

Fixation identification: The optimum threshold for a dispersion algorithm

PIETER BLIGNAUT

University of the Free State, Bloemfontein, South Africa

It is hypothesized that the number, position, size, and duration of fixations are functions of the metric used for dispersion in a dispersion-based fixation detection algorithm, as well as of the threshold value. The sensitivity of the I-DT algorithm for the various independent variables was determined through the analysis of gaze data from chess players during a memory recall experiment. A procedure was followed in which scan paths were generated at distinct intervals in a range of threshold values for each of five different metrics of dispersion. The percentage of points of regard (PORs) used, the number of fixations returned, the spatial dispersion of PORs within fixations, and the difference between the scan paths were used as indicators to determine an optimum threshold value. It was found that a fixation radius of 1° provides a threshold that will ensure replicable results in terms of the number and position of fixations while utilizing about 90% of the gaze data captured.

1. Introduction

Cognitive processing takes place during fixations and allows the human to “see” (Just & Carpenter, 1984), but the eyes are never still. There is constant tremor of the eyes in the form of nystagmus, drifts, and microsaccades (Rayner, 1998). “A fixation . . . may be thought of as the mean X and Y position coordinates measured over a minimum period of time during which the eye does not move more than some maximum amount. More simply stated, point-of-gaze must continuously remain within a small area for some minimum time” (Eyenal, 2001, p. 12).

Irwin (1992) found the theoretical minimum duration for a single fixation to be 150 msec, whereas Manor and Gordon (2003) argued that 100 msec can also be justified. Rayner (1998) indicated that the mean duration of a single fixation may depend on the nature of the task. The mean fixation duration during silent reading is about 225 msec. During visual search, however, the mean fixation duration is 275 msec, and for tasks that require hand–eye coordination, such as typing, the mean fixation can be 400 msec (Rayner, 1998).

Saccades are rapid eye movements with velocities as high as 500°/sec used to reposition the central focus area of the eye (fovea) to a new location (Rayner, 1998). The duration of a saccade is influenced by the distance covered. For example, a 2° saccade, which is typical during reading, lasts about 30 msec, whereas a 5° saccade (typical of scene perception) lasts about 40–50 msec (Rayner, 1998).

The fixation, as the fundamental unit of eyetracking analysis, depends on both the algorithm used to separate fixations from saccades and the parameters employed for the algorithm (Shic, Chawarska, & Scassellati, 2008). Results produced by different algorithms may vary sig-

nificantly (Spakov & Miniotas, 2007). One of the greatest restrictions of the available algorithms for fixation detection is the fact that the parameter settings are crucial. It has been shown by Shic et al. (2008), as well as Manor and Gordon (2003), that a change in the parameters may cause a change in the reported fixation durations. Choices made by researchers during analysis can lead to different interpretations of the same eyetracking data (Shic et al., 2008).

In this article, the dispersion threshold algorithm for fixation identification (I-DT) of Salvucci and Goldberg (2000) is implemented, and an attempt is made to determine the optimum dispersion threshold for each of four different measures of dispersion. Various characteristics of the identified fixation sequences, also known as scan paths, are analyzed at a range of thresholds. The optimum threshold for each measure is considered to be the one at which a number of indexes (the percentage of PORs included, the number of fixations identified, the position of the fixations, and their spatial distribution) are optimized to such an extent that the position and duration of a viewer’s actual gaze is represented as accurately as possible.

It is hypothesized that all of the indexes listed above (dependent variables) are functions of the metric used for dispersion as well as of the threshold value (independent variables). Specifically, the sensitivity of the dependent variables of the I-DT algorithm for the various independent variables was analyzed.

If the hypothesis holds, it would mean that eyetracking research should always report the algorithm that was applied, together with the metric and threshold that was used, in order to facilitate the replication of results—an opinion shared by Salvucci (as quoted in Karn, 2000).

P. Blignaut, pieterb.sci@ufs.ac.za

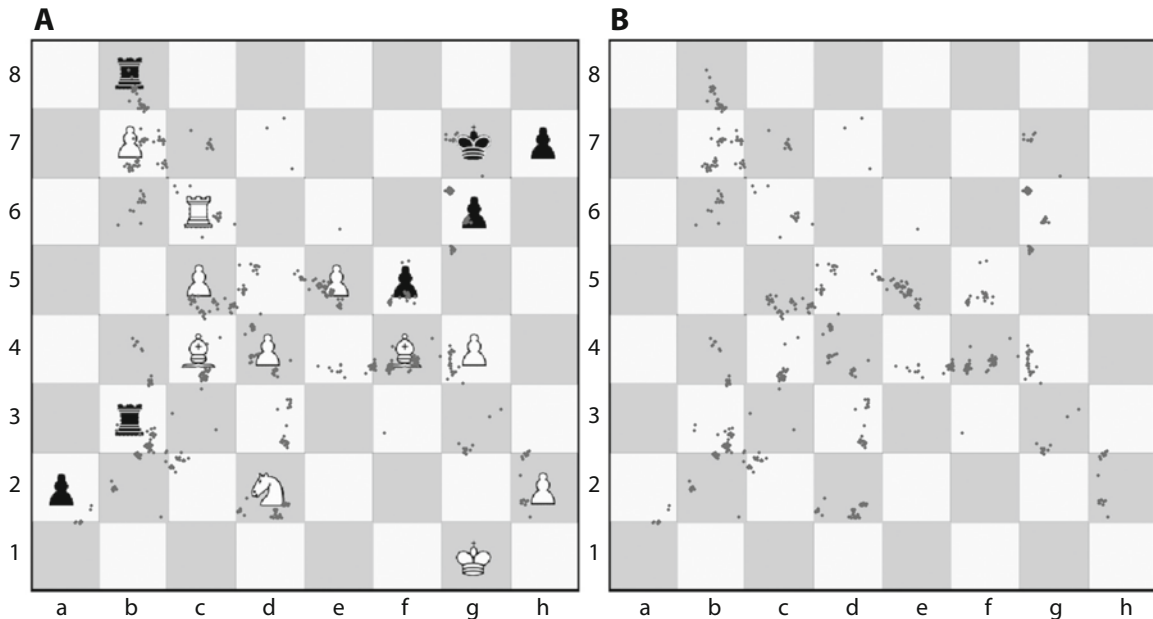


Figure 1. (A) Raw gaze data of a participant during a memory recall experiment. (B) Panel A repeated, but with the chess pieces removed.

The present article is organized as follows. Section 2 will provide some background on the identification of fixations from raw data, and Sections 3 and 4 will give an overview of the literature on existing algorithms for identifying fixations and the critical nature of the threshold values that are used in these algorithms. Section 5 will deal with the methodology used in the present study. In Section 6, the results will be analyzed with respect to the hypothesis as formulated above. The quality of the fixation sequences that are identified by the various metrics at a range of threshold values will be evaluated in Section 7. Section 8 will resolve the findings of Sections 6 and 7 in an attempt to determine an optimum dispersion threshold for the I-DT algorithm of Salvucci and Goldberg (2000). The findings will be summarized in Section 9, and the article will be concluded with some suggestions for future research in Section 10.

2. Identifying Fixations From Raw Gaze Data

The raw data generated by planar eyetracking equipment consists of tuples (x_i, y_i, t_i) where x_i and y_i refer to the so-called point-of-regard (POR)—that is, the gaze coordinates in the plane of a two-dimensional stimulus (e.g., a computer screen) at time t_i . The interval at which data are sampled varies with different eyetrackers but typically can be 8–20 msec.

Figure 1A shows the raw gaze data (individual PORs) for a chess player looking at a configuration of pieces for 15 sec, whereafter he had to reconstruct the configuration for the purposes of a memory recall experiment. Figure 1B shows the same board with the pieces removed in order to reduce cluttering and improve visibility of the PORs. It is clear that the PORs are grouped together and that these clusters could possibly be identified by a human observer

as fixations. It is, however, not always clear whether PORs that are spatially close to each other are also temporally adjacent or how PORs that are not spatially close to other PORs should be handled. Furthermore, manual inspections are extremely tedious and not feasible when working with a large data set. An algorithm is thus needed that can be automated to extract fixations from the raw gaze data in such a way that the viewer's intent (whether conscious or subconscious) is extracted as accurately as possible.

Figure 2A shows a recording of the same stimulus for another participant. The clustering is much less evident, and a human observer would have difficulty identifying the fixations as indicated by the enclosing convex hulls in Figure 2B. PORs that were regarded by the fixation detection algorithm as noise or intermittent gaze captures during saccades are not included in a fixation. It is also clear that some clusters that might have been mistaken for a single fixation by a human observer are identified by the algorithm as separate fixations, since they are not temporally adjacent. For longer recordings, the issue of overlapping fixations would become even more pronounced.

3. Overview of Existing Algorithms

Two basic conditions exist for a cluster of PORs to constitute a fixation: The total duration must be long enough, and the PORs must be spatially close enough to one another while forming a temporal sequence. These conditions can be more precisely defined in terms of a duration threshold and a distance or velocity threshold.

