Cluster: A program for the identification of eye-fixation-cluster characteristics

RAMAKRISHNA S. PILLALAMARRI, B. DIANE BARNETTE,
DEBORAH BIRKMIRE, and ROBERT KARSH

U.S. Army Research Laboratory, Aberdeen Proving Ground, Maryland

Innovations in software for the analysis of eye movements have not kept pace with the development of hardware for collecting samples of eye position (Scinto & Barnette, 1986). Eye fixation and duration have been the primary measures for gleanings of knowledge of subjects' performance while the subjects are engaged in cognitive visual tasks. The program Cluster was developed as a means of investigating the dynamics of target examination characteristics that do not lend themselves to traditional methods of eye-movement analysis. This tool has proved to be a valuable means of assessing visual activity at a micro level, in comparison with the gross measures of distribution of visual attention in various areas of the visual field. In this report, we describe the history surrounding the development of Cluster as an analytical tool, the source of input required for its execution, the mechanics of the execution as an interactive process, the replicability of raters' judgments, the program's products of visual displays and data file output, and the potential application of such a tool for analyzing visual activity.

Scientific research does not always adhere to experimental protocol in a simple, straightforward manner. In the statistical analysis of the data obtained from even a well-designed research study, unexpected aspects of human behavior and performance are often exposed. In the present paper, we describe a computer program designed and written to handle just such a situation, which arose during a research endeavor (Birkmire, Karsh, Barnette, Pillalamarri, & Breitenbach, 1991) of the U.S. Army Human Engineering Laboratory (HEL) to aid designers of military tracked vehicles at the U.S. Army Tank-Automotive Command (TACOM). Specifically, these designers were interested in determining what vehicle design features (e.g., gun barrels, turrets, etc.) were more or less critical in the identification of a vehicle viewed through an electronic imaging device, such as a forward-looking infrared (FLIR) sensor.

HEL's visual performance team developed a research program to ascertain whether eye movements could be useful for the attempt to determine what vehicle features might be critical for target discrimination. We designed an experiment in which we used a method developed and reported by Gerhart, Graziano, and Carter (1983), whereby a stored image of a vehicle silhouette could be combined with a background of digitally generated Gaussian noise. The result is a quantifiable simulation of a non-specific sensor system image. The subject's task was to visually examine a series of these images to identify one of eight different targets embedded in varying levels of target signal-to-background noise ratios (SNR).

Eye-movement records were recorded at the rate of 60 Hz while each subject was engaged in the visual task. The data were subsequently processed off line into sequential eye-fixation records (Karsh & Breitenbach, 1983) characterized by fixation spatial location coordinates and time durations. Literal graphic representations of these records, called fixation scan paths, were scaled and superimposed over the respective source images. An examination of these composites showed that fixations tended to be grouped into a cluster near the target as expected. However, it was determined that the apparent locations of fixation clusters were at sites some 1.7° to 3.0° distant and in varying directions from the actual target sites, even though the expected locations of such clusters would have been within 1.0° of the targets. These unexpected offset distances made it impossible for traditional mapping techniques to be used to specifically align fixation locations with actual target features.

Various hypotheses concerning possible machine- or task-induced errors were suggested to account for the source and magnitude of such offset distances (Birkmire et al., 1991). Hypotheses related to machine-induced errors included possibilities for poor system mirror and subject head alignments and/or calibrations, and possibly an introduction of noise into the data stream. Hypotheses related to task-induced errors concerned the type of the task (i.e., search as opposed to reading or casual picture viewing). The instrument used to collect the eye movements was carefully examined, and no mechanical problems were found. In a further test of the system, eye-movement data were collected for several other tasks in addition to the target discrimination task. Subsequent examination of these data failed to show offset distances of the size and magnitude found in the original data set. Consequently,

Correspondence should be addressed to B. Diane Barnette, U.S. Army Research Laboratory, Human Research and Engineering Directorate, ATTN: AMSRL-HR-SD, Aberdeen Proving Ground, MD 21005.

the hypothesis of machine error as a major source of the offsets was rejected. The offsets appeared to be task related, and since traditional mapping techniques were not usable with these data, the program Cluster was developed to analyze the dynamics of the fixation clusters, independently of their absolute locations in space.

