Attention to repeated images on the World-Wide Web: Another look at scanpath theory

SHEREE JOSEPHSON
Weber State University, Ogden, Utah

and

MICHAEL E. HOLMES
Ball State University, Muncie, Indiana

The scanpath theory of visual perception was tested using Web pages as visual stimuli. *Scanpaths* are repetitive sequences of fixations and saccades that occur upon reexposure to a visual stimulus. Since Internet users are exposed to repeated visual displays, the Web provides ideal stimuli to test this theory. Eye movement data were recorded for subjects' repeated viewings of three Web pages over three sessions. Resemblance of eye path sequences was measured with a string-edit method; multidimensional scaling and cluster analysis were used to group sequences. Support was found for the scanpath theory; some clusters included pairs of sequences from the same subject. A repeated measures analysis of variance revealed a statistically significant main effect for stimulus type, with a text-intensive news story page generating more similar sequences than a graphic-intensive advertising page. There was a statistically significant main effect for cross-viewing comparisons, reflecting a linear trend in which eye paths for the same subject became more alike over time.

Since news sites emerged on the Web scene in 1994, followed by banner advertising in late 1994 and portal pages in 1996, Internet users have been repeatedly exposed to certain visual displays of information on their computer screens. Users of on-line newspapers read multiple news stories, with screen after screen of text displayed in the same visual pattern. Heavy users of a product or service are likely to call up the corporate home page for updated product information or to make a purchase on line multiple times before the image is changed. And Web users access the Internet through the same portal page time after time, seldom, if ever, changing their starting-page default.

These common practices of Internet users make the Web a natural place to test a somewhat controversial and often-discussed theory of visual perception, that of the existence of *scanpaths.* Noton and Stark (1971a, 1971b) defined scanpaths as repetitive sequences of fixations and saccades. Fixations occur when the eye is relatively immobile and indicate the area where attention is being allocated (Rayner, 1995). Saccades are the quick jumps of the eye and serve the purpose of bringing a new visual region upon the fovea, the part of the retina where visual acuity is the greatest and where clear detail can be obtained (Rayner, 1978).

Noton and Stark's (1971b) scanpath theory predicts that a person scans a new stimulus during the first exposure and stores the sequence of fixations in memory as a spatial model, so that a scanpath is established. When the person is reexposed to the stimulus, the first few eye movements tend to follow the same scanpath that was established during the initial viewing of the stimulus, which facilitates stimulus recognition.

Research has also indicated that when a participant is presented with a blank screen and told to visualize a previously seen figure, the scanpath is similar to that established when he or she had actually viewed the figure (Stark, 1994). According to the scanpath theory, a spatial model—considered a precognitive, perceptual model—controls the sequences of eye movements. Noton and Stark (1971b) asserted that "the internal representation of a pattern in memory is a network of features and attention shifts, with a habitually preferred path through the network, corresponding to the scanpath. During recognition, this network is matched with the pattern, directing the eye or internal attention from feature to feature of the pattern" (p. 940).

Noton and Stark (1971b) argued that control of the eye by specific features in the visual stimulus is improbable, because of the differences in scanpaths of different viewers for a given pattern. They also rejected the explanation that viewers are driven by habits, because of the variation in scanpaths of a given viewer for different stimulus patterns.

Correspondence concerning this article should be addressed to S. Josephson, Communication Department, Weber State University, 1605 University Circle, Ogden, UT 84408-1605 (e-mail: sjosephson@weber.edu).

Ellis and Smith (1985) elaborated on Noton and Stark's (1971b) scanpath theory by suggesting that scanpaths can be generated by completely random, stratified random, or statistically dependent stochastic processes, but they did not test these conjectures. A completely random process assumes that each element of a visual stimulus has an equal probability of being fixated on during each fixation. A stratified random process assumes that the probabilities of visual elements' being fixated reflect the attentional attractiveness of those elements and do not depend on previous fixations. The statistically dependent stochastic process specifies that the position of a fixation depends on previous fixations. In view of the perceptual processes that are assumed to underlie eye movements, Rayner (1995) and Stark and Ellis (1981) believed that it is unlikely that saccades from one fixation point to another are generated by either completely random or stratified random processes and looked toward statistically dependent stochastic processes as an explanation.

Early studies on eye movements made while participants viewed scenes and pictures also have provided evidence that visual exploration or search is not random. Eye movements are related to the content of the scene (H. F. Brandt, 1940; Buswell, 1935; Llewellyn-Thomas, 1968; Yarbus, 1967), although, as was noted, they are not controlled by it (Noton & Stark, 1971b). The pattern of fixations and saccades can be changed by altering the pictures or the task. Content that contains unique detail also dramatically influences the pattern of fixations and saccades, since such detail draws more attention than does common or expected visual information. Viewers tend to fixate on unique regions of visual scenes sooner, more frequently, and for longer durations than on any other area of the visual scene (Antes, 1974; Mackworth & Morandi, 1967).

Some studies focus on the role that peripheral vision plays in determining where a person will look next. Parker (1978) speculated that peripheral vision might be the major force driving the scanpath. However, eye movement studies on ambiguous and fragmented figures showed that the same physical stimulus results in different scanpaths, depending on the changing perceptual representation of the viewer (Ellis & Stark, 1978; Stark & Ellis, 1981). Therefore, peripheral vision may not play a major role in generating the scanpath. S. A. Brandt and Stark (1997) pointed out that since there was no actual diagram or picture in their visual imagery study, "[i]nput from foveal or peripheral vision cannot play a role in generating scanpath eye movements during imagery" (p. 32).