The algorithms that can be used to identify fixations within raw gaze data have been well summarized (Duchowski, 2007; Jacob, 1990, 1993; Salvucci & Goldberg, 2000; Shic et al., 2008; Spakov & Miniotas, 2007), and several novel improvements and alterations have been sug-

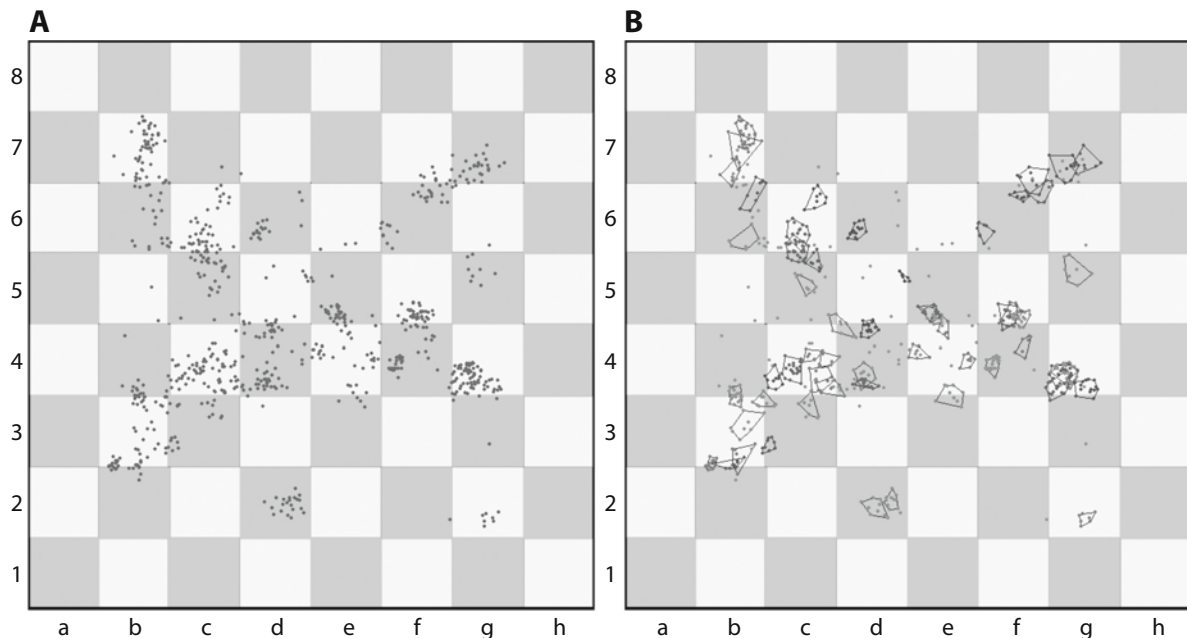


Figure 2. (A) Raw gaze data of a second participant showing the points of regard (PORs). (B) Panel A repeated, but with enclosing convex hulls that group PORs together in fixations. PORs that are not assigned to a fixation are visible outside the hulls.

gested in the past (Kumar, Klingner, Puranik, Winograd, & Paepcke, 2008; Santella & DeCarlo, 2004; Urruty, Lew, Ibadaddene, & Simovici, 2007). Also, several tools exist that employ these algorithms (Camilli, Nacchia, Terenzi, & Di Nocera, 2008; Eyenal, 2001; Gitelman, 2002; Heminghaus & Duchowski, 2006; Salvucci, 2000; Spakov & Miniotas, 2007; Tobii Technology AB, 2008).

The algorithms can be categorized roughly in terms of the way in which the above-mentioned conditions for fixations and the corresponding thresholds are handled. The velocity threshold algorithm discussed by Salvucci and Goldberg (2000) and Kumar et al. (2008) separates fixation points (PORs belonging to a fixation) and saccadic points (PORs that do not belong to a fixation) on the basis of their point-to-point velocities. The velocity of a fixation point is less than a chosen threshold value, whereas a saccadic point has a velocity that is larger than or equal to the threshold. Thereafter, consecutive fixation points are collapsed into fixations, and saccadic points are discarded. Another velocity-based algorithm is based on hidden Markov models (HMMs). The algorithm “uses a two-state HMM in which the states represent the velocity distributions for saccade and fixation points” (Salvucci & Goldberg, 2000, p. 73). It is more robust than the velocity threshold algorithm but uses a set of eight parameters—namely, two observation (mean and variance of the distribution) and two transition parameters for each of the two states. These parameters can be adjusted dynamically through a process of reestimation. The process of reestimation can, however, be difficult to implement (Salvucci & Goldberg, 2000).

The dispersion threshold algorithm was originally proposed by Widdel (1984), and adaptations and implemen-

tations thereof are discussed in Camilli et al. (2008), Salvucci and Goldberg (2000), Shic et al. (2008), and Urruty et al. (2007). The algorithm utilizes the fact that fixation points, because of their low velocity, tend to cluster close together. Fixations are identified as groups of consecutive PORs within a particular dispersion or maximum separation. The dispersion can be measured in terms of the distance between points in the fixation that are the farthest apart (Salvucci & Goldberg, 2000), the distance between any two points (Shic et al., 2008; Spakov & Miniotas, 2007), the distance between points and the center of the fixation (Camilli et al., 2008; Shic et al., 2008), the standard deviation of the x - and y -coordinates (Anliker, 1976; Eyenal, 2001), or a minimum spanning tree of the points in a fixation (Salvucci & Goldberg, 2000; Urruty et al., 2007).

Recently, two promising approaches to fixation detection have been proposed. The mean shift procedure proposed by Santella and DeCarlo (2004) searches for a local maximum in a d -dimensional space by shifting each point of the space toward higher density areas in order to separate clusters until such movements involve a small number of points. Urruty et al. (2007) proposed a clustering algorithm in which clusters are formed by projections of eye-tracking data. Urruty et al.’s approach constructs clusters in subspaces of lower dimensionality and then uses these to assist in identifying clusters in the original data set.

In this article, the focus will be on the dispersion threshold algorithm only. This algorithm is quite robust with regard to identified fixation sequences, as opposed to other algorithms; for example, velocity-based algorithms may produce inconsistent results at or near threshold values (Salvucci & Goldberg, 2000) or at slow eye movements

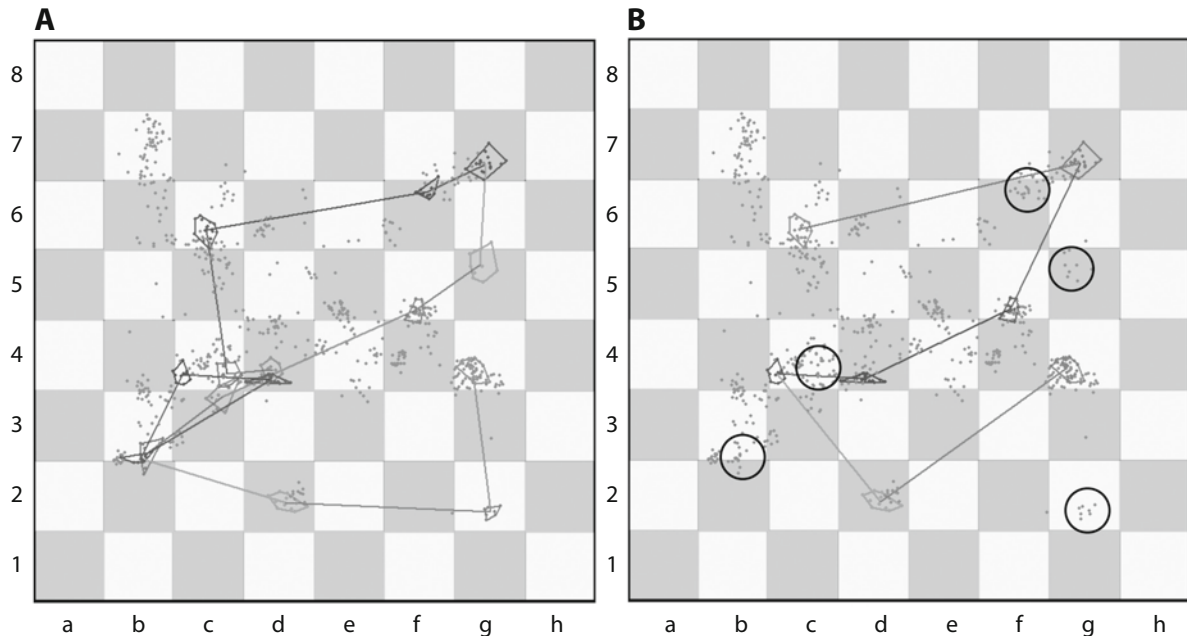


Figure 3. Scan paths of a participant as identified with the I-DT algorithm with a radius threshold of 1.0° and a duration threshold of 100 msec (A: d4, c4, c6, f6, g7, g5, f5, c3, b3, d4, c4, b3, d2, g2, g4) and 200 msec (B: c6, g7, f5, d4, c4, d2, g4). Black circles indicate differences.

(Urruty et al., 2007). Furthermore, the I-DT algorithm is simple and easy to implement, and end users have little difficulty comprehending the meaning of the parameters and relating them to published recommendations. The algorithm is also commonly used in analysis tools (Camilli et al., 2008; Gitelman, 2002; Salvucci, 2000; Tobii Technology AB, 2008). The algorithm is, however, very sensitive to parameter settings (Salvucci & Goldberg, 2000), which necessitates an effort to establish the optimum settings.

4. Criticality of Threshold Values

One of the biggest restrictions of the available algorithms for fixation detection is the fact that the parameter settings are crucial. Karsh and Breitenbach (1983) have shown that the different algorithms for detecting fixations can lead to totally divergent results. Shic et al. (2008) indicated that mean fixation duration is a linear function of the parameters chosen. Shic et al. also showed that specific findings of an eyetracking analysis can be made insignificant or even reversed by changing parameter settings. The fact that existing commercial eyetracking systems suffer from a lack of transparency regarding the algorithms they implement while having a strong dependency on the parameters entered by users (Lankford, 2000) further aggravates the situation. As was stated by Wooding (quoted in Karn, 2000), the standardization of algorithms for the extraction of fixations and saccades, in order to ensure rigor during research, is long overdue.

The threshold values for velocity and dispersion that are normally used during research are motivated by physiological characteristics. Depending on the nature of the task, it has normally been recommended that the threshold for minimum fixation duration should be 100–200 msec (Manor &

Gordon, 2003), and the dispersion threshold should include 0.5° – 1° of visual angle (Camilli et al., 2008; Eyal, 2001; Jacob & Karn, 2003; Salvucci & Goldberg, 2000).

If the duration threshold is set too low, false fixations might be identified; if it is too high, actual fixations might be missed (Camilli et al., 2008). Manor and Gordon (2003) have also shown that significantly more fixations are identified with a duration threshold of 100 msec than with a threshold of 200 msec. Figure 3 shows the raw gaze data for a participant looking at a stimulus for 15 sec. The scan path during the first 3 sec as identified with the I-DT algorithm of Salvucci and Goldberg (2000) is indicated. In Figure 3A, the duration threshold was set at 100 msec, and the maximum fixation radius at 1.0° . In Figure 3B, the same dispersion threshold was used (fixation radius 1.0°), but the duration threshold was set to 200 msec. It is clear that several clusters of PORs that a human observer might have regarded as fixations were not identified in Figure 3B—for example, c4, f6, g5, b3, and g2 (circled in Figure 3B).

