SESSION IV
EYE MOVEMENTS: A SYMPOSIUM

Richard K. Olson, *Presider*

Evaluating and reporting data quality in eye movement research

GEORGE W. McCONKIE
University of Illinois, Champaign, Illinois 61820

In order to judge the degree of confidence one should have in the results of an experiment using eye movement records as data, it is necessary to have information about the quality of the eye movement data themselves. Suggestions are made for ways of assessing and reporting this information. The paper deals with three areas: characteristics of the eye movement signal, algorithms used in reducing the data, and accuracy of the eye position data. It is suggested that all studies involving eye movement data should report such information. Appendices include linear interpolation algorithms for mapping from the eye movement signal to stimulus space and a way of obtaining an index of accuracy for each data point.

In recent years, there has been an upsurge in the use of eye movement data in psychological research (Levy-Schoen & O'Regan, 1979, Monty & Senders, 1976; Rayner, 1978). There has also been considerable development in eye movement monitoring technology, and at present there are a number of techniques in use for collecting such data (e.g., see Young & Sheena, 1975). The process of obtaining reliable and accurate eye movement records is not an easy one, and there are many potential sources of error of various types. This makes it particularly important that reports of eye movement research include information that allows knowledgeable readers to assess the quality of the eye movement data obtained in the study and, hence, to be able to judge the degree of confidence that they should place in the results of the study. So far, no general format has been proposed for reporting this kind of information. In fact, while it is obvious that information about the accuracy and reliability of the data should be presented, it is often not clear just how to make such a report. The purpose of this paper is to

suggest what information investigators should report about the quality of their eye movement data and to recommend ways of reporting this information, so that others can more effectively evaluate their research.

It would be inappropriate to set standards for what is and is not sufficiently good eye movement data for research purposes. The degree of reliability and accuracy of the data that is needed for investigating different questions varies dramatically. Some studies need only information about whether the eyes moved in one direction or another, other studies need data on the durations of fixations, and still others need an accurate indication of exactly where in the stimulus pattern the eyes were directed on a given fixation. The first of these puts few constraints on the quality of data, whereas the last requires great precision that is difficult to achieve. Thus, rather than attempting to adopt standards concerning what constitutes acceptable data, it will be more useful to make a list of items that might be reported in studies involving eye movement data. This would provide a more or less standard basis for making comparisons among studies. An example of such a list is given in Table 1. Not all items will be appropriate for every study; rather, investigators should include those items that are necessary for evaluating the quality of those aspects of the data that are used in their particular studies or research program.

The information that might be reported falls into three categories: characteristics of the signal itself, algorithms used for reducing the data, and accuracy of the data from which the results of the experiment are obtained. Each of these topics will be discussed below,

0005-7878/81/020097-10$01.25/0

Table 1
A List of Items to Report in Indicating
Eye Movement Data Quality

A. Characteristics of the signal
  1. Sampling rate
  2. Delay of signal
      a. Time required for obtaining information to calculate eye position
      b. Further delay until eye position signal is available for sampling
      c. Further delay until sample is obtained
      d. Additional time required for converting the sample to a usable form
  3. Maximum tracking rate of the eye movement monitoring equipment
  4. Noise characteristics of the signal
  5. Drift
B. Algorithms used in reducing data
  1. Algorithm for identifying beginning of a saccade
  2. Algorithm for identifying end of a saccade
  3. Algorithm for identifying where the eyes are directed during a fixation
      a. Nature of the calibration task
      b. Nature of the calibration table
  4. Algorithm for identifying disturbances in the eye movement data
C. Accuracy of the eye position data
  1. Degree of short-term repeatability
  2. Accuracy of the mapping function
  3. Degree of longer term repeatability

with suggestions for the types of measures that would be appropriate. This discussion will be simplified by assuming the monitoring of only a single dimension, the horizontal component of eye movements. Corresponding information should be reported for the vertical component, if that is important in the study. When the inclusion of both dimensions creates special problems in assessing or reporting the quality of data, this will be noted.

In order to standardize the data quality measures, it will be assumed that the stimulus display region is divided into a large number of small rectangular areas, all of the same size, by laying an imaginary grid over it. Each area will be referred to as an L-area. The width and height of each of these L-areas will be referred to as horizontal and vertical L-units, and these will be taken as the units for measurement of the data quality. In reading research, for instance, each L-area holds a single letter. The L-areas moving horizontally across the page are referred to as letter positions, and the L-areas moving vertically, as lines. In most picture perception studies, there are no such convenient elements in the stimulus display itself and the grid producing L-areas must be arbitrarily created. The use of L-areas and L-units permits the quality indexes to be reported in a more standardized fashion and thus permits easier interpretation of the indexes and easier comparison among studies.

The first thing a report should include, then, would be actual width and height of the L-areas in millimeters, thus defining the horizontal and vertical L-units for the study. The average width and height in degrees of visual angle from the position of the subject should also be reported. This will be an average, because the actual number of degrees will vary slightly over a flat surface since different locations will be slightly different distances from the eyes. The viewing distance and the visual angle of the entire display should also be reported. Finally, the experimenter should calculate the amount of change in the eye movement monitor (EMM) signal that typically results when subjects move their eyes a distance equivalent to 1 L-unit. Thus, if the EMM provides an analog signal that is digitized for storage, this computation would indicate the typical movement in these A/D values that occurs with a movement of the eyes of 1 L-unit. If this varies considerably over different parts of the visual field, or for different subjects, some indication of the range of this variation should also be reported.