Although scanpath theory has not been studied extensively in recent years, several new studies have provided support for the scanpath theory. Pieters, Rosbergen, and Wedel (1999) found that scanpaths remain constant across advertising repetitions and across experimentally induced and naturally occurring conditions, and like S. A. Brandt and Stark (1997), Zangemeister, Oechsner, and Freksa (1995) and Gbadamosi, Oechsner, and Zangemeister (1997), also demonstrated firm evidence for scanpath sequences in the viewing of both real and imagined stimuli.

TESTING THE SCANPATH THEORY

In recent years, researchers have used Markov models and string-edit methods to test the scanpath theory. For example, Pieters et al. (1999) used Markov models to compare scanpaths of repeated exposures with print advertisements. Stark and Ellis (1981) also used Markov analysis to quantify the similarity of eye movements. S. A. Brandt and Stark (1997) applied string-edit analysis to compare the viewing pattern across an irregularly checkered grid displayed on a computer screen with the eye movements made while subjects imagined that particular grid. Using string-edit analysis, Zangemeister et al. (1995) and Gbadamosi et al. (1997) found evidence for scanpath sequences in their subjects' eye movements while similarly performing real viewing and visual imagery.

A Markov process is a stochastic model for the probabilities that the viewers' eyes will move from one visual element to another. The assumption is generally that scanpaths across visual elements can be described by a first-order Markov process—that is, each eye fixation depends only on the previous one. In addition to Markov dependence, there are two other constrained stochastic conditions that are possible: reversibility and stationarity. Reversibility means that saccades from Elements A to B occur as often as saccades from B to A (Ellis & Smith, 1985), and stationarity predicts that the scanpaths of viewers exposed repeatedly to the same visual stimulus will remain constant across exposures.

Pieters et al. (1999) concluded that scanpaths remain constant across repeated exposure to advertising stimuli and across experimentally induced and naturally occurring conditions. They concluded that scanpaths obey a stationary, reversible, first-order Markov process.

Using a different means of comparing scanpaths—the string-edit method—S. A. Brandt and Stark (1997) also found evidence supporting scanpath theory. Specifically, they found that eye movements during imagery are not random but reflect the content of the visualized scene. They concluded, therefore, that an "internalized, cognitive perceptual model must be in control of these scanpaths" (p. 32).

Abbott and Hrycak (1990) noted several advantages of string-edit methods for studying event sequences and outlined several limitations of Markovian sequence models. First and foremost, they argued, the sequence-generating process may have a longer history than the immediate past typically used in Markov analysis. Second, Markov models describe the stochastic processes that generate observed sequences and can be used to explore the goodness of fit of a predicted model but do not address the questions of whether there is a typical event sequence for a given process. Abbott and Hrycak argued that the direct testing of the Markov model—in terms of actual resemblance between generated and observed sequences—requires a technique for assessing similarity between sequences, categorizing sequences, and identifying typical sequences. String-edit analysis affords all of these techniques.

We tested Noton and Stark's (1971b) scanpath theory on different kinds of images widely used on the World-Wide Web: a news page, an advertising page, and a portal page. We compared recorded scanpaths by using a string-edit method, a technique that measures resemblance between sequences by means of a simple metric based on the insertions, deletions, and substitutions required to transform one sequence into another (Sankoff & Kruskal, 1983); this generates a *distance index*, or measure of dissimilarity.

Although several researchers, such as S. A. Brandt and Stark (1997) and Salvucci and Anderson (2001), have used string-edit methods to study eye path sequences, relatively few studies in which this method has been used have been reported, despite the fit between eye-tracking data and string-edit methodology. To our knowledge, this is the first study in which repeated exposures to Web page visual stimuli have been examined using eye tracking for measurement and string-edit methods for analysis.

String-edit sequence comparison, in conjunction with scaling and clustering techniques suited for the resulting proximity data, provides a visual display of intersequence distances and identifies clusters of similar sequences. If scanpaths are stable over repeated viewings and are not driven wholly by stimulus features (and are, therefore, variable across subjects), sequences for a given subject and stimulus should group together in neighborhoods in the scaling and should share cluster membership. Sequences for a given subject should not be in separate neighborhoods and clusters, although a given neighborhood or cluster may contain multiple subjects, reflecting the influence of stimulus features across subjects.

METHOD

Subjects

The subjects were 8 students at a large western university (4 males and 4 females). Their average age was 22.5 years. They were compensated for participating in the three-session study. All of the subjects were regular users of the Internet, reporting an average of almost 9 h a week of usage.

Apparatus

The eye movement data were collected using an ISCAN RK-426PC Pupil/Corneal Reflection Tracking System (ISCAN, Inc. [1998], Burlington, MA), which uses a corneal reflection system to measure the precise location of a person's eye fixations when he or she is looking at a visual display on a computer monitor. The eye-tracking equipment does not require attachments to the head (e.g., no bite bar or chinrest).

The ISCAN RK-426PC system collects data at 60 Hz, or about every 16.7 msec, and at a manufacturer-reported resolution of $\pm 0.3°$.

