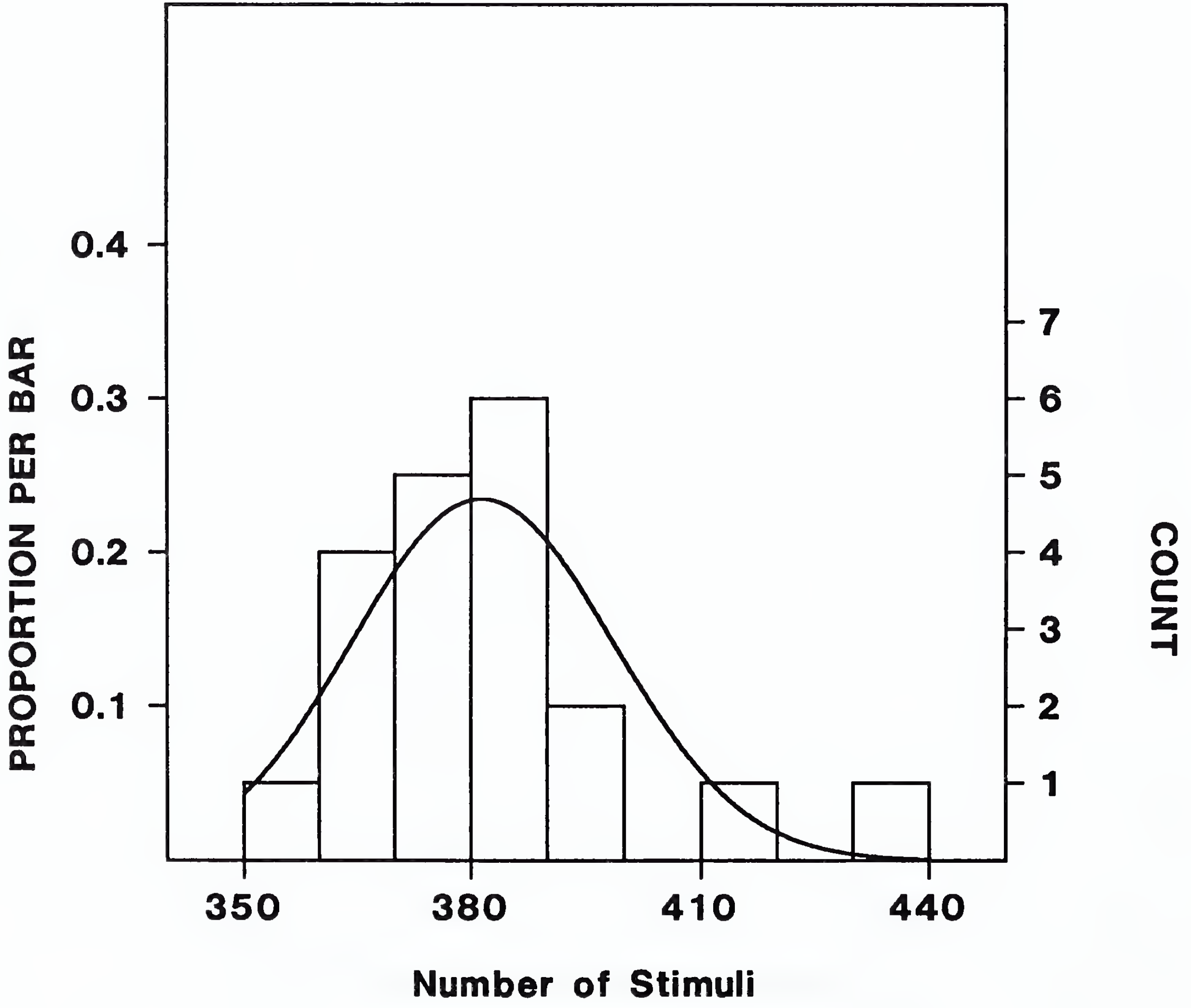
PROPORTION PER BAR
0.1
0.2
0.3
0.4
COUNT
1
2
3
4
5
6
7
350
380
410
440
Number of Stimuli

Figure 8.