For future reference, the typical amount of change in the EMM signal resulting from moving the eyes 1 L-unit is referred to as a Tinker, in honor of a prominent eye movement researcher. Thus, the Tinker is the unit of movement in EMM data space equivalent to a movement of 1 L-unit in the stimulus space. Of course, with two-dimensional eye tracking, there will be both horizontal and vertical L-units and Tinkers.

In some systems, the EMM output is given directly in terms of the stimulus space, using internal processing to map from the original eye position signal to the visual display. In this case, the units provided can be adopted as L-units, and Tinker units would then be on the same scale.

CHARACTERISTICS OF THE EYE MOVEMENT SIGNAL ITSELF

There are five characteristics of the raw eye movement signal that should be investigated and reported: the sampling rate, the delay, maximum tracking rate, noise characteristics, and drift.

Sampling Rate
The time in milliseconds between taking successive samples of the eyes' position should be reported.

Delay in the Signal
When information about the eyes' location becomes available for sampling, this information is necessarily lagging behind the actual location of the eyes. A good estimate of the delay in this signal is important for evaluating some types of research, particularly that involving eye movement-contingent stimulus control. The amount of this delay is not always easy to estimate. However, an estimate can be made on the basis of four facts about the eye movement recording apparatus and associated equipment. First, how long does it take the equipment to obtain the information needed to compute

the eye's location? For instance, if a TV monitor is being used to record eye movements, it may take 16 msec for the camera to complete a scan of the eye. In the case of limbus reflection techniques, the information is almost immediately available. Other techniques typically lie between these extremes. Second, how much time transpires between the moment the information needed to compute the eye's location is available and the moment at which the eye position information actually becomes available to be recorded or sampled by the computer? Delays may be induced at this stage by filters or signal processing requirements. Third, how long is it after the information becomes available before the computer or other recording device actually has the sample? Delays at this stage may result from slow sampling rates, from time required for digitizing an analog signal, or from averaging over repeated samples for the purpose of reducing noise in the signal. Fourth, if the data are provided in one form (say, as values indicating eye position in the EMM space) but to be used must be transformed to some other form (say, as values indicating where the eyes are centered in the stimulus array), the time required to make this transformation should also be included in calculating the delay in the signal.

If the maximum tracking rate of the EMM equipment is too low, this can also contribute to a delay in the signal during and immediately following saccadic eye movements. This problem will be dealt with in more detail in the next section.

In systems that give a stimulus position directly as output, these functions are handled internally and may not be available for test. In this case, the manufacturer should provide precise indications of the delays involved.

Information concerning delay in the signal is of importance for studies in which stimuli are being manipulated in real-time in response to characteristics of eye movements. When no such eye movement-contingent stimulus control is taking place, signal delay need not be reported.

Maximum Tracking Rate of the
Eye Movement Equipment

During saccadic eye movements, the eyes reach velocities as great as 830 deg/sec (Alpern, 1971). Peak velocities vary with the lengths of the saccades. If the signal produced by the eye movement equipment is not capable of changing fast enough to respond at the peak velocity rates of the eye movements typically observed in the task being studied, this can have several effects. A delay in the signal will occur during saccadic movements. The eye movement velocity pattern obtained during saccades may be inaccurate, at least for saccades above a certain length. The time duration of saccadic movements may be inflated, and as a result, the durations of fixations may be underestimated, especially for fixations following longer saccades.

A lower maximum tracking rate can result from electronic filtering of the signal in an attempt to reduce noise, from equipment requiring mechanical movement in eye tracking, or from other sources.

Investigators should report the maximum tracking rate of the equipment they are using. This should be obtainable from the manufacturer, or it can be assessed by monitoring the movements of an artificial eye that can be accurately moved at different rates.

Noise Characteristics of the Signal

There are two types of noise in the eye movement signal that should be reported: local noise and repetitious patterns. Local noise concerns the amount of variation in the EMM signal from one sample to the next when the eyes are in a fixation. It should be recognized, of course, that during a fixation there is some degree of movement of the eyes, and it would not be a reasonable goal to attempt to obtain a signal that shows no change at all during a fixation. However, this movement tends to be very small with respect to the amount of noise found in the signal of most EMM equipment.

In order to estimate the amount of local noise present in the signal, a series of fixations should be selected, and within these each successive data value should be subtracted from the value obtained previously, to yield a difference value. The absolute value of these differences should then be obtained. Information concerning the distribution of these values should be reported. This can be done by reporting the median and the 90th percentiles of this distribution, for instance. Dividing these indexes by the value of a Tinker will transform them into a measure based on L-units and will indicate the level of noise obtained relative to stimulus space units appropriate for the experimental situation. If the amount of this variability changes from one part of the stimulus display to another (e.g., if greater variability is found as the eyes move into regions that yield the highest EMM value), then distributions should be reported from both the low-variability and high-variability regions.