Fixation and Saccade Criteria

The minimum fixation duration was set at 100 msec. Although there are no definitive studies establishing 100 msec as the minimum amount of time necessary for a pause to be considered a fixation, 100 msec has become a widely accepted "rule of thumb" that numerous researchers believe produces reliable categorization of fixations from raw data. Researchers studying media images comparable to the Web pages in this study include Fischer, Richards, Berman, and Krugman (1989), who used this measurement in a study on print advertisements, and Baron (1980), who used it in a study on television. Lohse and Johnson (1996) used 100 msec as the minimum fixation duration in a marketing study. Researchers studying reading have concurred with this benchmark number (Stark, 1994). They have argued that a duration of 100 msec is a good criterion for distinguishing true fixations from corrective movements and eye drift.

In addition, the maximum shift that operationally defined a saccade in the present study was an area of 10×6 pixels, a measurement that was derived by slightly adjusting the manufacturer's suggested spatial settings, which were more precise. We increased the setting slightly because our study did not require such a small parameter.

Stimuli

The three Web pages used as stimuli were chosen for a number of reasons. Each page represents a distinct category of visual imagery on the Web. The portal page, used as a starting point for content on the Web, consists of a large number of hyperlinks and dialogue boxes for search functions and e-mail. The advertising page is highly visual and extremely colorful and is used to "build the brand" and sell the product. The news page displays mostly typography of various sizes for headlines, bylines, and body copy and is designed to convey information in an efficient manner.

The portal page and the on-line advertisement were completely contained in the first view, not forcing the viewer to scroll, thus simplifying the data analysis. On the news page, scrolling was required to view the entire news story, but the remainder of the Web page content remained stationary through the use of frames. Only the area containing the body of the story was scrollable.

Each subject's eye was about 22 in. (± 1 in.) from the screen of the computer monitor while he or she viewed the visual stimuli. Although each image was different, the stimuli subtended a visual angle of approximately 20° on the computer screen.

Specifically, the size of the visual stimuli viewed at a screen resolution of 800×600 was as follows: portal page, 774×407 pixels; on-line advertisement, 527×410 pixels; and news page, 768×401

A. Portal page   B. Advertising page   C. News page
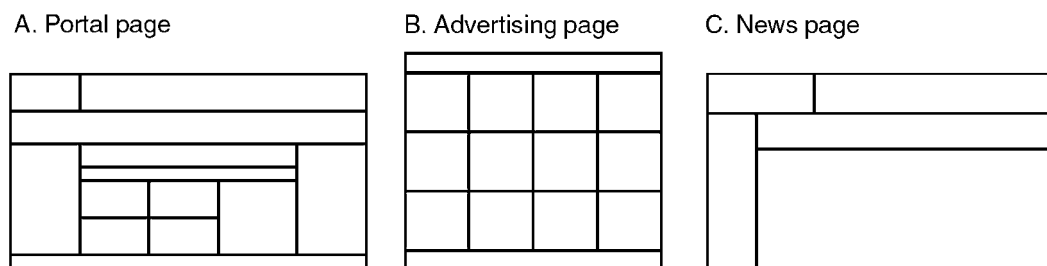

Figure 1. Target region grids for portal, advertising, and news Web page stimuli.

pixels in the first view and an additional 583 pixels of depth in the scrollable area. These three Web pages easily facilitated superimposing simple grids over their images (see Figures 1A, 1B, and 1C). This was necessary since string-edit analysis requires defining a sequence alphabet—in this case, a set of target regions in each stimulus. The Web pages selected were deemed to be relatively simple in layout, yet contained enough complexity that the sequential processing of the subfeatures could occur, thus producing the necessary sequential eye movements that define a scanpath. S. A. Brandt and Stark (1997) emphasized that the stimulus materials used in this sort of analysis required "a set of subfeatures whose positional encoding required careful review of the spatial layout by the subjects" (p. 34).

The use of grids was also logical because of the use of Web pages as stimuli. In the news and advertising fields, for example, it is usually more useful to know what elements are looked at and how frequently they are viewed than to know the exact coordinates of each fixation. In addition, defining target regions avoids the complex question of exactly how much of the visual field is perceived in each fixation.

However, since the three stimuli shown in Figure 1 differ considerably in the size, variety, and configuration of their target regions, the sequence comparisons can be performed only for scanpaths within each stimulus; cross-stimulus sequence comparisons are not meaningful. The differences in stimuli may also influence overall results in sequence similarity. A stimulus with a large region that may capture the extended attention of the viewer, such as the text region on the news page (Figure 1C), may result in lower sequence dissimilarities, owing to less variety in scanpaths.

Data Gathering

Each of the 8 subjects reported to the eye-tracking lab three different days separated by 48 h—on a Tuesday, a Friday, and a Monday. One week was selected as an appropriate length of time because basic research on scanpath theory finds stable effects, across exposure durations, of up to a week (Stark, 1994).

During each visit, the subjects viewed the same three Web pages while their eye movements were recorded. They viewed the Web pages in a different order of exposure on each visit.

At the lab, a subject was seated in front of the computer monitor and was told that he or she would be looking at three Web pages for a brief period of time. Each subject was instructed to imagine that he or she had been surfing the Web and had encountered the page. The subject was told to view the page as he or she would in that situation. The subjects were not given specific tasks in viewing the Web pages, because we were interested in scanpaths independent of directed search behavior, as might be found in casual Web surfing. Each subject was also told not to click on any links. After completing a task during which the scanning equipment was calibrated, the subject was shown a Web page. The calibration was checked after each subject had viewed the first two Web pages.