OBJECTIVES

An objective of this research was to study various aspects of the dynamics of the fixation clusters in an effort to discover any underlying relationship between a fixation cluster's spatial characteristics and its offset distance in comparison with the actual target characteristics. We therefore produced a computer program to identify and quantify cluster characteristics.

CLUSTER

An algorithm was developed to allow the user, given a scan path record from a target identification task, to isolate and code the fixations that were believed to be associated with the target examination part of the record, to characterize the nature of the spatial area that the resultant cluster occupied, and to measure the offset distance of the center of the cluster area from the center of the actual location of the target embedded in the noise field. A full description of the development, mechanics, execution, and potential uses of the program follows.

Method

Materials. For input, Cluster requires a file of fixation records that have already been reduced from the original data collected at the time the subject was performing

the visual identification task. Each fixation file (one per subject) contains records representing the eye-movement history for that subject's visual behavior during the presentation of 16 images or stimulus slides. Figure 1 represents data for the first of 16 slides. The initial four records in the fixation file contain summary information for those 16 slides. The information in these four records includes a subject identification code, the total number of slides for which eye-movement records exist, and the number of fixations made for each of the 16 slides presented. Following the initial summary records are 16 segments of records (one segment per slide) containing three records of slide summary information and detailed records of fixation information for each fixation made while the subject viewed the slide. Slide summary information includes the actual slide number, the number of fixations for the slide, the x and y coordinates of the actual target location, the target location code, the total exposure time, the mean fixation duration time, the mean spatial distance between fixations, the target type, the signal-to-noise ratio, the target response code, and the type of response code. Each fixation record includes the fixation number, x coordinate, y coordinate, fixation duration, and distance from the previous fixation.

Apparatus. The unprocessed, frame-by-frame real-time data were collected and stored on a Digital Equipment Corporation PDP-11/84 minicomputer, interfaced to a Tektronix 4014 cathode-ray tube (CRT) terminal, running under the RT-11 operating system. Only the initial levels of processing to create subject fixation files were completed on this system. The fixation files were downloaded, via ASCII file transfer in a communications software package, to a Zenith Data Systems ZWX-248 microcomputer for subsequent levels of processing. This

1100	16	8	6	2 ---	Subj.; No. of slides; No. of fixations on slides 1-3
10	10	6	13	3 ---	No. of fixations on slides 4-8
7	10	10	9	9 ---	No. of fixations on slides 9-13
7	8	8		---	No. of fixations on slides 14-16
60	8	16	101	22 ---	Image no.; no. of fixations; target x,y; target location code
303	0.573	63	5	12 ---	Exposure time; mean fixation duration; mean spatial distance; target type; signal-to-noise ratio
3	1			---	Target response code; type of response
1	27	17	16	0 ---	Fixation no.; fixation x,y; fixation duration; distance from previous fixation
2	72	88	13	84	
3	88	102	62	21	
4	56	102	22	32	
5	100	89	37	45	
6	55	94	31	45	
7	19	-3	9	103	
8	77	96	85	114	

Figure 1. Sample of fixation file record format. (Sample includes fixation information for Subject 10, Slide 1.)

transfer enabled us to take advantage of graphics capabilities available on personal computers that typically do not exist on larger systems with CRT terminals. The advent of personal computers, with their capability for graphical display, makes the development of software like Cluster possible. The Zenith system is configured with the MS-DOS operating system, dual internal floppy disk drives, a 20-MB internal hard drive, a red-green-blue color monitor, and a Hewlett-Packard LaserJet printer. Cluster was written with a Microsoft GW-BASIC (Version 3.2) software package.