If the dispersion threshold for the I-DT algorithm of Salvucci and Goldberg (2000) is too low, the algorithm might exclude fixations of people with a large amount of tremor—for example, older people or people wearing glasses. If the dispersion threshold is too high, intermediate PORs that are actually part of saccades might be mistaken as part of a fixation, or clusters of PORs that actually represent separate fixations could be merged. Figure 4 shows a scan path for the same participant as the one in Figure 3. The duration threshold was set at a fixed 100 msec, and the dispersion threshold was set to 0.5° (Figure 4A) and to 1.5° (Figure 4B). As compared with Figure 3A, several fixations were not identified in Figure 4A (e.g., c4, f6, g5), whereas the fixations for c4 and d4 were combined in Figure 4B.

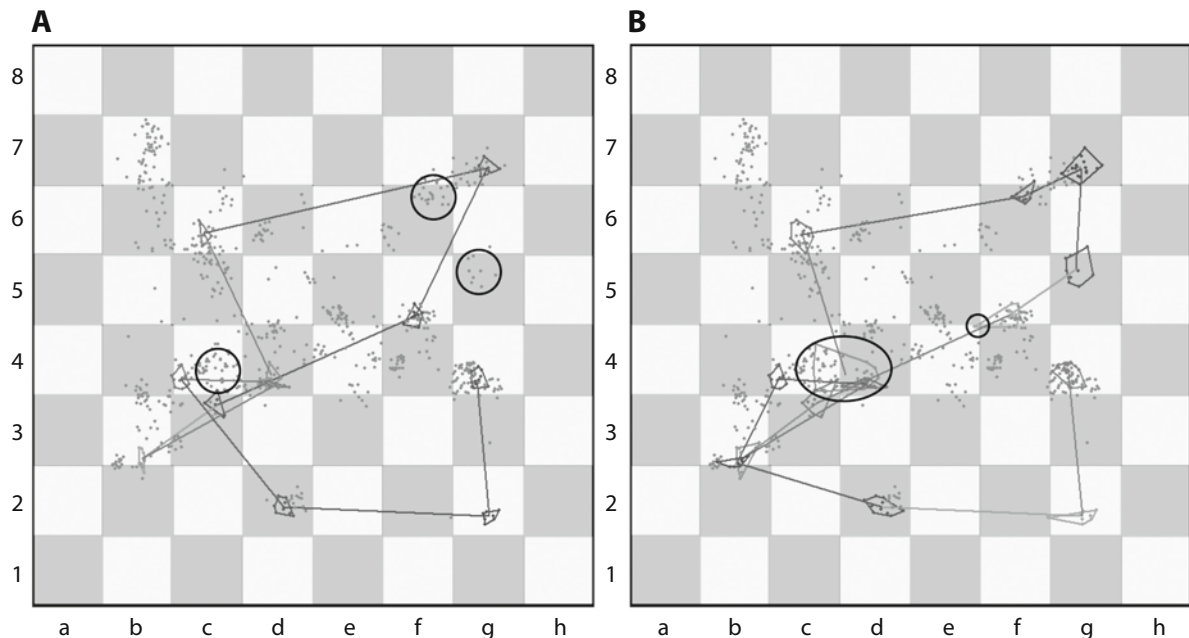


Figure 4. Scan path of a participant as identified with the I-DT algorithm with a duration threshold of 100 msec and a radius threshold of 0.5° (A) and 1.5° (B). The differences are highlighted with black circles.

Also, in Figure 4B, the saccadic point on the edge of e4 en route from f5 to c3 was included in f5.

Duchowski (2007) indicated that parameters may be determined empirically and referred also to Tole and Young (1981), who suggested an adaptive approach to overcoming the criticality of threshold values by recalculating the thresholds on the basis of recently observed noise. It might also be possible to personalize the threshold for individual users, in order to accommodate individual differences in tremor and drift (refer to Figures 1 and 2).

With reference to the hypothesis mentioned in Section 1, the aim of this study was to determine the impact of the metric and threshold used in a dispersion-based algorithm. Whereas the duration threshold was kept at a constant 100 msec in the study, an attempt was also made to determine the optimum dispersion threshold.

5. Methodology

One dispersion-based algorithm for fixation detection, the I-DT algorithm of Salvucci and Goldberg (2000), was implemented with five different measures of dispersion. Various characteristics of the fixation sequences that were identified by the algorithm were analyzed at a range of thresholds. The optimum threshold for each measure was considered to be the one at which a number of indexes (the percentage of PORs included, the number of fixations identified, the position of the fixations and their spatial distribution) were optimized to such an extent that the position and duration of a viewer's actual gaze was represented as accurately as possible.

5.1. Participants, stimuli, and experimental setup.

Thirty-one chess players were approached between rounds of a chess tournament to take part in the study. Participa-

tion was voluntary, and all the participants had normal or corrected-to-normal vision. The sample included 27 males and 4 females, with an average age of 31.0 years ($SD = 13.2$). Since no participant showed signs of bad health, trembling, or weakened vision, age was not considered to have an effect. Chess expertise varied from novice (ELO¹ rating 1,000) to expert (ELO rating 2,400), with an average ELO rating of 1,880 ($SD = 445$). Although chess expertise could have an influence on scan paths, it was not considered to have an impact on fixation identification. Therefore, the sample was considered to be homogeneous in all respects that relate to the aim of this study.

Four different board configurations or stimuli, one of which is shown in Figure 1, were displayed to each participant onscreen for a period of 15 sec each. After every 15-sec exposure, players had to reconstruct the configuration for purposes of a memory recall experiment. The recall performance of the participants is beyond the scope of this study, and only the eyetracking data that were captured during the 15-sec exposure time were analyzed.

Two of the boards represented midgame setups from actual games with 18 and 28 pieces, respectively. For purposes of the memory recall experiment, several pieces on these two boards were randomly interchanged to create the other two boards. The random interchanging of pieces meant that the overall configuration of the boards remained the same but that invalid positions might have been created. Although the validity of the boards or the number of pieces could certainly have an effect on the scan paths, especially for better players, it was not considered to have an impact on fixation identification.

5.2. Equipment. Data were captured with a Tobii 1750 eyetracker (www.tobii.com). The eyetracker has a fre-

quency of 50 Hz, which means that the PORs were captured and written to the underlying database every 20 msec. The spatial resolution, or frame-to-frame variation of the recorded PORs (also referred to as noise), of the eyetracker was approximately 0.25° (Tobii Technology AB, 2003).

The eyetracker had a 17-in. screen, and the stimuli were displayed with a resolution of $1,024 \times 768$ pixels, with an eye-to-screen distance of ± 650 mm. Therefore, 1° of foveal angle was represented by about 33 pixels, or 10.5 mm. The individual squares of the chess board spanned about 20 mm ($\approx 2^\circ$) on screen, and each piece was displayed at about 7×8 mm ($< 1^\circ$).

Calibration was done by displaying five dots at known positions in the same area where the boards were displayed. One participant could not be calibrated satisfactorily, and his recordings were therefore not included in the final analysis.

5.3. Algorithm and metrics. For each participant–stimulus combination, the I-DT algorithm of Salvucci and Goldberg (2000) was applied with a slight addition (Line 6) to ensure that the identified fixation stayed within the confines of the threshold:

1. points \leftarrow All PORs in recording
2. **while** there are still points
3. Initialize window over first points to cover duration threshold
4. **if** dispersion of window points \leq threshold
5. Add additional points to the window until dispersion $>$ threshold
6. Remove last point added
7. Note a fixation at the centroid of the window points
8. Remove window points from points
9. **else**
10. Remove first point from points
11. **end while**

Since the primary aim of this study was to determine an optimum dispersion threshold for the various metrics of dispersion, the duration threshold (Line 3) was set at a constant 100 msec. This is in line with Manor and Gordon (2003), who found 100 msec to be a useful and practical balance between the theoretical minimum and maximum limits of fixation duration.

Salvucci and Goldberg (2000) explained the use of one particular measure of dispersion of a fixation (Lines 4 and 5) but also indicated that alternative dispersion metrics can be used. Five different metrics were used in turn to measure dispersion in this study, in order to determine whether the number, position, size, and duration of fixations that are returned by the algorithm are functions of the specific metric used. The metrics also served as verifications of one another, especially in terms of intermetric agreements of the scan paths that were returned.

1. For the distance dispersion (*DD*) metric (Shic et al., 2008), the distance between each point in a fixation and every other point should be less than a threshold value.

2. The distance threshold (*DT*) algorithm (Spakov & Miniotas, 2007) compares the distance between succes-

sive points with a threshold. Since the eyetracker samples points at a constant rate, it can also be referred to as a *velocity threshold*.

3. The largest distance from the center of a fixation to any POR in the fixation is referred to as the fixation *radius* (*Rd*) and should be less than a radius threshold value. This is essentially the same as Anliker's (1976) centroid-distance method, which requires that a predetermined percentage of points should be no further than a threshold from the centroid of a fixation.

4. The Salvucci dispersion (*Slv*) (Salvucci & Goldberg, 2000) is a measure of the maximum horizontal and vertical distance covered by the PORs in a fixation. For this study it was defined as $[(\text{Max } X - \text{Min } X) + (\text{Max } Y - \text{Min } Y)]/2 \leq \text{Threshold}$, which is different from the metric in Salvucci and Goldberg in that the average of the horizontal and vertical dispersion was used and not the sum. Although this change would have no effect on the fixations that would be identified, the order of magnitude of thresholds would be comparable to that of the other metrics (e.g., *DD*, *DT*, and *Rd*), especially for long, narrow fixations.

5. The standard deviation (*SD*) of the distances of all points from the center of a fixation should be less than a threshold value (Eyenal, 2001). This is referred to as the *position-variance* method (Duchowski, 2007) and is a variant of Anliker's (1976) centroid-distance method.

When implementing the *Rd* and *SD* metrics, the centroids of fixations were recomputed after each addition, and it was decided to allow no points to exceed the set threshold; that is, 100% of the points should be no further than the threshold distance from the centroid of a fixation.

Figure 5 shows seven consecutive PORs, with the *DD*, *DT*, *Rd*, and *Slv* measures of dispersion indicated. The distance between Points 4 and 7 is the largest of all the interpoint distances, whereas the distance between Points 5 and 6 is the largest of all differences between two consecutive points.