Eye movement data for a 15-sec exposure to each page were recorded. Previous research (Loftus, 1976) has established that eye fixations on essential information in a visual display occur within the first few seconds. S. A. Brandt and Stark (1997) gave their subjects 20 sec of viewing for familiarization and 10 sec for imagining. In this study, a pretest was used to determine that 15 sec was enough time for the subjects to examine the Web pages. Fifteen seconds also provided sufficient data to compare the eye paths with string-edit tests.

When the subjects returned on subsequent days, they were told to imagine they were surfing the Web and had chosen to revisit each site. They were instructed to view the pages as they would in a repeated visit to the site.

Procedure

Sequence comparison. The first step in comparing the eye path sequences was to define a sequence alphabet for each Web site. This was accomplished by assigning each target area on each Web page

an alphabetic code. The second step was to define the eye path sequence for each subject's viewing of each Web page by recording the sequence of fixations by the defined target area within which the fixation occurred (called "target tracing" by Salvucci & Anderson, 2001). For example, a viewing beginning with a single fixation in Area A followed by three fixations in Area C would generate a sequence beginning ACCC. . .

Optimal matching analysis (OMA) was used to compare these coded sequences. OMA is a generic string-edit tool for sequence comparison when each sequence is represented by well-defined elements drawn from a relatively small sequence alphabet (Holmes, 1997)—in this case, visual areas. OMA produces a numerical index—the Levenshtein distance—of the dissimilarity between any two sequences, computed as the smallest possible cost of elementary operations of insertion, substitution, and deletion of units required to align or transform one sequence into another (Abbott & Forrest, 1986; Sankoff & Kruskal, 1983; see Figure 2 and Figure 3). Similar sequences will, when compared, have smaller dissimilarity indexes; the more different two sequences, the greater the distance. Bradley and Bradley (1983) provide a formal definition of the OMA algorithm:

$$S^{ab} = \frac{S^{ab}_{mn}}{\max(m,n)},$$

where

$$S^{ab}_{ij} = \min\left[S^{ab}_{i-1,j-1} + D^{ab}_{if}, S^{ab}_{i-1,j} + C^a_i, S^{ab}_{i,j-1} + C^b_i \right],$$

with initial conditions

$$S^{ab} = 0, S^{ab}_{i0} = \sum_{k=1}^{i} C^a_k, S^{ab}_{0j} = \sum_{k=1}^{j} C^b_k$$

and, as adapted for scanpath comparison, S^{ab} = the distance from scanpath a to scanpath b, D^{ab}_{ij} = the substitution cost for fixation target i in scanpath a and fixation target j in scanpath b, C^a_k = the insertion cost and deletion cost for fixation target k of scanpath a, m = the length of scanpath a, and n = the length of scanpath b.

The dynamic programming implementation of this algorithm creates a table of string-edit costs in which one sequence defines the marginal left column and the other sequence defines the marginal top row of the table. Each cell contains four values: three subcell values (the substitution cost for the sequence elements defining that cell and their insertion and deletion costs) and the accumulated cost of the sequence alignment up to and including that element pair (Figure 4). Zhu and Ungar (2000) provided a succinct description of the dynamic programming operation:

> This table implicitly contains the edit cost of every permutation of edit rules required to transform one string into another. The table is a two-dimensional array whose values are calculated from the upper left diagonally to the bottom right. . . . The edit distance at each point is calculated from the surrounding values to the left and above the current position. These surrounding values represent the minimum edit distance required to reach that particular position within the two strings. The edit cost to obtain the current position from one of the surrounding positions is first calculated and then added to the value at that surrounding position. . . . Performing these operations produces three edit distances. The minimum edit distance is assigned to the current position. This algorithm guarantees that the minimum edit distance will always be the value at the lower right corner. (p. 2)

Given the range of insertion/deletion and substitution costs used in this study and the normalization of the distance index by the length of the longer of the two sequences, the dissimilarity index ranges from 0 for identical sequences to 1 for maximally dissimilar sequences. Additional examples, illustrations, and applications can be found in Kruskal (1983).

It should be noted that alignments may use a combination of substitutions and indels (insertions and deletions) to produce the Levenshtein distance. In their application of the string-edit method, S. A. Brandt and Stark (1997) set equal substitution costs for all pairs of sequence elements. In this study, substitution costs could
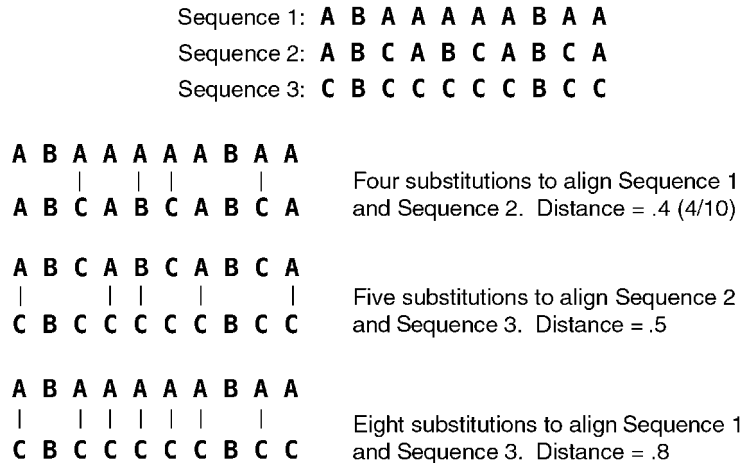
Sequence 1: **A B A A A A B A A**
Sequence 2: **A B C A B C A B C A**
Sequence 3: **C B C C C C B C C**