Design. In general, a scan path record of a subject's eye-movement history derived from a visual identification task, in which the target is visible but embedded in varying levels of Gaussian noise, typically consists of three types of looking. Some initial fixations relate to the eye's moving toward the target site. There are some brief orienting fixations in which the eye stabilizes at or near the target location, after which fixations accumulate or cluster at the target site where examining for target identification takes place. A scan path depicting such eye-movement behavior and the target area is shown in Figure 2. A zoomed view of the fixation cluster and target area is also shown. Lines are drawn between fixation vectors to preserve for the viewer the order of fixation occurrence. The allocation of which fixations go with the target examination phase of the task is not well defined and cannot be modeled as a deterministic process. Subjective judgments are made about factors of spatial and temporal relatedness of fixation coordinates constituting a cluster, as well as about the proximity and direction of that cluster from the target site. Another dimension of the subjective judgment of relatedness of fixations within a

cluster is fixation duration. Fixations associated with target search have shorter durations (Scanlan & Agin, 1978) than do fixations associated with the kind of examining that creates clusters at target sites. Although some criteria can be defined and possibly formulated with some difficulty as a set of rules determining the inclusion or exclusion of fixations that may be functionally related, subjective assessments take many factors into account that are not easily articulated or easily formulated into such rules. The formulation of deterministic rules requires that one establish thresholds of event characteristics that can be used to classify the event. Many events, however, have characteristics whose thresholds fall on the decision boundaries. Formulating decisions that use fuzzy rules works best in environments with such phenomena, but the problem of determining the amount of fuzziness and its nature still has to be specified deterministically. Thus, a decision was made against developing an automated, deterministic process for classifying a fixation as being included in or excluded from the target identification cluster.

Cluster's user interface was written with the desire to establish a simple and easy-to-use approach to a complex problem. The user is able to assign and/or change fixation classifications (i.e., included or excluded) at will and as many times as is necessary until he/she is satisfied that the appropriate fixations have been selected to constitute the cluster related to target examination. This feature is particularly important given that a subjective judgment about a fixation's classification may hinge on characteristics of the fixation(s) following, as well as preceding, the current point. Again, such a process would be extremely difficult if not impossible to include in a deterministic process. Once the target cluster has been defined, a number

Figure 2. Fixation scan path for Subject 10, Slide 1 (zoomed to show detail).

of parameters describing the cluster are calculated. Let us now take a closer look at the mechanics of the program Cluster.

Procedure. In the following discussion, the procedures for levels of user interaction will be considered first; then we will turn to the algorithms embedded in the program for calculating the distance, cell distribution, and convex polygon parameters.

Upon execution, Cluster first prompts the user to enter the number of the subject to be processed. This number is embedded in the name of the subject's fixation data file. A search is made within the current working directory for the fixation file for this subject; when the file has been located, it is opened, and the user is prompted to enter the number of slides to be skipped. Some studies include slides between actual stimulus slides, which require verbal responses, or perhaps the user may be interested in looking only at slides that meet certain criteria. This option allows the user to skip over slides in order to get to the ones of interest. When the records for the first slide of interest have been located, the fixation data for that slide are read and subsequently displayed on the monitor. Figure 3 is a reproduction of what is shown on the user's monitor. Only the fixation x and y coordinates and the duration are displayed. The top and the right side of Figure 3 show several variables for which values do not yet exist. At this point, the user enters a return to advance to the next stage of processing. This stage displays the subject's scan path for the current slide along with a nine-cell box displayed directly over the actual target area. The right side of the screen displays the duration data for fixation number one. The program is waiting for the user to make a decision about whether the fixation should be included in the target examination cluster. The current

fixation of interest is highlighted in the scan path so that the user is given a visual placement of the fixation in relation to other potential cluster fixations. The user presses the return key to include it or the space bar to exclude it. After the user has responded in the same manner to all the fixations that follow, he/she is prompted to indicate whether he/she chooses to change the status of any of the fixations. The user may randomly pass through the fixation list as many times as is necessary in order to feel comfortable with the selection of fixations to create the target examination cluster. When this selection phase has been completed, the user enters a return to proceed to the next level of processing. At this stage, the nine-cell target box is relocated from the actual target position and drawn directly over the center of the user-selected target examination cluster (see Figure 4).