The experimenter should also examine the raw data for repetitious patterns that may be present but do not show up in sample-to-sample differences. For instance, a 60-Hz noise pattern resulting from changes in light intensity in the experimental room, or line noise, should be noted, together with an indication of its extent. Again, the size of this noise should be checked at both the low and high regions of the EMM signal, and if there is a difference this should be mentioned. As before, the range of this noise can be converted to a more useful form by dividing it by the value of a Tinker.

Drift

The final aspect of the eye movement signal itself that needs to be assessed and reported is the drift. Often, the EMM signal will change over time with no change in the stimulus conditions, simply because of

temperature changes or other factors that affect the electrical characteristics of the equipment. This should be assessed by establishing some type of standard stimulus situation that can be held constant for a period of time. This may involve the use of a stationary artificial eye, for instance. The equipment should then be adjusted to provide an output signal in the low range of the EMM signal, and it should be sampled regularly, say every 15 sec, over a period of time equal to that typically required for a subject to complete the experimental task being studied. This same test should be repeated with the equipment adjusted to yield an output at the high end of the EMM signal range. The timing of this test should be similar to the typical use of the equipment for data collection. That is, if data are typically collected immediately after the EMM equipment is turned on, the test should be made the same way; if the equipment is typically allowed to warm up for a period of time, the test should be done after similar warm-up. Data from this test should be included in the description of EMM signal characteristics.

Summary

The report of suggested information concerning sampling rate, delay, maximum tracking rate, noise, and drift in the signal will help readers understand some of the problems encountered by the experimenter in making decisions about when fixations began and ended, where the eyes were directed, and so on. Some of the problems involved are discussed further by McConkie, Zola, Wolverton, and Burns (1978).

ALGORITHMS USED IN REDUCING THE DATA

Eye movement research often requires four algorithms that convert the raw data to data showing a series of fixations at particular stimulus locations. Some studies do not need all four types of information and, hence, do not require algorithms of all four types. The algorithms are for (1) identifying the beginning of a saccade (or end of a fixation), (2) identifying the end of a saccade (or beginning of a fixation), (3) identifying where in the stimulus display the eyes were directed during that fixation, or identifying the direction and extent of a saccade, and (4) identifying disturbances in the eye movement data that suggest that the data should not be used (e.g., blinks, squints, or other irregularities). The nature of the algorithm that must be used to accomplish each of these depends greatly on the characteristics of the signal itself, particularly the level of noise, and on the nature of the calibration task used and the type of information it provides for use in transforming the data from EMM signal space to stimulus display space.

The algorithms used for these purposes, insofar as they are applicable to the study being reported, should be described, or reference should be made to some source in which they are publicly available.

Examples of algorithms for taking a linear interpolation approach to map from EMM signals to stimulus locations (i.e., to indicate where in the stimulus the eyes were directed at any given moment) are given in Appendices A and B. Appendix A presents a common simple algorithm for use in one-dimensional eye tracking and Appendix B presents an algorithm for use when both horizontal and vertical components of eye movements are being monitored.

The use of algorithms such as these requires that the subject be engaged in some sort of calibration task that yields a set of EMM signal values that correspond to a set of known stimulus locations. The algorithm for mapping EMM signals to stimulus locations (which will be referred to here as a mapping algorithm) simply provides a means of interpolating between these known points to assign stimulus locations to other EMM signal values. The calibration task used here as an example is to have the subject look directly at each of a series of points and, while looking at each, press a button. This causes the computer to sample the EMM signal value corresponding to each stimulus location and to store the value in a table, referred to as the calibration table. Following the calibration task, this table of numbers is used by the mapping algorithm. Other tasks can be used, of course, and this may change the nature of the algorithm used for mapping (e.g., see O'Regan, 1978). The nature of the calibration task and of the resulting calibration table should be reported.

Note that linear interpolation approaches of the type described in these appendices make two strong assumptions. First, they assume complete repeatability of the EMM values obtained during the calibration task. Second, they assume that, within each stimulus region bounded by adjacent points used in the calibration task, the distances between real fixation locations and the differences between the EMM values corresponding to each of these locations are linearly related. To the extent that these assumptions are violated, the accuracy of the data, in terms of where the eyes are being directed in the stimulus or the absolute lengths of saccades, is brought into question.

Some suggestions can be made for improvement of the accuracy of this aspect of the data. First, great care should be made in obtaining repeatable EMM values for each fixation target location during the calibration task. Subjects must often be trained to exercise care in this aspect of an experiment. One way of doing this is to consistently provide them with feedback concerning the degree of repeatability they are showing. In this way subjects can be engaged in a sort of game of improving their own performance on this task. Another technique that can be used is to have the subject fixate each target location more than once during the calibration period. Then if the EMM values obtained from the same fixation target location are not sufficiently similar, the subject can be required to fixate that location additional times until successive values are close enough to meet the

criterion set. In this way, spurious values are rejected and greater consistency is obtained. If this technique is used in an experiment, the investigator should report the criterion used for accepting EMM values during calibration.

It has been our experience that one source of spurious values during calibration arises from subjects' tendency to move their eyes away from the fixation point too quickly. If the task is to look directly at a point and press a button, subjects often initiate a saccade before the button is pressed. This tendency can be greatly reduced by having only a single fixation target available at any one time. After each EMM sample is taken, the target is then moved to a new location. In addition, a tendency for the subjects to anticipate the move of the target, again making saccades prior to pressing the button, can be reduced by leaving the target in its present location for about 500 msec after the button is pressed and only then moving it to its next location.