A B A A A A B A A
| | | | | Four substitutions to align Sequence 1
A B C A B C A B C A   and Sequence 2. Distance = .4 (4/10)

A B C A B C A B C A
| | | | | Five substitutions to align Sequence 2
C B C C C C B C C   and Sequence 3. Distance = .5

A B A A A A B A A
| | | | | | | Eight substitutions to align Sequence 1
C B C C C C B C C   and Sequence 3. Distance = .8

Figure 2. Sequence alignment or matching through insertions, deletions, and substitutions.

have been set in at least three ways: (1) at a value equal for all substitutions, following S. A. Brandt and Stark; (2) by distinctions between content forms of the regions, so that costs for regions in related content categories would be lower (e.g., top-level site navigation menus vs. page content navigation menus), whereas costs between categories would be higher (e.g., navigation regions vs. page text regions); and (3) by a measure of distance such that regions "closer" to each other could be substituted at less cost than regions further apart. We believed that uniform substitution costs lacked requisite variety; that is, uniform substitution costs treat the difference between two fixations in adjacent target areas as equivalent to the difference between two fixations in widely separated areas. The second approach was unsuitable for testing a theory of physical scanpaths, since the distance indexes would emphasize differences in content viewed rather than differences in the actual paths through which it was viewed. We applied the third approach; substitution values were based on the inverse of the physical adjacency of target areas—that is, the number of target regions in a direct path between the two points. For example, a pair of contiguous target areas, on a page for which the longest direct path between the centers of two target areas traversed five regions, would be assigned a substitution cost of .20. This assumes that for the purposes of comparing scanpaths, the difference between a fixation in a given target region and a fixation in a contiguous region is less than that between fixations in noncontiguous regions.

The contribution to the Levenshtein distance by the length of the compared eye path sequences (defined by the number of fixations in each) is an issue that has to be considered in OMA. To adjust for the role of sequence length in shaping the total cost of alignment, the intersequence distance was determined by dividing the raw sum alignment cost by the length of the longer sequence in the sequence pair. This made the distance relative to length and comparable across pairs of varying lengths.

Next, WinPhaser software (Holmes, 1996) was used to generate a sequence distance matrix of distance indexes for each possible pair of sequences for each stimulus. WinPhaser's OMA package uses a dynamic programming algorithm by Andrew Abbott of the University of Chicago. UCINET software (Borgatti, Everett, & Freeman, 1992) was used to perform nonmetric multidimensional scaling and hierarchical cluster analysis on the distance matrices. Scaling arranges the sequences in *n*-dimensional space such that the spatial arrangement approximates the distances between sequences; cluster analysis helps to define *neighborhoods* of similar cases within that *n*-dimensional space.

RESULTS

Twenty-four scanpath sequences were generated for the portal stimulus and for the advertising stimulus (three se-
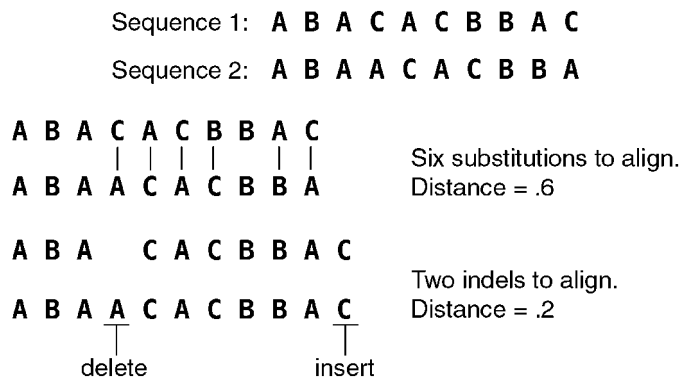
Sequence 1: **A B A C A C B B A C**

Sequence 2: **A B A A C A C B B A**

A B A C A C B B A C
  | | | | | | Six substitutions to align.
A B A A C A C B B A   Distance = .6

A B A   C A C B B A C

A B A A C A C B B A C   Two indels to align.
      ⊤   ⊤   Distance = .2
    delete   insert

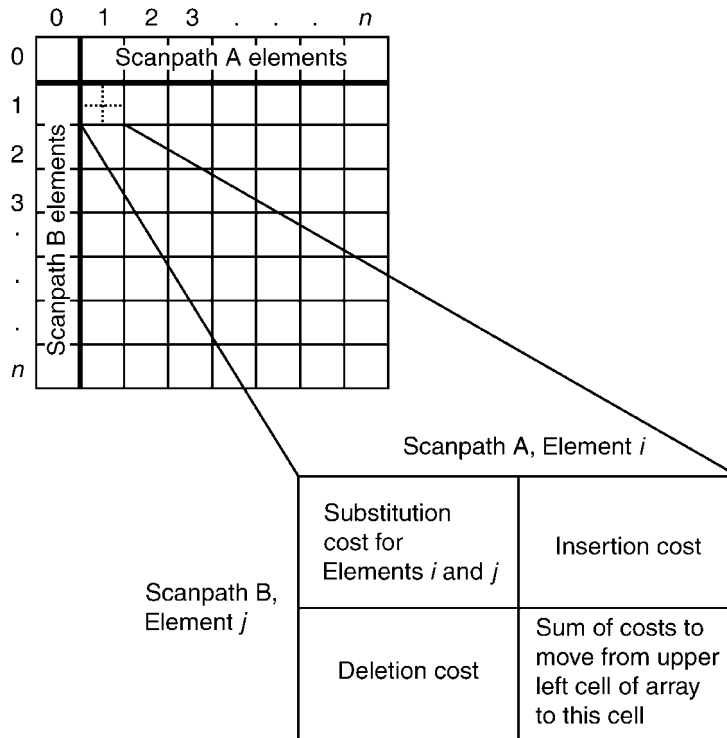**Figure 3. Levenshtein distance: the lowest cost alignment.**