In addition to the nine-cell target box over the cluster, a rectangular box is drawn that strictly encompasses only the smallest possible area of the target examination cluster. Subsequently, a convex polygon is drawn around the cluster's outer points. This polygon represents what one would see if a pin were stuck into each of the cluster points and a rubber band were placed around the bordering pins. Figure 5 provides a zoomed view of the cluster and target areas shown in Figure 4.

We now turn to the computer algorithms. The distance parameters are calculated on the basis of the distance between the center of the actual target position and the center of the target examination cluster. The center of the actual target position is calculated by defining a polygonal shape (or more basically a rectangle) that encompasses the outer boundaries of the vehicle silhouette, and by determining the center vector of that rectangle. The center of the target examination cluster, as defined by the fixa-

```
# OF FIXATIONS: 9      BOX LOCATION: ( 0, 0 ) CELL SIZE: 36 STEP SIZE: 10
FIXATION FILE: FIXNS.10 SLIDE: 1 / 0 TARGET CODE:      WINDOW SIZE: ±
```

READING FIXATION FILE: FIXNS.10, SLIDE:1 TO CONTINUE, PRESS ANY KEY				NUM	NFX	NDR
NUM	X-POS	Y-POS	DURATION	Area: 1		
1	27	17	16	Area: 2		
2	72	88	13	Area: 3		
3	88	102	62	Area: 4		
4	56	102	22	Area: 5		
5	100	89	37	Area: 6		
6	55	94	31			
7	19	-3	9			
8	77	96	85			

```
Target Offset: ( , )
Offset Distance:
Fixations, Inside:
Total Duration:
X-Range: to
Y-Range: to
```

Figure 3. Monitor display of fixation data for Subject 10, Slide 1.

Figure 4. Monitor display showing a nine-cell rectangle placed directly over the target examination cluster for Subject 10, Slide 1.

Figure 5. Monitor display showing a nine-cell rectangle placed directly over the target examination cluster for Subject 10, Slide 1 (zoomed to show detail).

tion points associated with the target identification task, can be calculated according to at least three different methods. The first is to calculate the mean fixation position (x,y) , where

$$(x,y) = \frac{1}{n} [\sum(x_i), \sum(y_i)], i = 1, 2, \dots, n.$$

A second method is to calculate the weighted mean fixation position, in which each fixation position is weighted by the fixation duration at the point (x,y) where

$$(x,y) = \frac{1}{\sum(d_i)} [\sum(x_i d_i), \sum(y_i d_i)], i = 1, 2, \dots, n.$$

This measure lets a fixation with a long duration have more influence on the center cluster position than a fixation with a shorter duration. Finally, a third method is to calculate the center of the convex polygon that encompasses the fixations of the target examination cluster. The center of the cluster is calculated as the weighted (by area) mean position (x,y) of the centers of the triangles that, when joined, form the convex polygon.

The following calculations are made to compute that area.

Let (x_i, y_i) , in which $i = 1, 2, \dots, m$, be the coordinates of the convex polygon. The first of the $(m-2)$ triangles will be composed of points 1,2,3 of the convex polygon. The center of the area of this triangle is

$$(X_1, Y_1) = \frac{1}{3} (x_1 + x_2 + x_3, y_1 + y_2 + y_3),$$

and the area of the triangle is

$$A_1 = 0.5 |x_1(y_2 - y_3) + x_2(y_3 - y_1) + x_3(y_1 - y_2)|$$

Thus, the area of the polygon is

$$A = A_1 + A_2 + \dots + A_{m-2},$$

and the center of the area of the polygon is

$$(x,y) = \frac{\sum A_i(x_i, y_i)}{A} \quad i = 1, 2, \dots, m-2.$$

The area of the convex polygon is always smaller than the area of the rectangle, except when the polygon and the rectangle are identical. Below, an algorithm is described that determines an m -sided convex polygon, its area, and its center for a given set of n points.