5.4. Procedure. In order to compare the various metrics of the I-DT algorithm on a range of thresholds, an applica-

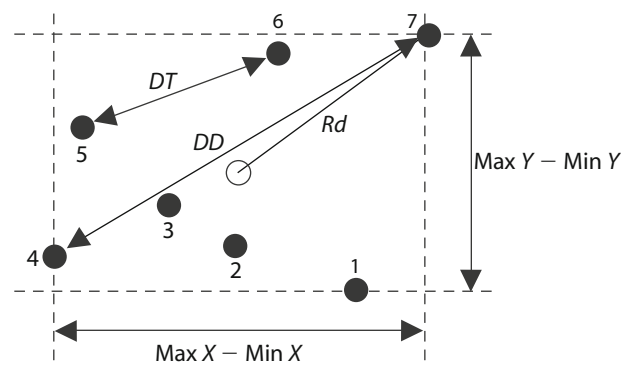


Figure 5. Seven consecutive points of regard with different measures of dispersion indicated. *DT*, distance threshold; *DD*, distance dispersion; *Rd*, radius; *Slv*, Salvucci dispersion.

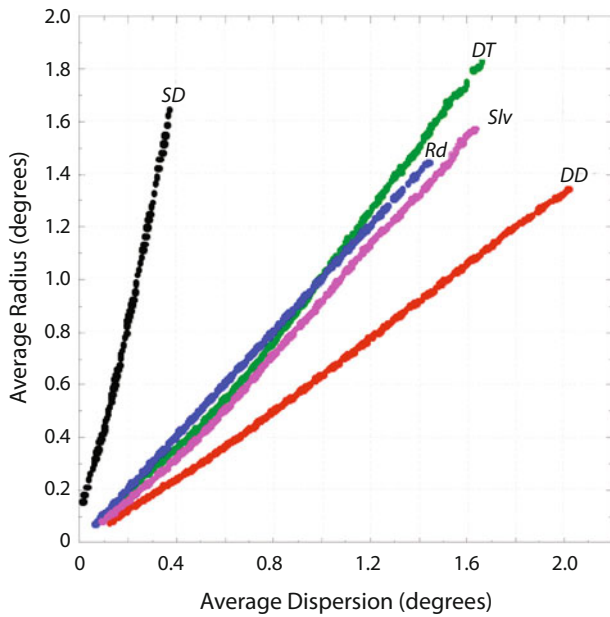


Figure 6. Average radius against average dispersion per metric. *SD*, standard deviation; *DT*, distance threshold; *Rd*, radius; *Slv*, Salvucci dispersion; *DD*, distance dispersion.

tion was developed that identifies fixations according to each one of the metrics in turn.

During data capturing, a separate recording of the gaze data was saved for each participant and stimulus. Applying the I-DT algorithm to each recording with each of the five metrics, one after the other, fixation sequences were generated for a range of threshold values. Thus, for each metric, a total of $p \times s \times t$ fixation sequences were generated, where p denotes the number of participants (30), s the number of stimuli (4), and t the number of thresholds for the respective metric. The value of t ranged from 70 for the *SD* metric to 320 for the *DD* metric (see Section 5.6 below).

For each fixation in every sequence that was generated, the exact dispersion (not to be confused with the threshold) was saved, along with its duration, radius, and nearest neighbor index (NNI; see Section 7.1 below). For every sequence of fixations, the percentage of PORs in the recording that was included in fixations was saved, as well as the averages over all fixations for dispersion, radius, duration, and the NNI.

Fixation sequences that were generated at equivalent thresholds (see Section 5.5 below) for each one of the different metrics were compared with one another in terms of (1) the number of fixations that were identified, (2) the percentage of PORs included in fixations, (3) the spatial distribution of the PORs that defined a fixation, and (4) the amount of difference between the fixation sequences. For each metric, it could thus be determined which threshold or range of thresholds would (1) lead to the most fixations, (2) include the most PORs in the fixations, (3) provide the highest measure of regularity in the spatial arrangement of the PORs within fixations, and (4) have the largest inter-metric agreement of fixation sequences.

5.5. A common basis for comparisons. Since the metrics differ from one another with regard to the entity that is measured for dispersion, it follows that they cannot be compared with one another at the same threshold. A common dispersion measure is needed to enable comparisons.

Consider a set of tuples (d_t, r_t) , where d_t represents the average dispersion of all fixations that were identified for all the participants and stimuli at threshold t , and r_t represents the average radius of the same fixations. For each metric, r_t is plotted against d_t for each t in the range of thresholds (Figure 6). The linear relationships are almost perfect (Table 1), confirming that the two constructs measure the same thing on different scales.

The linear relationships between the average radius and the average dispersions as applicable to the other metrics will be used to determine an equivalent radius threshold for the respective dispersion thresholds. Indicators from the various metrics will then be compared at equivalent radius thresholds. For example, the fixation sequences returned by the *Rds* metric at a threshold of 1.0° , the *DD* metric at 1.53° , the *DT* metric at 0.98° , the *Slv* metric at 1.08° , and the *SD* metric at 0.24° can be compared with one another, since these values can all be expressed in terms of an equivalent radius threshold of 1.0° .

5.6. Threshold range. With reference to the procedure described above, each metric was applied for a series of threshold values starting at 0° and incremented by 0.01° after each application. It is important to note that no conclusions can be based on thresholds below the eyetracker’s spatial resolution—that is, 0.25° (Tobii Technology AB, 2003), which is equivalent to a fixation radius of 0.125° .

A decision had to be made as to when to stop incrementing the threshold values. The percentage of PORs that are included in fixations was used as a barometer. This value largely depends on the threshold that is set: The larger the threshold, the more PORs will be included. Of course, given a threshold that is large enough to span the entire stimulus, 100% of PORs will be included. The PORs that are not included in fixations are mostly intermittent data captures during saccades, but they can also be noise due to participant or equipment variability. Occurrences of clustered PORs that do not meet the threshold for minimum duration (100 msec for this study) can also not be regarded as a fixation.

An obvious choice as to when to stop the process would be when all PORs are included in fixations. This would, however, unnecessarily prolong an already time-consuming process. Figure 7 shows graphs of the percent-

Table 1
Results of Linear Regression for Radius Versus Dispersion per Metric

Metric	Gradient	Intercept	r	R^2	p
<i>DD</i>	0.6817	-0.0416	.9996	.9993	<.01
<i>DT</i>	1.1354	-0.1105	.9972	.9945	<.01
<i>Rd</i>	1.0000	0	1.0000	1.0000	-
<i>Slv</i>	0.9976	-0.0797	.9992	.9983	<.01
<i>SD</i>	4.2225	0.0026	.9976	.9953	<.01

Note—*DD*, distance dispersion; *DT*, distance threshold; *Rd*, radius; *Slv*, Salvucci dispersion; *SD*, standard deviation.

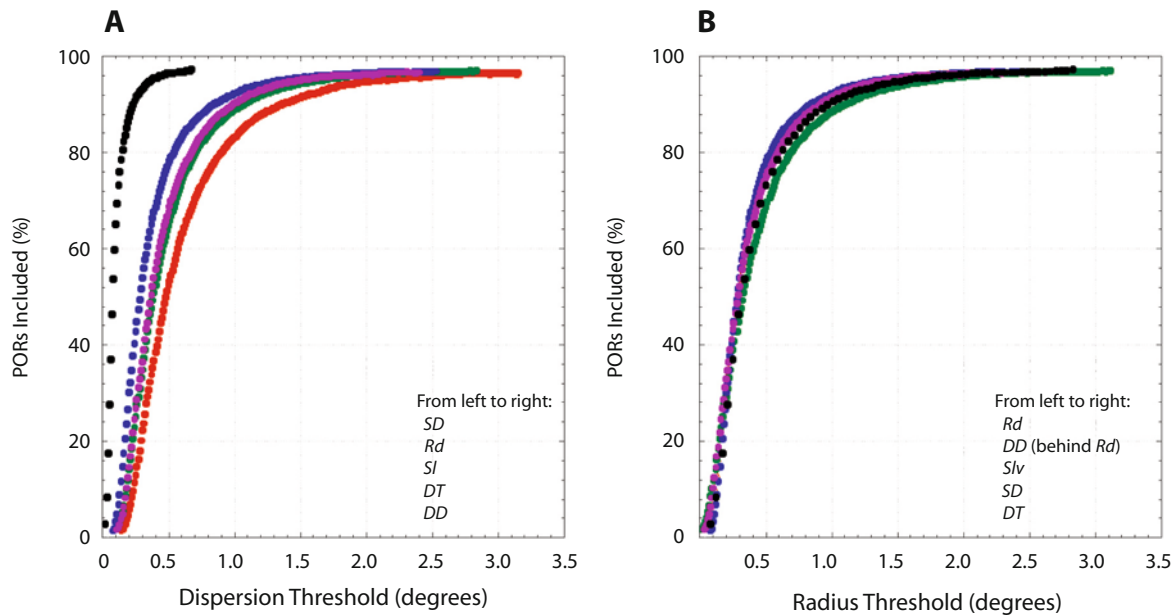


Figure 7. Percentages of points of regard (PORs) included in fixations against dispersion threshold of the particular metric (A) and the common radius threshold (B). *SD*, standard deviation; *Rd*, radius; *Slv*, Salvucci dispersion; *DT*, distance threshold; *DD*, distance dispersion.

ages of PORs that were included in fixations against the threshold value as applicable for each metric (Figure 7A) and against the common radius threshold (Figure 7B). As in Figure 6, each point in the graphs represents an average over all participants and stimuli at the specific threshold. The percentage of PORs increases rapidly at low radius thresholds and starts leveling out once the radius threshold reaches 1.5° . At this point, the process could have been halted, but to be on the safe side, it was decided to stop incrementing the threshold values for a specific metric when 98% of the PORs were included in fixations. For the *SD* metric, this level was reached at a dispersion threshold of about 0.7° , whereas for the *DD* metric, it was reached only at about 3.2° .

6. The Effect of Metric and Threshold on Fixation Identification

The next step was to test the main hypothesis of the study. In order to do this, the eyetracking data were analyzed to determine whether the metric and threshold used in a dispersion-based fixation detection algorithm have an effect on the number, size, and duration of fixations that are identified. The influence of the metric and dispersion threshold on the position of fixations will be dealt with in Section 7.2.

6.1. The influence of dispersion threshold on duration and size of fixations. Figures 8A and 8B show the graphs of average duration and average radius of fixations, respectively, against radius threshold. Linear regression showed that all lines were linear, with R^2 significant at the 99% level.