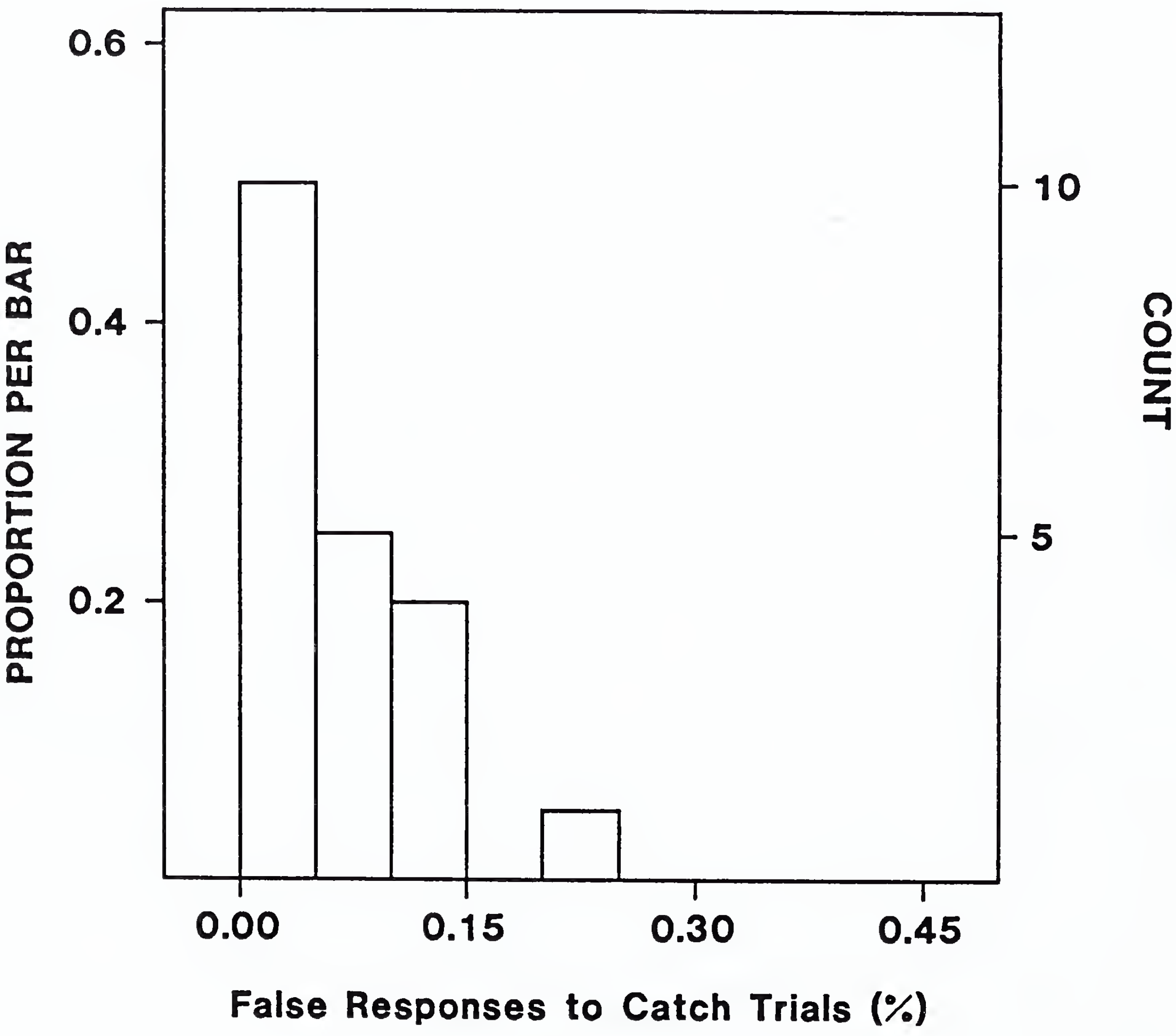
PROPORTION PER BAR
0.6
0.4
0.2
10
5
COUNT
0.00
0.15
0.30
0.45
False Responses to Catch Trials (%)

Figure 9.