The offset distance from the center of the actual target position to both the weighted mean fixation position of the rectangular area and the center (x,y) position of the convex polygon are now calculated in the program.

The target area is divided into nine cells produced from three rows and three columns. The bordering cells of the matrix are expanded 0.5° to account for fixation over- and undershoot to the feature visually examined. This nine-cell matrix is then overlaid onto the center of the target examination cluster, and a new monitor display is created to show this. Cluster then calculates the distribu-

tion of fixations and their durations of the target examination cluster into the nine cells. The frequency count of fixations and the sum of their durations for each cell are then calculated as both an absolute measure of visual attention to specific target areas and a proportion of the total visual looking time spent on the target.

The algorithm that defines the convex polygon takes the given set of arbitrarily ordered n points in a two-dimensional space that constitute the target examination cluster and finds an m -sided ($m \leq n$) convex polygon that includes all the points. The set (p,q) is said to be convex if and only if any point r , in which

$$r = \beta p + (1-\beta)q,$$

is a member of the set, for

$$0 \leq \beta \leq 1.$$

Generalizing this, a convex polygon can be mathematically defined as the geometrical area that always includes the (x,y) pair,

$$(x,y) = [\beta x_i + (1-\beta)x_j, \beta y_i + (1-\beta)y_j]$$

where (x_i, y_i) , (x_j, y_j) are two points in the fixation cluster, and β is a real number, $0 \leq \beta \leq 1$.

Again the resulting geometrical shape resembles what would appear if pins were stuck in all the n points and a rubber band were placed around the entire set of n points. The following iterative algorithm determines such a convex polygonal shape.

1. Of the n points in the cluster, find the leftmost point. Alternately, the algorithm can be modified to work with the rightmost, topmost, or bottommost point. If there is a tie, any one of the competing points may be chosen. This point is the first vertex of the convex polygon.

2. With reference to a vertical line passing through this vertex, calculate the amount of angular shift as the line tilts clockwise to meet the i th point in the cluster. Let this angle be θ . Of all the $(n-1)$ such angles, the smallest angle is selected. If it belongs to the j th point in the cluster, it is labeled the second vertex of the convex polygon.

3. The first and second vertices of the convex polygon are joined, extending the line beyond the points. Using the second vertex as the pivot, the angular shifts are calculated as the line is rotated clockwise to meet each of the other $(n-2)$ points. The cluster point associated with the smallest of the $(n-2)$ angles is labeled the third vertex of the convex polygon. This process is repeated until the vertex selected corresponds to the first vertex. At this point, the polygon is complete, and m vertices, with ($m \leq n$), will have been defined.

4. The resulting polygon of m vertices is broken up into $(m-2)$ triangles. The area and center of each triangle are calculated. The center of the convex polygon is defined as the weighted mean center of all the triangles.

When Cluster has performed all these calculations, a final summary of cluster parameters is written to both an offset data disk file and the Hewlett-Packard LaserJet printer. The summary contents include the (1) horizontal

and vertical range, center, and enclosed area (by a rectangular box) of the cluster; (2) mean fixation position of the cluster; (3) weighted mean fixation position of the cluster; (4) convex polygon enclosing the cluster fixations; (5) center and area of the convex polygon; (6) distance (offset) from the center of the polygon to the center of the actual target location; and (7) distribution of fixations in a nine-cell grid area covering the area of the target.