Figure 4. Array used in the dynamic programming algorithm for optimal matching analysis, with example of cell subvalues.

quences for each of 8 subjects). There were 23 sequences generated for the news stimulus; 1 subject's second viewing of the stimulus was not recorded owing to equipment failure. Scanpath sequences for the portal stimulus ranged in length from 21 to 62 fixations ($M = 39.5, SD = 10.3$). Scanpath lengths for the advertising stimulus ranged from 21 to 54 fixations ($M = 39.3, SD = 7.9$). The scanpaths for the news stimulus ranged from 22 to 70 fixations in length ($M = 49.8, SD = 13.5$).

Sequence distances for the portal stimulus scanpaths ranged from .27 to .75 ($n = 276, M = .47, SD = .09$); the advertising stimulus scanpaths ranged from .33 to .74 ($n = 276, M = .50, SD = .08$). The news stimulus scanpaths, although showing lower overall distance indexes ($n = 253, M = .41, SD = .16$), produced a greater range of sequence distances, from .06 to .74. As was noted above, if the viewer reads the text block, the scanpath sequence settles into a series of fixations in the same target region. Pairwise comparisons of such sequences generate lower distance indexes, because few operations are required to align the sequences.

The distance indexes for the multiple viewings of each stimulus by each subject are displayed in Table 1. Two features should be noted: First, the results for the news stimulus reveal consistently low distance indexes for viewings by 3 subjects. Their scanpaths are dominated by fixations in the text region of the stimulus, suggesting close reading. In contrast, other subjects devoted more initial fixations to regions surrounding the text or continued to move

between the text region and other regions, rather than having their attention captured by the text.

Second, for each stimulus, the distance index for the comparison of the second and the third viewing tends to be the lowest of the three possible within-subjects comparison pairs (first to second viewing, second to third viewing, first to third viewing), suggesting an initial drift and subsequent stabilizing of the scanpath over time.

A 3×3 repeated measures analysis of variance (ANOVA) was performed on the distance indexes reported in Table 1. Mauchly's test of sphericity was not statistically significant for stimulus [portal, advertising, and news; $W(2) = .632, p = .317$] or for comparison pair [first vs. second viewing, second vs. third viewing, and first vs. third viewing); $W(2) = .494, p = .171$] but was statistically significant for the interaction of stimulus type and comparison pair [$W(9) = .011, p = .026$]. Owing to this finding, the Greenhouse–Geisser correction was applied in the ANOVA. The main effect of stimulus was statistically significant [$F(2,12) = 11.676, \epsilon = .731, p = .007$]. The main effect of comparison pair was also significant [$F(2,12) = 5.80, \epsilon = .664, p = .031$]. The interaction of stimulus type and comparison pair was not statistically significant [$F(2,12) = 0.156, \epsilon = .433, p = .829$].

Post hoc pairwise comparisons of stimuli, with Bonferroni correction for multiple comparisons, revealed a statistically significant difference between the distance index means for the advertising stimulus ($M = .484, SD = .046$), and the news stimulus ($M = .305, SD = .305$; dif-

Table 1
Sequence Distance Indexes by Scanpath Pairs
for Repeated Viewings of Portal, Advertising,
and News Stimuli by Each Subject

Subject	1st to 2nd	2nd to 3rd	1st to 3rd	Mean
		Portal Stimulus		
1	.32	.43	.40	.38
2	.37	.37	.42	.39
3	.46	.34	.51	.44
4	.44	.33	.42	.40
5	.46	.58	.41	.48
6	.44	.27	.46	.39
7	.46	.38	.38	.41
8	.49	.41	.51	.47
M	.43	.39	.44	
		Advertising Stimulus		
1	.57	.44	.60	.54
2	.43	.39	.48	.43
3	.48	.44	.47	.47
4	.57	.57	.59	.57
5	.63	.34	.66	.54
6	.43	.47	.49	.46
7	.46	.37	.47	.43
8	.44	.46	.52	.47
M	.50	.44	.53	
		News Stimulus		
1	–	–	.16	–
2	.14	.19	.17	.16
3	.42	.50	.43	.45
4	.39	.34	.41	.38
5	.49	.14	.48	.37
6	.43	.25	.52	.40
7	.20	.25	.10	.18
8	.21	.22	.13	.18
M	.32	.27	.30	

straints on viewing differences among subjects. The news stimulus (Figure 1C), in contrast, contains a large textual region that may, as was noted, drive greater similarity between subjects in their viewing of the page.

Post hoc tests showed no statistically significant pairwise differences among the comparison pairs; however, if paths stabilize over time, distance indexes should follow a pattern of (distance of viewing 1 vs. viewing 3) > (distance of viewing 1 vs. viewing 2) > (distance of viewing 2 vs. viewing 3). This was supported by a within-subjects polynomial contrast that revealed a statistically significant linear trend for comparison pairs [$F(1,6) = 6.817, p = .040$].