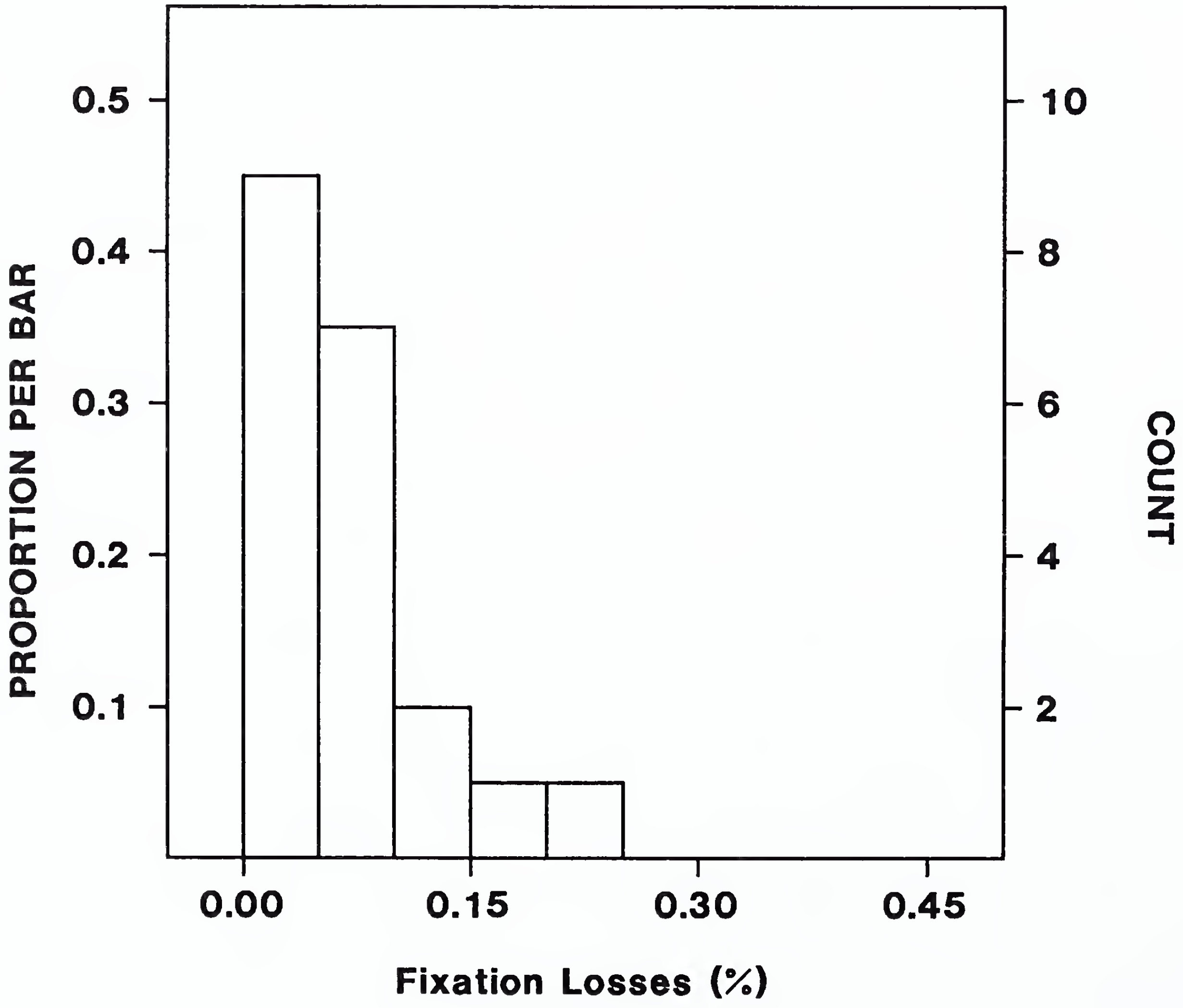
PROPORTION PER BAR
0.1
0.2
0.3
0.4
0.5
COUNT
2
4
6
8
10
0.00
0.15
0.30
0.45
Fixation Losses (%)

Figure 10.