Replicability. Since the procedure for determining the inclusion of fixations into a cluster is a subjective process, it was necessary to assess the degree of replicability of the process. Therefore, a study was conducted in which eight independent raters were shown scan paths from the target discrimination research (Birkmire et al., 1991) described previously and asked to determine which fixations belonged to the target cluster. After the purpose and procedures of this study were explained, the subjects were shown six scan paths to familiarize them with the task. While each subject viewed the first scan path, the experimenter told him/her which points to include or exclude from the cluster and stated the reasons for the decision. Each of the next five practice scan paths were shown to each subject, who either included or excluded each fixation in the target cluster and verbally gave the reason for the decision to the experimenter. Following this familiarization phase, each subject was shown 16 different scan paths from the target discrimination study and made assignments of fixations to target clusters.

To check interrater agreement for the assignment of fixations to target clusters, a coefficient of agreement (κ , or kappa) developed by Cohen (1960) was calculated between all 28 pairs of raters. The κ statistic ranged from .54 to .92, with a mean of .75 ($p < .001$). Percent agreement (number of actual agreements/number of possible agreements $\times$ 100) ranged from 79% to 97%, with a mean of 89%.

CONCLUSIONS

Although the development of Cluster may have been somewhat serendipitous, as a new tool for extracting meaning from eye-movement data it has proved to be more than useful (Birkmire et al., 1991).

Historically, advances in eye-movement research have primarily emphasized the development of techniques and equipment to record eye movements (Young & Sheena, 1975). Furthermore, the use of eye-movement records to

assess cognitive processing strategies has primarily involved fixation locations and time durations with respect to the placement of objects and text in the visual field. The program Cluster, however, takes a step beyond that level of assessment of visual attention into the visual behavior occurring during the examination stage of target identification. Specifically, given the nature of the data, with respect to the unexpected offset magnitude, traditional mapping techniques were not applicable. Cluster provided a means of assessing visual behavior in the absence of mapping target examination clusters to the absolute location of targets. Finally, the high degree of agreement among a large number of raters indicates a reliable, albeit subjective, process. Such a technique should be applicable to the assessment of any data that can be mapped in a visual two-dimensional space in which the subjective judgment of the x,y position determines the membership of a data point in a group or cluster.

Copies of the software can be obtained from B. Diane Barnette, U.S. Army Research Laboratory, Human Research and Engineering Directorate, Aberdeen Proving Ground, MD 21005 (e-mail: dbarnett@he14.brl.mil).

REFERENCES

- BIRKMIRE, D. P., KARSH, R., BARNETTE, B. D., PILLALAMARRI, R. S., & BREITENBACH, F. W. (1991). *Perceptibility of military vehicle silhouettes* (Technical Memorandum 2-91). Aberdeen Proving Ground, MD: U.S. Army Human Engineering Laboratory.
- COHEN, J. (1960). A coefficient of agreement for nominal scales. *Educational & Psychological Measurement*, 20, 37-46.
- GERHART, G. R., GRAZIANO, J. M., & CARTER, M. J. (1983, August). *Vehicle perceptibility analysis*. Paper presented at the annual meeting of the KRC Symposium on Ground Combat Signatures.
- KARSH, R., & BREITENBACH, F. W. (1983). Looking at looking: The amorphous fixation measure. In R. Groner, C. Menz, D. F. Fisher, & R. A. Monty (Eds.), *Eye movements and psychological functions: International views* (pp. 53-64). Hillsdale, NJ: Erlbaum.
- SCANLAN, L. A., & AGIN, A. K. (1978). *A behavioral model of target acquisition in realistic terrain* (Tech. Rep. P78-70R). Culver City, CA: Hughes Aircraft Co.
- SCINTO, L. F. M., & BARNETTE, B. D. (1986). An algorithm for determining clusters, pairs or singletons in eye-movement scan-path records. *Behavior Research Methods, Instruments, & Computers*, 18, 41-44.
- YOUNG, L. R., & SHEENA, D. (1975). Survey of eye movement recording methods. *Behavior Research Methods & Instrumentation*, 7, 397-429.

(Manuscript received April 24, 1992;
revision accepted for publication September 30, 1992.)