As the threshold for inclusion of a POR into a fixation was increased, more and more PORs were included, to the effect that both the fixation duration and radius

of PORs increased. With reference to the first hypothesis mentioned above, it can thus be concluded that both fixation duration and fixation size are linear functions of the dispersion parameter chosen. The findings of Shic et al. (2008) with regard to fixation duration are thus confirmed.

6.2. The influence of dispersion threshold on the number of fixations identified. Graphs of the average number of fixations that were identified per participant and stimulus at a specific threshold value are shown in Figure 9. These graphs should be interpreted together with Figure 7B. For a small threshold, only a small number of PORs were included per fixation—to such an extent that many fixations did not meet the requirement of minimum duration (100 msec in this study) and, thus, were discarded. As the threshold increased, more and more fixations met the duration threshold, and the number of both valid fixations and PORs that were included increased. At some stage, the increase in the number of PORs that were included became very small (Figure 7B), and any further increase in threshold caused more PORs per fixation, with the result that fewer fixations were identified. Therefore, a maximum number of fixations were identified at a certain critical threshold value that coincided with the elbow in Figure 7B (radius threshold $\approx 0.72^\circ$).

6.3. The influence of metric used on the duration, size, and number of fixations identified. From Figure 9, it is evident that the radius threshold at which the maximum number of fixations were identified was about the same for all metrics ($0.72^\circ \pm 0.04^\circ$). Figure 10 shows graphs of the average number, duration, and radius of fixations for all the participants and stimuli at this threshold. Table 2 shows the results of a one-way ANOVA for each one of these dependent variables against the metric used. All

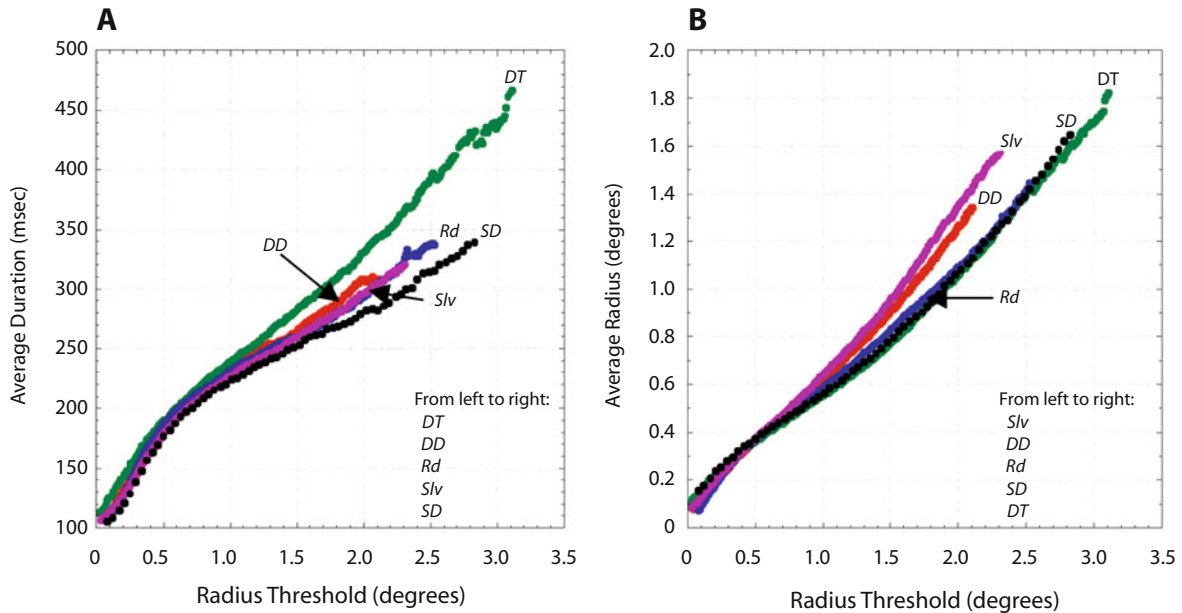


Figure 8. Average duration (A) and average radius of fixations (B) against the common radius threshold. *DT*, distance threshold; *DD*, distance dispersion; *Rd*, radius; *Slv*, Salvucci dispersion; *SD*, standard deviation.

three variables were significantly ($p < .05$) influenced by the metric used.

Using Tukey's post hoc test, it was established that the number of fixations returned by the *DT* metric differed significantly ($p < .01$) from that for all the other metrics, whereas the number of fixations returned by the other metrics did not differ significantly ($p < .01$) from one another at a radius threshold of 0.72° . Although the overall effect of metric on fixation duration was significant ($p < .05$), none of the individual intermetric differences in duration proved to be significant ($p < .05$). Most of the intermetric differences in fixation size—namely, *DD/DT*, *DD/Rd*, *DT/Slv*, *DT/SD*, *Rd/Slv*, *Rd/SD*, and *Slv/SD*—were significant ($p < .01$). Referring back to Figure 8, it is evident that these differences will become more pronounced at higher thresholds.

Taking into account the results of Sections 6.1, 6.2, and 6.3 and with reference to the hypothesis, it can thus be concluded that the dispersion threshold value that is chosen by a researcher has an influence on the number, size, and duration of fixations that are identified. The influence of the metric used in the dispersion-based algorithm on these variables is not consistent.

These results confirm the assertions by Karsh and Breitenbach (1983) and Shic et al. (2008) that different values for dispersion thresholds can lead to inconsistent results, especially when an algorithm that is sensitive to parameter values such as I-DT is used. The issue in this article is to determine the optimal settings in order to negate or compensate for the sensitivity.

7. Evaluation of Fixation Sequences

It might be reasonable to expect that the same fixations should be identified for a particular recording irrespective

of the metric or threshold settings that are used. It is, however, difficult, if not impossible, to determine the correct set of fixations to use as a benchmark against which to evaluate the accuracy of a specific metric or threshold setting.

Since it was determined above that the number, size, and duration of fixations that are identified depend on the dispersion threshold that was chosen, the question arises as to what threshold would return the best fixation sequence. In order to determine this optimum threshold value for each of the metrics, the fixation sequences that they returned at each threshold (in the range of threshold values that were inspected) were examined in terms of four indicators: the percentage of PORs included in fixations (Section 5.6), the number of fixations that were identified (Section 6.2), the spatial distribution of PORs that define a fixation (Section 7.1), and average Levenshtein agreement (*LA*) with other metrics (Section 7.2). The optimum threshold for each metric was accepted to be the one at which these indicators are optimized to such an extent that the position and duration of a viewer's actual gaze is represented as accurately as possible. The nature of the optimization for each indicator is discussed in Section 8 below.

7.1. Nearest neighbor index. The NNI denotes the ratio between the average of the minimum distances between points and the mean random minimum distance that one would expect if the distribution was random (Clark & Evans, 1954). Although the index was originally intended for use in fieldwork in botany, it is still widely used in spatial analysis of data in agriculture, paleontology, and crime analysis.

Initially, the nearest neighbor distance for each point in a set is computed. The average of these distances (\bar{r}_A) is then divided by \bar{r}_E , the expected mean nearest neighbor distance if the distribution were random (Clark & Evans, 1954):

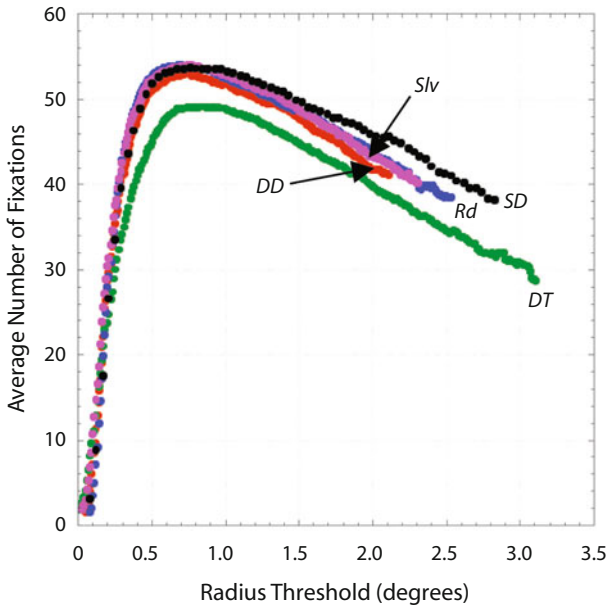


Figure 9. Average number of fixations against radius threshold. *Slv*, Salvucci dispersion; *DD*, distance dispersion; *Rd*, radius; *SD*, standard deviation; *DT*, distance threshold.

$$\bar{r}_E = 0.5\sqrt{A/N}$$

where A = the area covered by the points and N = the number of points.

The ratio \bar{r}_A/\bar{r}_E is equal to 1 when the distribution is random. A value lower than 1 indicates clustering, whereas a value higher than 1 indicates regular spacing. Theoretically, the NNI lies between 0 (maximum clustering) and 2.1491 (strictly regular hexagonal pattern).

Although this index was previously used in eyetracking to investigate the dispersion of fixations in a recording (Camilli et al., 2008; Di Nocera, Camilli, & Terenzi, 2007; Di Nocera, Terenzi, & Camilli, 2006; , the present study is interested in the dispersion of PORs within fixations. Since the tremor, drift, and microsaccades are in-

voluntary, it is accepted that these small eye movements are random, and therefore, the PORs within a fixation should have an NNI larger than 1. Clustering of PORs within a fixation might indicate the existence of separate fixations.

For the purposes of this article, the area covered by a fixation was computed by defining a convex hull around the comprised PORs, as in Figure 2B, and then calculating the area of that hull. In order to compensate for edge effects, Donnelly's (1978) adjustment was added to \bar{r}_E as follows:

$$\bar{r}_E = 0.5\sqrt{A/N} + (0.0514 + 0.041/\sqrt{N}) \times B/A,$$

where B = the perimeter of the convex hull.

A graph of the average NNI for the fixations that were identified against threshold value is shown in Figure 11A. The regularity of the dispersion of PORs within fixations increases sharply with the threshold and reaches a maximum before dropping again.

For all the metrics, the maximum NNI is reached at a radius threshold of $0.26^\circ \pm 0.03^\circ$. The equivalent dispersion threshold for individual metrics can be computed from this value by using the regression coefficients in Table 1. A one-way ANOVA indicated that the NNIs for the respective metrics at 0.26° differ significantly [$F(4, 1122) = 16.8, p < .01$] from one another. Tukey's post hoc analysis revealed that this effect was due largely to the low value of the *SD* metric, which differed significantly ($p < .01$) from all other metrics at 0.26° . None of the other intermetric differences in NNI were significant at the 99% level, although the difference between *DD* and *Slv* was significant at the 95% level.