Plots of the multidimensional scaling solution in two dimensions are displayed for the portal stimulus (Figure 5), the advertising page stimulus (Figure 6), and the news page stimulus (Figure 7). These are spatial representations of the structure of the distance indexes; each sequence is represented by a point in a multidimensional space. The points are arranged so that the distances between all pairs of points are an optimal fit to the dissimilarities of the sequence pairs. Similar sequences are represented by points that are close together, whereas more dissimilar sequences are represented by points that are farther apart.

The two-dimensional solution to multidimensional scaling was used for convenient display because we were more interested in recognizing neighborhoods than in defining dimensions. The figures also indicate the most central eye path sequence (i.e., the sequence or sequences with the least mean distance from other sequences in the multidimensional scaling solution). In addition, in each set, the most similar sequence pairs are noted, as well as the two most dissimilar sequence pairs.

It is worthwhile to examine several of these cases, since they underscore the operation of the string-edit method used here and aid in the interpretation of the results. Figure 8 displays eye path sequences for four viewings of the
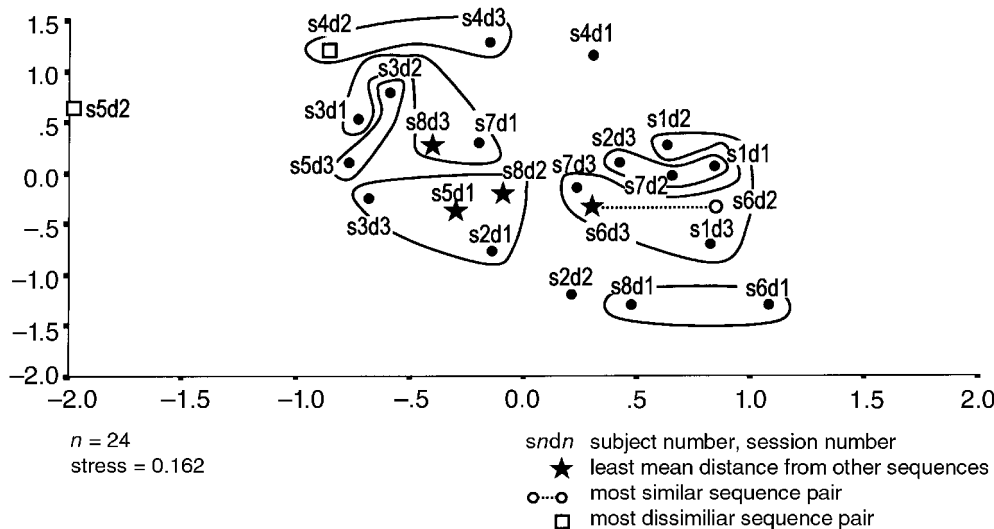
ference = .179, p = .012). The advertising stimulus (Figure 1B) is a regular grid of independent images and offers, of the three stimuli, the least content-based impetus to top–down, left–right viewing, thereby placing fewer con-

Figure 5. Multidimensional scaling solution in two dimensions for the portal page stimulus, with cluster analysis results superimposed.
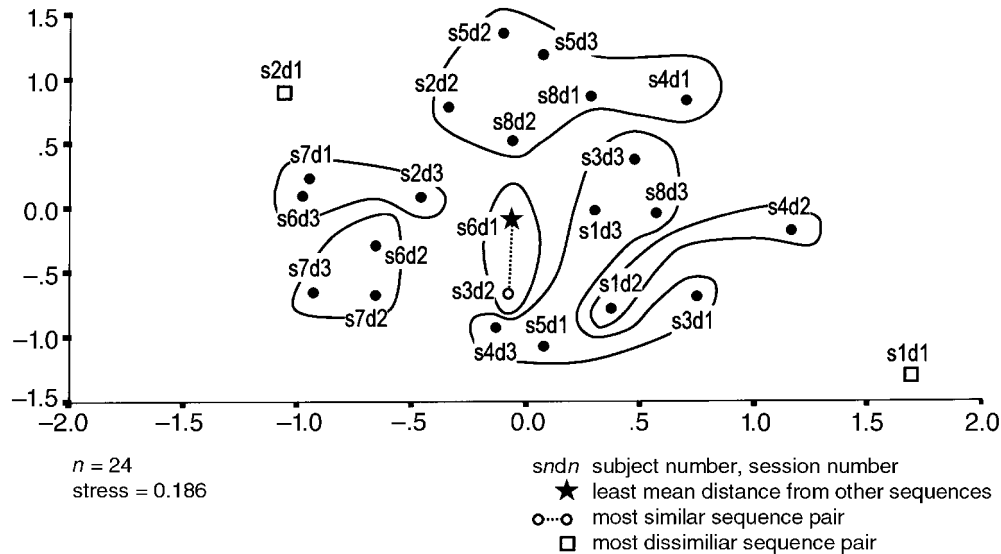
Figure 6. Multidimensional scaling solution in two dimensions for the advertising page stimulus, with cluster analysis results superimposed.

portal stimulus. According to our string-edit results, the paths for Subject 4 and Subject 5 are the most dissimilar in the set. Visual inspection suggests that this stems from different attention to upper and lower areas of the display. Conversely, the lower two paths in Figure 8 are more similar; they are from the same person (Subject 8) and represent "central" sequences in the set. They are characterized by more similarity than the previously noted paths and show attention distributed across the width of the middle tier of regions.