Given that reliable EMM values are recorded during the calibration task, there is still the problem of dealing with nonlinearity in the EMM signal. The presence of nonlinearity, when using a linear interpolation mapping approach like those presented in Appendices A and B, has the effect of producing error in the accuracy of mapping from EMM values to stimulus locations in those regions between the fixation target locations used during calibration. An approach to assessing the amount of this error in a given experimental situation is described in the next section. The amount of error can be reduced by using more fixation target locations during calibration and by concentrating the density of these locations in the regions of greatest nonlinearity.

The other approach to dealing with nonlinearity is to abandon the use of linear interpolation techniques. It is hoped that those researchers using curvilinear interpolation techniques for mapping from EMM values to stimulus locations will be encouraged to describe these techniques in print. O'Regan's (1978) smooth pursuit approach avoids all interpolation, given that movement in only a single dimension is being recorded. Having alternative approaches available will provide new investigators with a selection from which to choose the most appropriate for their purposes, given the constraints of their research (accuracy requirements, time or computer space limitations, etc.).

ACCURACY OF THE EYE POSITION DATA

There is often some confusion about the meaning of accuracy when speaking about eye movement data. O'Regan (Note 1) has suggested distinguishing between relative accuracy and absolute accuracy. Relative accuracy refers to the resolution or sensitivity of the EMM equipment; that is, how small a displacement of eye position can be reliably detected? Absolute accuracy refers to the ability of the system to determine the orientation of the eyes with respect to locations in the visual field. EMM equipment can have very high relative accuracy and yet be poor in absolute accuracy for a number of reasons. It is important that comments on accuracy indicate which type is being discussed. In the present context, the term accuracy refers strictly to absolute accuracy.

Sources of inaccuracy in eye position data can be grouped in three categories. First is error that reduces short-term repeatability of the eye movement signal. This includes noise in the EMM signal, inability of subjects to reposition their eyes accurately, and so on, and hence leads to variation in eye position values when the person is asked to successively fixate the same point. Second is error introduced in mapping from EMM values to stimulus position. This primarily results from using an algorithm that is inadequate to deal with nonlinearity in the calibration matrix. Third is error that develops over time during the experimental task and might be called longer term repeatability. Due to head movement, electronic drift, or other factors, calibration values obtained prior to the task may differ from those taken following the task.

Degree of Short-Term Repeatability

The ideal eye movement monitoring situation would be one in which the EMM signal returned to exactly the same value every time a subject was asked to look directly at the same location. That is an ideal that is not reached for a number of reasons. Some of the reasons were dealt with in a prior section: EMM signal noise and drift. However, other reasons could include varying lighting conditions in the experimental room, head movement, pupil size changes (which may result from changes in the amount of light emanating from a CRT display), changes in eyelid position (especially when eyelashes intrude into the sensed region), amount of fluid on the eye's surface (which may vary with the time elapsed since the last blink or with whether or not an air conditioner in the room is on at a given moment), various types of problems in the dynamic operation of the EMM itself, and lack of consistency in the position of the subject's eyes when asked to look directly at the same location. Thus an indication of the amount of variance in EMM signal values obtained when the subject looks repeatedly at the same point gives a general summary of the quality of the entire eye movement monitoring situation.

For a one-dimensional eye tracking situation, this can be done by conducting a task such as the calibration task described earlier, in which three to five fixation points are displayed at equal distances apart, with the extreme points being at the outside edge of the stimulus region within which eye movement monitoring occurs in the experimental situation. The subject is then asked to look directly at each point in succession. If a CRT is used as the display device, a target (say, a dot with a box around it) can be made to appear successively at each of these

points in succession or in some random order. The subject is asked to look directly at each dot in each location and press a button. The EMM signal value should be obtained corresponding to the time of each buttonpress (given that the eyes are in a fixation). If the signal is quite noisy, an average over several EMM values following the buttonpress should be obtained to indicate the EMM signal obtained when the eyes are directed to that point. This is done repeatedly until the subject has looked at each point, say, 10 times. Each successive EMM value can be subtracted from the previously obtained value corresponding to that point to yield a difference score. The standard deviation of the distribution of these error scores can be obtained. This standard deviation then becomes an indication of the degree of short-term repeatability of the data. Furthermore, if the standard deviation is then divided by the Tinker value, the index of repeatability is transformed to an L-unit scale.

There are three added complexities. First, different subjects may show different degrees of variability in such a measure of repeatability, since the measure depends on their ability to adjust their eyes to the same position when looking at the same location and, with some equipment, on their ability to keep their head motionless. Thus it may be best to have a range of standard deviations obtained from several subjects. Second, the amount of variability may be different at different regions in the visual field. Often, the EMM values obtained when a subject is looking to the outer areas of the region within which the eyes are being monitored tend to be less stable than those obtained when the subject is looking at the more central areas. Thus, some indication of the range of standard deviations obtained from different areas in the visual field should be indicated if there is substantial variability. Also, the experimenter should report any patterns observable (e.g., a tendency to have less repeatability in particular regions). Third, the task as described may not tap some sources of variability present in a given EMM system. For instance, if pupil-size changes affect the indication of eye position, then the eye position recorded may depend partially on the amount of light coming from a CRT display being viewed by the subject. This could occur in a reading experiment if one page of text were shorter than another, thus reducing the total illumination coming from the CRT. The effect of this variable could be assessed by having the subject look repeatedly at the same set of points, as indicated earlier, but also adding and deleting extraneous material on the CRT to change the total illumination available at different times. The effects of some other possible variables can be assessed in the same way.