The statistical significance of the NNI can be computed as well (Clark & Evans, 1954):

$$r_{SE} = .26136 / \sqrt{N^2/A}$$

$$z = (\bar{r}_A - \bar{r}_E) / r_{SE}.$$

If $z > 2.33$, the null hypothesis that the PORs are clustered can be rejected at the 99% confidence level (one-tailed). If

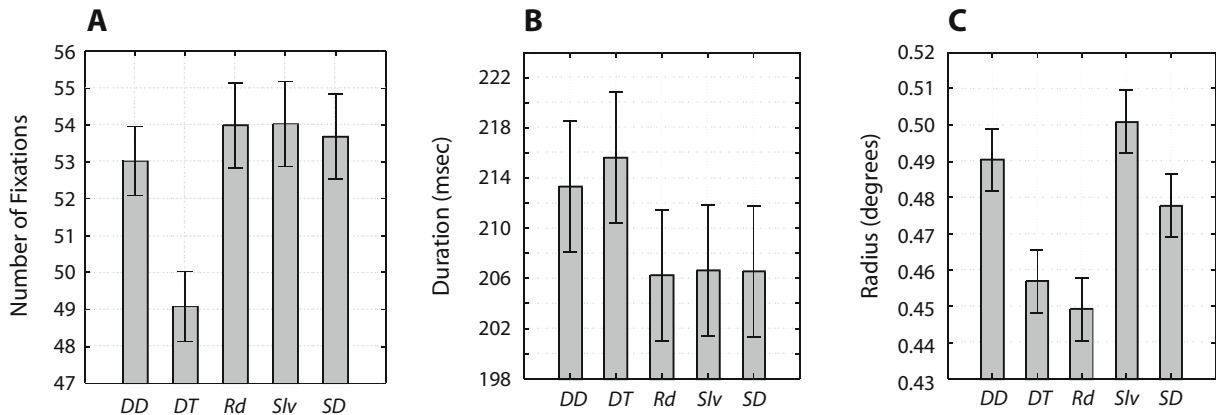


Figure 10. Number (A), duration (B), and radius of fixations (C) against metric at a radius threshold of 0.72° . *DD*, distance dispersion; *DT*, distance threshold; *Rd*, radius; *Slv*, Salvucci dispersion; *SD*, standard deviation.

Table 2
Means and Standard Deviations of Number, Duration,
and Radius of Fixations at a Radius Threshold of 0.72°

Metric	N	Fixations		Duration		Radius	
		M	SD	M	SD	M	SD
DD	230	53.0	8.3	213.3	39.1	0.49	0.06
DT	230	49.1	10.5	215.6	43.6	0.46	0.07
Rd	230	54.0	8.1	206.3	37.8	0.45	0.07
Slv	230	54.0	8.6	206.7	40.0	0.50	0.05
SD	230	53.7	8.0	206.6	40.2	0.48	0.09
		$F(4,1145) = 13.2$ $p < .01$		$F(4,1145) = 2.8$ $p < .05$		$F(4,1145) = 24.6$ $p < .01$	

Note—DD, distance dispersion; DT, distance threshold; Rd, radius; Slv, Salvucci dispersion; SD, standard deviation.

this null hypothesis is rejected, it implies that the PORs are either random or regularly dispersed. A graph of the average z value of the NNI for the various metrics against radius threshold is shown in Figure 11B. With the exception of very small threshold values, the dispersion of PORs within fixations was significantly nonclustered for the entire range of thresholds that was examined. Therefore, on the basis of NNI only, any dispersion threshold above $\approx 0.1^\circ$ will result in valid fixation sequences, although the best sequence will be attained at a radius threshold of $\approx 0.26^\circ$, where a valid fixation is regarded as one where its comprised PORs are significantly nonclustered.

7.2. Levenshtein agreement. The Levenshtein (1966) distance is a metric for measuring the amount of difference between two sequences. The Levenshtein distance (LD) between two character strings is given by the minimum number of operations, defined as insertion, deletion, or substitution, needed to transform one string into the other. The LD has no cost function, and every operation has equal weight. The metric has been applied previously in eyetracking research in the comparison of scan paths

for different participants (West, Haake, Rozanski, & Karn, 2006). This study differs in the sense that scan paths that were returned by different algorithms for the same recording of gaze data were compared.

For each threshold t , a sequence of fixations or scan path F_t is generated that is indexed according to the squares on the chess board—for example, d6, e6, f7, h7, f1. In seeking to identify the best fixation sequence for a specific recording, the sequences returned by different metrics at equivalent radius thresholds can be compared with one another. For example, if a sequence d4, c4, c6, f6, g7, g5 was reported by metric M1 for a specific recording of gaze data at a specific threshold, and a sequence d4, c4, c5, f6, g5 was reported for the same recording at the equivalent threshold with metric M2, the difference between the sequences would be 2: one substitution (c5 for c6) and one deletion (g7).

The question is to determine which of the reported scan paths is the closest to reality. In this study, it was accepted that if five different metrics return the same or nearly the same scan path, it is highly improbable that they are all wrong and that that scan path should

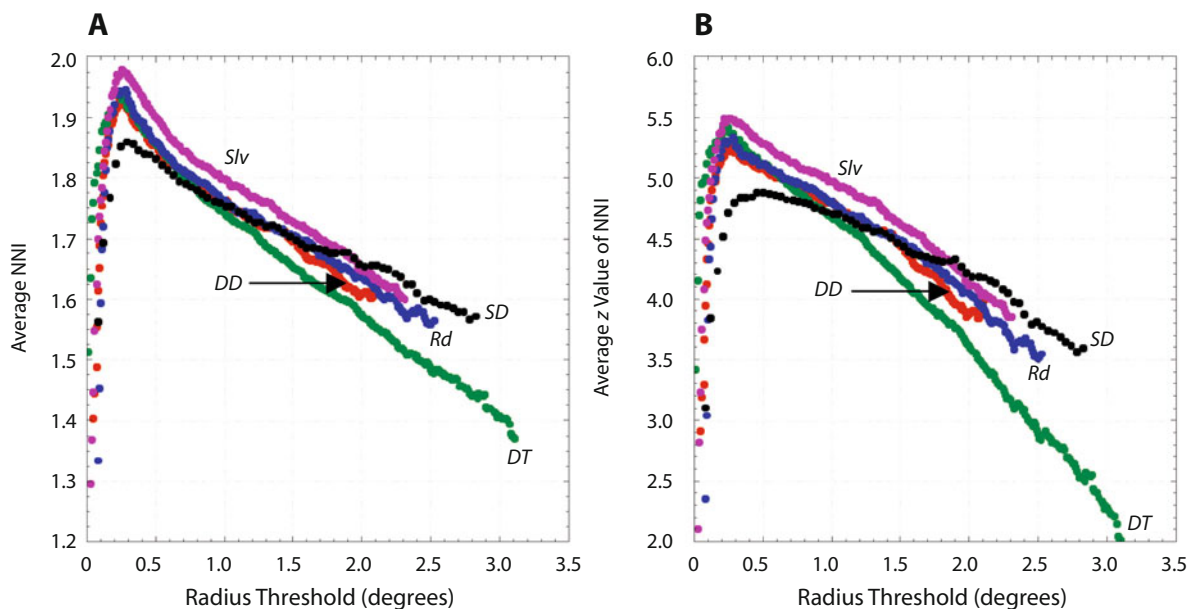


Figure 11. Graph of average nearest neighborhood index (NNI) (A) and average z value of NNI (B) against the radius threshold. Slv, Salvucci dispersion; DD, distance dispersion; Rd, radius; SD, standard deviation; DT, distance threshold.

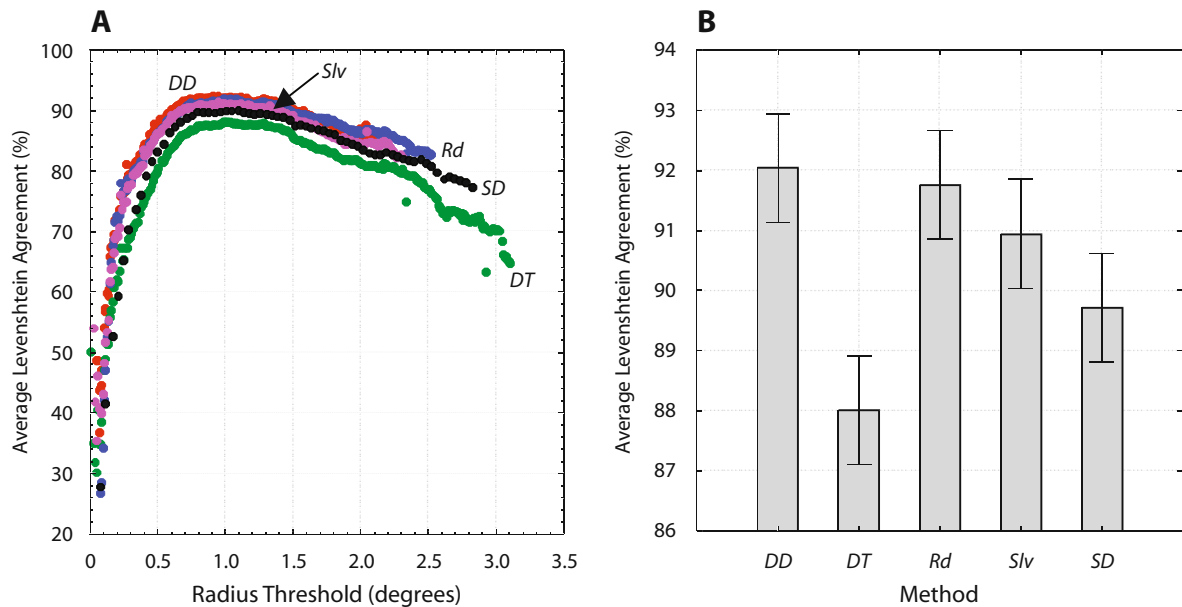


Figure 12. (A) Average Levenshtein agreement against radius threshold per metric. (B) Average Levenshtein agreement per metric at a radius threshold of 1.0°. *DD*, distance dispersion; *Slv*, Salvucci dispersion; *Rd*, radius; *SD*, standard deviation; *DT*, distance threshold.

be accepted as the best approximation of the correct sequence.

If the Levenshtein distance between two fixation sequences is expressed as a percentage of the length of the longest sequence, the *LA* between the two sequences can be expressed as $100 - LD$. For each metric and each participant recording, the average *LA* with every other metric was determined at each of the distinct thresholds. Finally, the *LA* at each threshold was also averaged over recordings. Figure 12A shows a graph of the average *LA* against radius threshold for each one of the metrics. A surprisingly low maximum of just over 90% was attained, indicating that even at the optimum threshold for each metric they still differed substantially with regard to the number and position of fixations that were determined. Another important observation was that, unlike the NNI, where the peaks were rather sharp, the *LA* stayed more or less constant between 0.7° and 1.3°.

For all metrics, a maximum *LA* was reached at a radius threshold of $1.0^\circ \pm 0.1^\circ$. Figure 12B shows a graph of the average *LA* per metric at a radius threshold of 1.0°. A one-way ANOVA indicated that the *LA* for the respective metrics at this threshold differed significantly [$F(4,1116) = 14.2$, $p < .01$]. Using Tukey's post hoc test, it was established that the *LA* differed significantly ($p < .05$) between the following pairs of metrics: *DT/DD*, *DT/Rd*, *DT/Slv*, *DD/SD*, and *Rd/SD*. With reference to the hypothesis, it can thus be concluded that the number and position of fixations are functions of both the metric used and the threshold value.

8. Finding the Optimum Threshold

Two indicators were used to evaluate the quality or "correctness" of the fixation sequences that are identified by the I-DT algorithm. The number of PORs that are

included in fixations and the number of fixations that are identified might also be indicative of the optimum dispersion threshold that should be used. Some of the indicators might be related or interdependent, whereas others might be considered to be less important or not important at all. In order to consider them as indicators of an optimum threshold, they must contribute to the identification of fixations so that the position and duration of a viewer's actual gaze is represented as accurately as possible.

The NNI succeeds in identifying an *absolute* lower limit for dispersion threshold—that is, a threshold below which clusters of PORs are unlikely to represent fixations. The *LA* is based on an assumption that the optimum threshold occurs at a value at which the agreement between the various metrics in terms of the positions of individual fixations is at a maximum. Unlike these two indicators, the percentage of PORs that are included in fixations does not show a definite maximum value, but the point at which the increase in PORs levels off (the elbow in Figure 7B) may indicate a *preferred* lower limit for the threshold—that is, a threshold above which a substantial number of PORs are included in fixations.

If all of the above candidate indicators are considered, an optimum threshold for the I-DT algorithm, when a specific metric is applied, would be a value that (1) will include as many fixations as possible, (2) will include as many PORs as possible, (3) will ensure maximum agreement with other metrics with respect to the number and position of fixations that are identified (*LA*), and (4) will return fixations in which the PORs are as regularly dispersed as possible.

To consider the threshold at which a maximum number of fixations are detected as an optimum threshold could be controversial, since it might be argued that the number of fixations is not important as long as the correct fixations

are identified. Eyetracking results are seldom based on single fixations but on aggregates of time spent on specific areas of interest (AOIs). Therefore, having one long fixation or two shorter consecutive fixations on the same AOI might be immaterial. It is, however, also true that less doubt about the validity of smaller fixations (which inevitably means more fixations) exists, provided that the radius threshold is above 0.3° (Figure 11). Furthermore, the centroid of longer fixations in which the PORs are clustered might be calculated at a point at which a participant has never actually looked! Also, the higher granularity of smaller fixations allows more distinct boundaries between AOIs. The one negative aspect of preferring shorter fixations is the fact that the correct setting of the duration threshold is more critical.

The percentage of PORs included does not show a definite maximum but approaches 100% as the threshold is increased. It would be preferable to utilize as many PORs as possible, while ensuring that noise or intermittent gaze captures during saccades are ignored. The elbow, at about 0.75° (Figure 7B), coincides more or less with the threshold at which a maximum number of fixations are identified and represents a point at which the increase in percentage of PORs becomes smaller as the threshold increases.

The *LA* between sequences may be the most intuitive indicator of an optimum threshold. In the extreme, if each of two metrics returns a single “fixation” that includes all PORs in a recording with centroid in the middle of a stimulus, the *LA* between them will be 100%, but the threshold at which this happens will be unacceptably high. If, however, the *LA* is high while there are a large number of fixations with a reasonable average duration (150–250 msec), it is highly unlikely that the agreement is due to chance. In this study, in which the intermetric agreements of five metrics were considered, the probability of accidental agreement was even less. Therefore, provided that a reasonable number of fixations of average duration is returned, this is considered to be the indicator of choice.

The importance of the NNI is difficult to assess. Of course, clustering of PORs within a fixation suggests separation of fixations, but it was shown above that as long as the radius threshold exceeds 0.1°, the significance of this index is such that clustering is never an issue (Figure 11).

The fact that the *LA* remains at more or less the same level between 0.7° and 1.3° could be utilized to maximize the other, albeit less important, indicators within this range as well. Within this interval, the number of fixations and the NNI would be at a maximum at 0.7°, whereas the number of PORs that are included would be at a maximum at 1.3° (Figure 13). It is thus a matter of deciding which is the most important: having the PORs within fixations non-clustered while also maximizing the number of fixations or utilizing as much of the available data as possible. For a longer duration threshold—for example, 200 msec—a researcher might consider placing higher emphasis on smaller fixations. If, however, the duration threshold is low (as was the case in this study), it might be better to opt for a higher dispersion threshold. Looking at the form of the graphs for the *LA*, it is less critical to err with a threshold that is too large than to have it be too small.

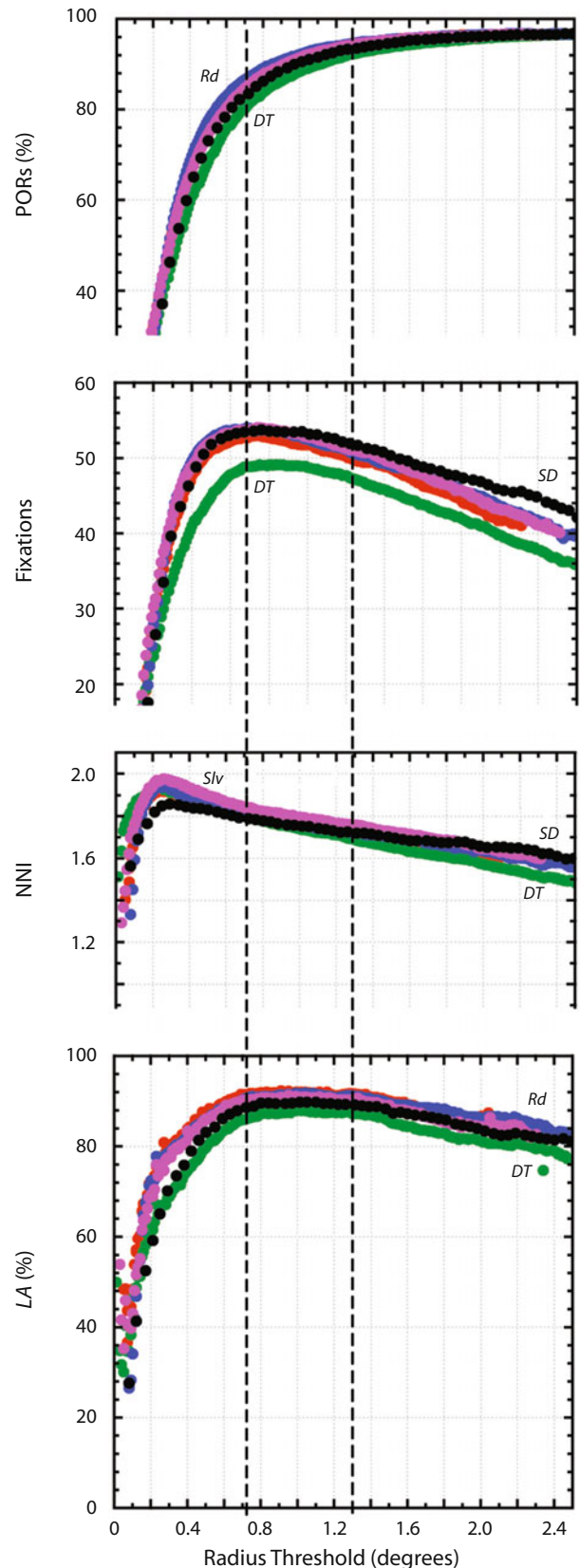


Figure 13. Per metric graphs of the various indicators of fixation accuracy against radius threshold. *Rd*, radius; *DT*, distance threshold; *SD*, standard deviation; *Slv*, Salvucci dispersion.

9. Summary

This study was undertaken to determine the criticality of the setting of the dispersion threshold in a dispersion-based algorithm for fixation detection. The study was also an attempt to determine the optimum dispersion threshold for each of five different measures of dispersion.

The radius of fixations was used as a common dispersion measure across the various metrics. A 1° radius threshold would, for example, be equivalent to a dispersion threshold of 0.24° for the *SD* metric.

The hypothesis that the number, position, size, and duration of fixations that are identified are functions of the metric used for dispersion proved to hold in most respects. The number of fixations identified by the *DT* metric was significantly lower than that identified by the other metrics, whereas the duration of fixations identified by the *DD* and *DT* metrics were generally longer than those identified by the *Rd*, *Slv*, and *SD* metrics. The *DT* and *Rd* metrics led to fixations that were smaller than those identified by the *DD*, *Slv*, and *SD* metrics. The *DT* and *SD* metrics returned scan paths that differed somewhat from those returned by the *DD*, *Slv*, and *Rd* metrics, indicating that the position of fixations is also a function of the metric used.

The fact that the number, position, size, and duration of fixations that are identified are functions of the threshold value could always have been accepted as obvious, but definite trends were identified in each case. The size and duration of fixations are linear functions of the threshold value, and the number of fixations reaches a maximum at some threshold, whereafter it decreases linearly with the threshold. The position of fixations also depends on the threshold, as indicated by the *LA* that is also maximized at some threshold value.