If the eyes are being monitored over a two-dimensional area, standard deviations should be calculated for both horizontal and vertical measures of eye position separately. Furthermore, this process should be repeated with

the row of fixation points occurring at three to five different vertical locations, in order to test repeatability over the entire area within which eye movement monitoring is taking place.

While this measure of repeatability gives some indication of the total system performance, it is of particular interest in dealing with data when the calibration procedure used in the actual research is similar to that described above. Information about repeatability provides one indication of the degree of faith one should have in the accuracy of the mapped data values and whether one can have more faith in the accuracy of data coming from certain regions of the display area than others. Other indications will be described later.

A second way of providing repeatability data is to collect the data during the experiment itself. In this approach, subjects are asked to look at each fixation target twice or more in the calibration prior to the experimental task and then twice or more immediately following the task, thus yielding at least two pairs of EMM values for each fixation target location. Difference scores are then computed by subtracting the first of each pair of successive values from the second. The standard deviation of the distribution of these differences can then be reported, divided by the Tinker value to convert to the L-unit scale, as described earlier.

Accuracy of the Mapping Function

The second of these sources of inaccuracy results from the mapping algorithm; these should also be assessed and reported. This can be done in the following manner. First, a calibration task is used in which the subject is asked to look directly at a series of points and press a button, with the computer sampling the EMM value corresponding to each stimulus location. This series of points should include those locations used in calibration in the normal experimental task, plus points halfway between each of these, which we will refer to as midpoints. Second, the mapping algorithm should then be used to assign stimulus locations to each of the midpoints, using only the calibration data corresponding to those points normally used in calibration in the experiment. Third, the location of each of these assigned stimulus locations should be subtracted from the actual locations of the corresponding midpoints to produce error scores. The distribution of these error scores then indicates the degree of combined error from the lack of short-term repeatability plus inaccuracy in the mapping. This can also be accompanied by some indication of the degree and nature of the nonlinearity typically found in the calibration table, so the reader can have some impression of the types of distortions with which the mapping algorithm was faced.

Degree of Longer Term Repeatability

The third source of inaccuracy has to do with those factors that can change over the period that data are

being collected during an experiment, including head movement, electronic drift, and so on. The degree of inaccuracy from these sources can be observed by engaging subjects in the calibration task both before and after data collection and comparing the calibration tables obtained by subtracting corresponding values from the pairs of tables. This yields a distribution of error scores reflecting both short-term and longer term repeatability. The mean and standard deviation of this distribution should be reported.

Often, this third source of error is the greatest contributor to total inaccuracy in the data, resulting primarily from the effects of head movement. If it can be demonstrated that the degree of inaccuracy resulting from the first two sources is relatively small, then it is possible to obtain an index for each data value that indicates its degree of inaccuracy due to this third source and its level of accuracy in general. Such an index can be particularly useful in reporting the level of accuracy of data for a particular experiment or in selecting only those data that show an acceptable level of accuracy required for the experiment being conducted. In order to obtain this measure, it is first necessary to perform a calibration task both before and after the experimental task. In this way, two sets of EMM values are obtained that correspond to particular stimulus locations, one prior to the experimental task and one following it.

Second, the assumption is made that during the task used in the experiment the EMM signal values associated with any given stimulus point range between those that would be assigned by the calibration table values obtained before the task and those which would be assigned by the calibration table values obtained after the task.

While this assumption is undoubtedly violated at times (e.g., the subject's head may move in one direction and then return before the end of the task, or drift in the signal may proceed in one direction and then return), we do not have direct evidence of such events, and they will be assumed to occur sufficiently rarely as to permit their being ignored.

Given this assumption, an accuracy indicator index can be obtained for any given data value. To do this, it is first recognized that three different sets of calibration values can be used to map a given data value onto a stimulus location: the values obtained prior to the experimental task, those obtained following it, and an average of these two. Using these, an EMM data value can be assigned three different stimulus locations, using some mapping algorithm, such as those presented in Appendices A and B. Since we do not know which of these locations is the most accurate (i.e., which corresponds most closely to the "true" position of the eyes at that time), the location obtained by using the averaged calibration data will be taken to indicate the best guess. However, taking the absolute value of the difference between the other two locations indicates the range of uncertainty of the location corresponding to this data point. Since the location obtained from the averaged calibration data is halfway between the other two locations, a simple indicator of data accuracy is computed by dividing the range of uncertainty by 2. It should be noted that this same value would be obtained by taking the absolute value of the difference between the location assigned by the averaged calibration data and either of the other locations. Thus, it is not necessary to compute all these locations. This index, which will be referred to as the index of accuracy (IA), indicates that the stimulus location assigned to that EMM data value by using the averaged calibration values may be off in either the positive or negative direction by an amount indicated by the index. Thus, if the three locations assigned to an EMM data value of 2,037 are 45.7, 47.2, and 48.7, we would take 47.2 to be the location of the eye (i.e., the eye is oriented to a location .20 of the distance from the left boundary of the 47th L-area to the right boundary of that L-area). However, we would also indicate that this location may be off by as much as 1.5 L-units to the left or right. If the experiment requires accuracy of 1 L-unit or less, this data point would be excluded as not having the needed level of accuracy.