It was found that there is no definite optimum dispersion threshold to be used in a dispersion-based fixation detection algorithm but that the various indicators point to an acceptable range for the radius threshold between 0.7° and 1.3°. Within this range, the *LA* between metrics is reasonably constant at about 90%. A lower threshold within this range will ensure more fixations with PORs that are less clustered within fixations, whereas a larger threshold will utilize more PORs. Using threshold values less than 0.7° will result in fixation sequences that differ considerably from one another. Threshold values above 1.3° are less of a danger, but since longer and larger fixations will be identified with PORs that can be more clustered within fixations, it is possible that fixations will be reported at positions that differ from a viewer's actual point of attention.

The 0.7°–1.3° range for the radius threshold is equivalent to a very small range of 0.17°–0.30° for the *SD* metric. The margin of error for the *DD* metric is somewhat bigger, with an acceptable range of 1.1°–2.0°.

The fact that the correct setting of the dispersion threshold is quite critical and that the number and position of fixations that are identified are dependent on the setting has important implications for eyetracking research. Researchers should be aware of the exact metric used in the tools that they are using and should set the threshold accordingly.

10. Future Work

A linear regression was done for the relationship between average dispersion according to a specific metric and fixation radius. A very good fit was found, but a study should be done with another type of stimulus to determine whether the set of coefficients can be applied generically.

This study was done using a Tobii 1750 50-Hz eyetracker while using stimuli that consisted of four chessboards of equal size and fixed granularity. It will be necessary to determine whether the results above for the range of optimum thresholds per metric can be replicated for an eyetracker with higher frequency and for other types of stimuli. It is especially important to repeat the research on stimuli with varying granularity, such as Web sites.

Furthermore, this study was confined to the threshold setting of the dispersion parameter only. Although theoretical evidence about the minimum duration of a fixation exists, a similar study can be undertaken to determine whether the theoretically determined values also hold in practice for the duration threshold of dispersion-based algorithms. Also, the study needs to be repeated for other algorithms in which one or more thresholds need to be set, such as the velocity threshold algorithm.

In this study, the *LA* with other metrics was used to determine the quality of the identified fixation sequence. A study could be constructed in which the position and order of fixations can be manipulated by the researchers—for example, by explicitly instructing participants to focus at specific items on a stimulus at specific moments. This set of fixations could then be used as a control set against which the correctness of the fixations identified by the algorithm–metric combination could be measured. It could also be possible to simulate fixation data that will be free of noise due to equipment or human variability and the exact position of every fixation is known, as well as the PORs that comprise it.

Although all precautions were taken to accurately calibrate each participant before and during a recording, the accuracy of calibrations was not considered in this study. It could be a topic for future research to determine the extent to which accuracy of calibration influences threshold settings.

In the present study, only the averages for all the participants were used. It could be an interesting topic to determine to what extent individual differences, such as near-sightedness, age, wearing of glasses, and so forth contribute to indicators such as number of fixations, the NNI, and the *LA*.

AUTHOR NOTE

I thank my colleagues, Tanya Beelders and Chung-Yen So, for their assistance in capturing the data for this study. A special word of thanks goes to Lyndon Bouah, who arranged it so that we could capture data during the South African Closed Chess Championships of 2007. My family, who motivated me to persist through the long tedious hours of data analysis, deserves special mention as well. I also thank Tanya Beelders, Marco Camilli, and Joseph Goldberg, as well as the anonymous reviewers, for their extremely valuable comments and feedback on the original manuscript. Correspondence concerning this article should be addressed to P. Blignaut, Department of Computer Science and Informatics, University of the Free State, P.O. Box 339, Bloemfontein 9300, South Africa (e-mail: pieterb.sci@ufs.ac.za).

REFERENCES

- ANLIKER, L. (1976). Eye movements: On-line measurement, analysis, and control. In R. A. Monty & J. W. Senders (Eds.), *Eye movements and psychological processes* (pp. 185-199). Hillsdale, NJ: Erlbaum.
- CAMILLI, M., NACCHIA, R., TERENCE, M., & DI NOCERA, F. (2008). ASTEF: A simple tool for examining fixations. *Behavior Research Methods*, **40**, 373-382.
- CLARK, P. J., & EVANS, F. C. (1954). Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology*, **35**, 445-453.
- DI NOCERA, F., CAMILLI, M., & TERENCE, M. (2007). A random glance at the flight deck: Pilot's scanning strategies and the real-time assessment of mental workload. *Journal of Cognitive Engineering & Decision Making*, **1**, 271-285.
- DI NOCERA, F., TERENCE, M., & CAMILLI, M. (2006). Another look at scanpath: Distance to nearest neighbor as a measure of mental workload. In D. de Waard, K. A. Brookhuis, & A. Toffetti (Eds.), *Developments in human factors in transportation, design, and evaluation* (pp. 295-303). Maastricht: Shaker.
- DONNELLY, K. P. (1978). Simulations to determine the variance and edge effect of total nearest-neighbor distance. In I. Hodder (Ed.), *Simulation studies in archaeology* (pp. 91-95). Cambridge: Cambridge University Press.
- DUCHOWSKI, A. T. (2007). *Eye tracking methodology: Theory and practice* (2nd ed.). London: Springer.
- ELO, A. (1978). *The rating of chess players past and present*. New York: Arco.
- EYENAL (2001). *Eyenal (eye-analysis) software manual*. Bedford, MA: Applied Science Group. Retrieved June 12, 2008, from www.csmbm.princeton.edu/resources/DocsAndForms/site/forms/Eye_Tracker/Eyenal.pdf.
- GITELMAN, D. R. (2002). ILAB: A program for postexperimental eye movement analysis. *Behavior Research Methods, Instruments, & Computers*, **34**, 605-612.
- HEMINGHOUS, J., & DUCHOWSKI, A. T. (2006). iComp: A tool for scanpath visualization and comparison. In *Proceedings of the 3rd Symposium on Applied Perception in Graphics and Visualization (APGV '06)* (p. 152). New York: ACM Press.
- IRWIN, D. E. (1992). Visual memory within and across fixations. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 146-165). New York: Springer.
- JACOB, R. J. K. (1990). What you look at is what you get: Eye movement-based interaction techniques. In *Proceedings ACMCHI '90 Human Factors in Computing Systems* (pp. 11-18). New York: ACM Press.
- JACOB, R. J. K. (1993). Eye movement-based human-computer interaction techniques: Toward noncommand interfaces. In H. R. Hartson & D. Hix (Eds.), *Advances in human-computer interaction* (Vol. 4, pp. 151-190). Norwood, NJ: Ablex.
- JACOB, R. J. K., & KARN, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 573-605). Amsterdam: Elsevier.
- JUST, M. A., & CARPENTER, P. A. (1984). Using eye fixations to study reading comprehension. In D. E. Kieras & M. A. Just (Eds.), *New methods in reading comprehension research* (pp. 151-182). Hillsdale, NJ: Erlbaum.
- KARN, K. S. (2000). "Saccade pickers" vs. "fixation pickers": The effect of eye tracking instrumentation on research. In *Proceedings of the 2000 Symposium on Eye Tracking Research and Applications* (pp. 87-88). New York: ACM Press.
- 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.
- KUMAR, M., KLINGNER, J., PURANIK, R., WINOGRAD, T., & PAEPCKE, A. (2008). Improving the accuracy of gaze input for interaction. In *Proceedings of the 2008 Symposium on Eye Tracking Research and Applications* (pp. 65-68). New York: ACM Press.
- LANKFORD, C. (2000). Gazetracker: A software designed to facilitate eye movement analysis. *Proceedings of the 2000 Symposium on Eye Tracking Research and Applications* (pp. 51-55). New York: ACM Press.
- LEVENSHTAIN, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Doklady Physics*, **10**, 707-710.
- MANOR, B. R., & GORDON, E. (2003). Defining the temporal threshold for ocular fixation in free-viewing visuo-cognitive tasks. *Journal of Neuroscience Methods*, **128**, 85-93.
- RAYNER, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, **124**, 372-422.
- SALVUCCI, D. D. (2000). An interactive model-based environment for eye-movement protocol analysis and visualization. In *Proceedings of the 2000 Symposium on Eye Tracking Research and Applications* (pp. 57-63). New York: ACM Press.
- SALVUCCI, D. D., & GOLDBERG, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research and Applications* (pp. 71-78). New York: ACM Press.
- SANTELLA, A., & DECARLO, D. (2004). Robust clustering of eye movement recordings for quantification of visual interest. In *Proceedings of the 2004 Symposium on Eye Tracking Research and Applications* (pp. 27-34). New York: ACM Press.
- SHIC, F., CHAWARSKA, K., & SCASSELLATI, B. (2008). The incomplete fixation measure. In *Proceedings of the 2008 Symposium on Eye Tracking Research and Applications* (pp. 111-114). New York: ACM Press.
- SPAKOV, O., & MINIOTAS, D. (2007). Application of clustering algorithms in eye gaze visualizations. *Information Technology & Control*, **36**, 213-216.
- TOBII TECHNOLOGY AB (2003). Product description: Tobii 1750 Eye-tracker. Available from www.tobii.com.
- TOBII TECHNOLOGY AB (2008). *Tobii Studio 1.2 user manual version 1.0*. Tobii Technology Analysis Software. Retrieved October 18, 2008, from www.tobii.com/archive/pages/18058/downloads.aspx.
- TOLE, J. R., & YOUNG, L. R. (1981). Digital filters for saccade and fixation detection. In D. F. Fisher, R. A. Monty, & J. W. Senders (Eds.), *Eye movements: Cognition and visual perception* (pp. 185-199). Hillsdale, NJ: Erlbaum.
- URRUTY, T., LEW, S., IHADADDENE, N., & SIMOVICI, D. A. (2007). Detecting eye fixations by projection clustering. *ACM Transactions on Multimedia and Computing, Communications and Applications*, **3**, 23:1-23:20.
- WEST, J. M., HAAKE, A. R., ROZANSKI, E. P., & KARN, K. S. (2006). eyePatterns: Software for identifying patterns and similarities across fixation sequences. In *Proceedings of the 2006 Symposium on Eye Tracking Research and Applications* (pp. 149-154). New York: ACM Press.
- WIDDEL, H. (1984). Operational problems in analysing eye movements. In A. G. Gale & F. Johnson (Eds.), *Theoretical and applied aspects of eye movement research* (pp. 21-29). New York: Elsevier.

NOTE

1. The ELO system was developed by Arpad Elo as a means for measuring and rating the average playing ability of chess players (Elo, 1978).

(Manuscript received August 19, 2008;
revision accepted for publication December 18, 2008.)