Of course, the same procedure can be followed for calculating the IA on the vertical dimension for any data value. In two-dimensional eye tracking, a data point may be rejected because of failure to obtain sufficient accuracy on either of the two dimensions.

A formula is presented in Appendix C for directly calculating the IA for any data value when dealing with linear interpolation with unidimensional eye tracking, without having to calculate multiple stimulus locations for each data value. With more complex mapping functions, it will often be necessary to calculate the IA in the manner described above.

Note that in packaged EMM systems that do not make calibration information available to the researcher but simply use it internally to map eye position data onto the stimulus space, it is not possible to obtain such an index. It is particularly important that accuracy be carefully assessed with these systems, using techniques similar to those described earlier, since inaccuracies are often not readily apparent in data normally collected for experiments.

CONCLUSION

The present paper attempts to encourage standards in the report of psychological research involving eye movement data. It argues that it is not appropriate to adopt standards concerning what is acceptable data; that varies with the nature of the questions being studied. However, it is appropriate to list the information that should be reported by researchers, so that others can judge the adequacy of their data. Thus, this is an argument for standards in the report of data, rather than standards concerning the data themselves.

If investigators who are engaged in eye movement research will use these suggestions to make a rather complete report of the quality of eye movement data obtained in their research, there should be several desirable results. First, other investigators will have a basis for judging the adequacy of the data reported in an experiment, given the nature of the question being investigated. Second, individual investigators will begin to have standards in the published literature against which they can judge the adequacy of their own data. Third, this is likely to put pressure on both investigators and equipment manufacturers to increase the data quality of their EMM equipment.

In addition, it is our hope that these suggestions will provide the beginnings of a vocabulary for discussing the quality of data being obtained in this burgeoning research area.

REFERENCE NOTE

1. O'Regan, J. K. Personal communication, 1980.

REFERENCES

ALPERN, M. Effector mechanisms in vision. In J. W. Kling & L. A. Riggs (Eds.), *Woodworth and Schlosberg's experimental psychology* (3rd ed.). New York: Holt, Rinehart, & Winston, 1971.

LEVY-SCHOEN, A., & O'REGAN, J. K. The control of eye movements in reading. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of visible language* (Vol. 1). New York: Plenum, 1979.

McCONKIE, G. W., ZOLA, D., WOLVERTON, G. S., & BURNS, D. D. Eye movement contingent display control in studying reading. *Behavior Research Methods & Instrumentation*, 1978, 10, 154-165.

MONTY, R. A., & SENDERS, J. W. (Eds.). *Eye movements and psychological processes.* New York: Erlbaum, 1976.

O'REGAN, J. K. A new horizontal eye movement calibration method: Subject controlled "smooth pursuit" and "zero drift." *Behavior Research Methods & Instrumentation*, 1978, 10, 393-397.

RAYNER, K. Eye movements in reading and information processing. *Psychological Bulletin*, 1978, 85, 618-660.

YOUNG, L. R., & SHEENA, D. Survey of eye movement recording methods. *Behavior Research Methods & Instrumentation*, 1975, 7, 394-429.

APPENDIX A
LINEAR INTERPOLATION APPROACH TO MAPPING AN EMM VALUE ONTO A STIMULUS LOCATION IN ONE DIMENSION

It is assumed that a calibration task has been performed that provides EMM values that correspond to certain known points on a single dimension in the stimulus array. The location of these stimulus points is given in a vector L. The location of each of these points is given on a scale of L-units, in which the left boundary of the ith L-area has a location value of i. Thus a point at the center of the fifth L-area from the left of the display has a location of 5.5. This type of scale allows for easier computation with the data later, since taking the integer value of any location, without rounding, indicates the L-area within which that point lies.

The EMM values corresponding to the locations in L are contained in Vector E. Thus, E_i contains the EMM value corresponding to Stimulus Location L_i.

In mapping, or translating, a given EMM data value D to a stimulus position S, it is first necessary to locate the last value in E that is equal to or smaller than D. This value will be labeled E_m, and $E_m \leqslant D < E_{m+1}$. This also indicates that $L_m \leqslant S < L_{m+1}$. The value of S can be obtained by using the following common linear interpolation formula:

$$S = L_m + \frac{D - E_m}{E_{m+1} - E_m} (L_{m+1} - L_m).$$

At times, an EMM value may be obtained that falls outside the bounds given by E_1 and E_n, where n indicates the number of entries in E and L. When $D < E_1$, the interpolation can be successfully carried out with m = 1, and when $D > E_n$, the interpolation can be carried out with m = n - 1. Of course, the accuracy of the resulting S locations becomes more suspect the farther they fall outside the region within which calibration data were obtained.

APPENDIX B
LINEAR INTERPOLATION APPROACH TO MAPPING AN EMM VALUE PAIR ONTO A STIMULUS LOCATION IN TWO DIMENSIONS

It is assumed that a calibration task has been performed that yields a set of EMM values that correspond to certain known points in the stimulus array. For simplicity, we will assume that these locations form a grid over the stimulus, being arranged in regular columns and rows. As in Appendix A, the locations of these columns and rows are given in L-units. A column of stimulus points at the leftmost boundary of the ith column of L-areas is given a horizontal location of i; a row of stimulus points at the bottom boundary of the ith row of L-areas (counting from the bottom) is given a vertical location of i. Thus a point at the center of the bottom left L-area has a horizontal location of 1.5 and a vertical location of 1.5.

The horizontal and vertical locations of each of the points for which EMM values are known are assumed to be contained in two vectors: LV, which contains the vertical location of each of these points, and LH, which contains the horizontal location. LV contains r values, the number of rows on which calibration values were obtained. LH contains c values, the number of columns in the calibration task.

The horizontal EMM values associated with each of these stimulus locations are assumed to be contained in a matrix, EH, having r rows and c columns. A second matrix of the same size, EV, contains the vertical EMM values associated with each stimulus location. Thus, the horizontal and vertical EMM values corresponding to the jth calibration point in the ith row are contained in EH_{ij} and EV_{ij}. These values can be used to plot each calibration point in EMM value space, as shown in Figure B1. Here, the scale on the X-axis is the horizontal EMM values, and the scale on the Y-axis is the vertical EMM values. It can be seen that while the original calibration stimulus locations may have been arranged in a rectangular grid pattern, the corresponding locations in the EMM space may not be. The fictitious data shown in Figure B1 are highly nonlinear.

Figure B1 also shows a particular EMM data point D, having vertical position D_v and horizontal position D_h, for which a corresponding stimulus location S, having vertical position S_v and horizontal S_h, is desired. An algorithm for mapping D onto S, using a linear interpolation approach, will now be described.

First, it is necessary to determine which region of the EMM value space shown in Figure B1 contains the location D. This region is shaded in the figure and can be found by using a step-
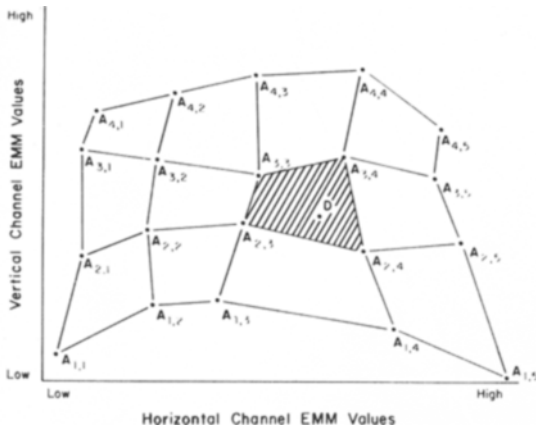
Figure B1. Plotting of hypothetical EMM data from a calibration table obtained by having a subject look directly at 20 stimulus points arranged rectangularly in four rows of five points each. A highly nonlinear pattern is shown to illustrate the types of nonlinearity that can occur. The shaded region corresponds to the area shown in Figure B2.
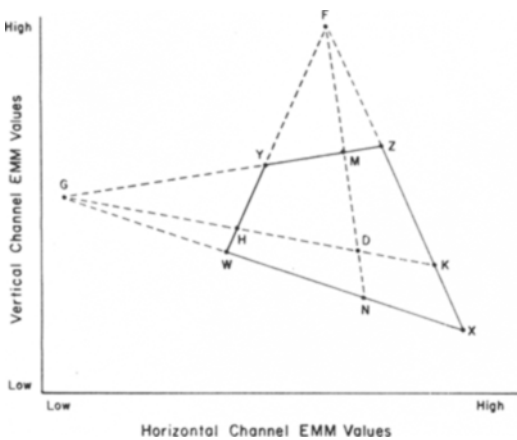

Figure B2. Plotting of shaded region of Figure B1, showing basis for mapping a data point D onto the stimulus region, using the two-dimensional linear interpolation approach described in Appendix B.

ping algorithm. In order to use this algorithm, it is necessary to calculate the slope and intercept for each line segment shown in Figure B1 connecting two successive calibration points in either the horizontal or vertical direction.

The stepping process begins at the lower left corner of the pattern, at $A_{1,1}$, and the first step is in the horizontal direction to $A_{1,2}$. At that point, we ask whether the data point D lies above, on, or below the line passing through $A_{1,1}$ and $A_{1,2}$. If it lies above, i should be incremented for the next step; if it lies on or below, j should be incremented. In this case, it lies above, and the next step goes to $A_{2,2}$. Again, the data point D is compared with the line just traversed ($A_{1,2}$; $A_{2,2}$), this time to determine whether D lies to the right of, on, or to the left of that line. If to the left, i is decremented (if possible); if on or

to the right, it is incremented. Thus, on each step, the data value D is compared with a line passing through the last arc traversed, and the next step is in the direction of the data point from the line; that is, i or j is either incremented or decremented appropriately. When movement in that direction is impossible (as when D lies below the bottom line of the calibration pattern), movement continues in the same direction as the last step. At corners, movement goes in the only direction possible. Moving back to the immediately preceding step is not permitted.

At the same time, a history of the points visited in this stepping is kept. When the algorithm results in a return to a point previously visited, this point and the prior three points visited will be found to specify the region within which D lies, or which should be used for mapping when D lies outside the calibration area. The stimulus location of the calibration point at the lower left corner of this area will be labeled LV_m, LH_n for the remainder of this section. It has a corresponding EMM space location of $EH_{m,n}$, $EV_{m,n}$. The four lines bounding this area, the four points defining those lines, and the four corresponding points in the calibration stimulus array are used to map D onto a stimulus location S.

In actual practice, while it is necessary to use the stepping algorithm just described to find the location of the first EMM data value, each successive value can typically be properly located by testing whether this new value has crossed the boundary of the region containing the last value, in the direction it has moved from the prior value.

Once the region within which D is located has been found, mapping to a stimulus location S proceeds by linear interpolation, as shown in Figure B2 and described in the steps described below. For ease of communication, each point has been given a single letter designation. It is assumed that the slope and intercept of the lines bounding the region, WX, WY, XZ, and YZ, have already been calculated and stored in a table.

(1) Find F, the point at which WY and XZ intersect. If WY and XZ are essentially parallel, flag F instead. (2) Find G, the point at which WX and YZ intersect. If WX and YZ are essentially parallel, flag G instead. (3) Find the slope and intercept of Line FD. If F is flagged, take the average of the slopes and intercepts of WY and XZ instead. (4) Find M, the point at which FD intersects YZ. (5) Find N, the point at which FD intersects WX. (6) Find P_v, the distance from N to D, as a proportion of the total distance from N to M. (7) Vertical position of S is given by $S_v = LV_m + P_v(LV_{m+1} - LV_m)$, where S_v is the vertical position of the fixation in the stimulus space, measured in L-units from the bottom row of L-areas in the stimulus space. LV_m = the vertical location, in L-units, of the fixation targets corresponding to Points W and X in Figure B2. LV_{m+1} = the vertical location, in L-units, of the fixation targets corresponding to Points Y and Z in Figure B2. (8) Find slope and intercept of Line GD. If G is flagged, take the average of the slopes and intercepts of WX and YZ instead. (9) Find H, the point at which GD intersects WY. (10) Find K, the point at which GD intersects XZ. (11) Find P_h, the distance from H to D, as a proportion of the total distance from H to K. (12) Horizontal position of S is given by $S_h = LH_n + P_h(LH_{n+1} - LH_n)$, where S_h is the horizontal position of the fixation in the stimulus space, measured in L-units, from the leftmost column of L-areas in the stimulus space. LH_n = the horizontal location, in L-units, of the fixation targets corresponding to Points W and Y in Figure B2. LH_{n+1} = the horizontal location, in L-units, of the fixation targets corresponding to Points X and Z in Figure B2.

APPENDIX C
CALCULATING THE ACCURACY INDICATOR (AI) FOR AN EMM VALUE

It is assumed that a translation matrix has been obtained both before and following the task by getting EMM values

resulting from looking at certain stimulus locations. For the present, we will deal with eye movement monitoring on a single dimension, assumed to be horizontal. Thus, the calibration matrix in this case will be in the form of a vector.

First, there is a vector L, containing values indicating the location of the points in the stimulus for which calibration information is obtained. L will contain as many values as there are points on this dimension for which corresponding EMM values were obtained. The values in L will be in L-units.

Next, vectors of EMM values corresponding to each of these stimulus locations are defined. Vector EA, with values EA_1 ... EA_i ... EA_c, where there are c stimulus locations used in calibration, is the vector of calibration values obtained before the experimental task. Vector EC, with a similar number of cells, contains the calibration information obtained after some intervening task. A vector EB is obtained by averaging the corresponding values of EA and EC, and another vector, ED, is obtained by subtracting each value of EA from the corresponding value in EC. Thus ED is a difference vector.

Finally, there is an EMM value, D, for which we wish to obtain an accuracy index (AI).

The first step is to find the last value of EB that is smaller than D. This value will be referred to as EB_i. Hence, $EB_i \leqslant D < EB_{i+1}$.

Next, the following formulas are used to calculate AI for S: $Q = (EB_{i+1} - EB_i)/(L_{i+1} - L_i)$, where Q is the number of EMM values corresponding to a single L-unit in this region of the stimulus pattern, or the local Tinker value. $R = (S - EB_i)/(EB_{i+1} -EB_i)$. $W = R (ED_{i+1} - ED_i)$. $AI = (ED_i + W)/2Q$.

The result indicates that the translated value S corresponding to Data Value D is accurate to plus or minus AI L-units, if short-term repeatability is high and the mapping algorithm yields minimal error.

If two-dimensional eye tracking is being carried out, a similar technique may be employed to yield AI values for both horizontal and vertical components. In this case, however, it is necessary to think of the EMM data space as being divided into quadrangles, with four corners defined by data values corresponding to the four points used in the calibration task. An EMM data pair (horizontal and vertical values) must then be located as being within one of these quadrangles. From there, two stimulus locations can be obtained using before and after calibration information, as above, and their distance apart can be found. These distances on the horizontal and vertical dimensions are each then divided by 2, and these products are divided by appropriate scaling values to yield a plus or minus AI value, indicating the accuracy of that data point in horizontal and vertical L-units.