Visual examination of the spatial arrangement of the sequences reveals support for scanpath stability. In the case of the portal stimulus (Figure 5), for Subjects 1, 3, 4, 6,

and 7, we find the eye paths from the three separate viewings colocated in relatively small areas. In addition, two of the three eye path sequences are neighbors for Subject 2. The advertising page (Figure 6) stimulus reveals somewhat less stability of paths, since only Subjects 2, 6, and 8 appear to form small neighborhoods for all three sequences. The news page (Figure 7), which invites top–down, left–right processing of its textual contents, displays tighter colocation of most cases. This is especially notable for Subjects 2, 7, and 8; however, Subjects 1, 4, and 5 all provide a pair of closely located sequences as well.

Cluster analysis results provide another approach to discerning neighborhoods of similar sequences and are
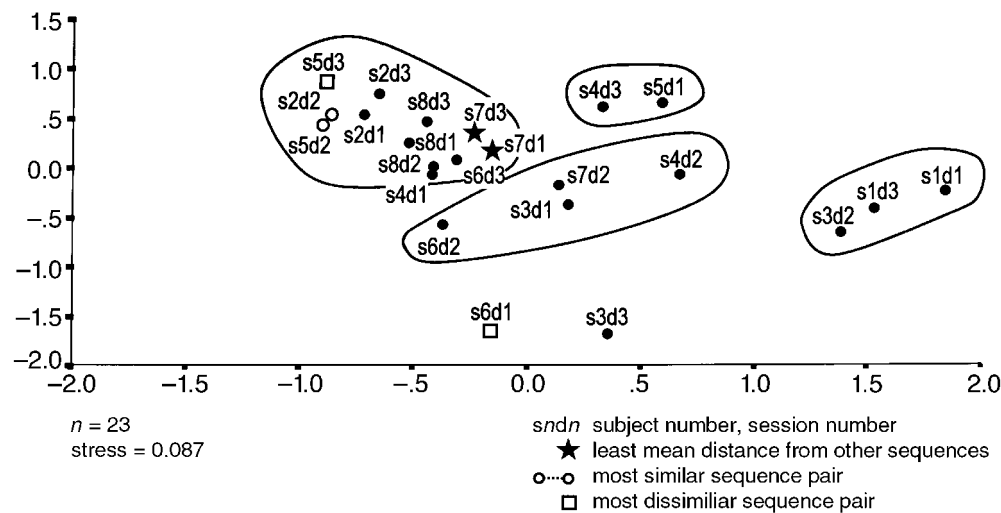

Figure 7. Multidimensional scaling solution in two dimensions for the news page stimulus, with cluster analysis results superimposed.

Subject 4, Day 2

Subject 5, Day 2

Subject 8, Day 2
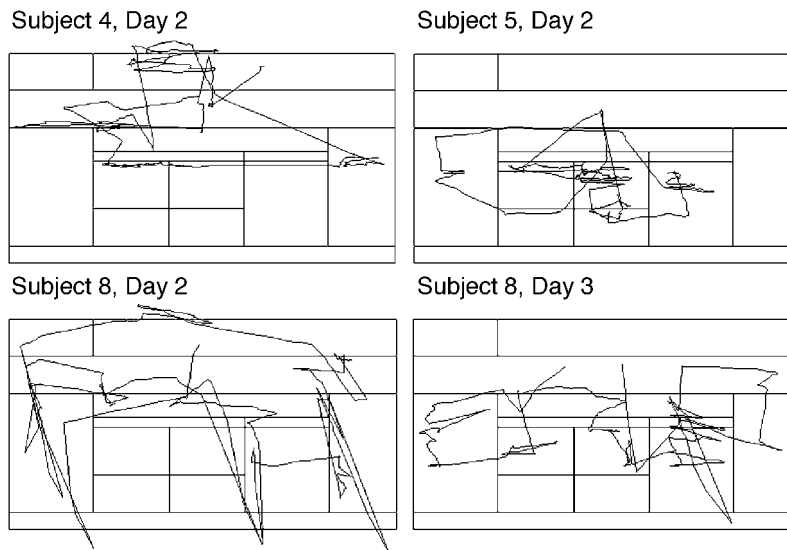
Subject 8, Day 3

Figure 8. Example eye path sequences for the portal stimulus.

superimposed on the multidimensional scaling results in Figures 5, 6, and 7. The hierarchical clustering of the portal sequences yielded seven clusters at Step 13 of the 23-step clustering solution. Three sequences had not joined a cluster at this step. The advertising sequences formed six clusters at Step 14 of the 21-step solution with two sequences not clustered, and the news sequences formed four clusters at Step 16 of the 21-step solution with two sequences not clustered.

The clustering results suggest that the three-dimensional solution may have been superior, since the "worm-shaped" clusters in some of the diagrams may reflect the greater stress of forcing the items into the less-optimal fit of two-dimensional space (i.e., a cluster that would fit into a small three-dimensional space is distorted or stretched in a two-dimensional representation). The clustering reveals families of sequences across subjects; note, for example, the cluster composed of a pair of eye path sequences from Subject 1 and a pair from Subject 6 in Figure 5 (portal stimulus). Similar features are found in the other plots. Figure 6 (advertising stimulus) reveals a larger cluster to be composed of sequences from 4 subjects (Subjects 2, 4, 5, and 8). In Figure 7 (news page stimulus), the largest cluster includes all three viewings for 2 subjects, and two viewings for 2 additional subjects; yet one of the other clusters is composed of single sequences from 4 different subjects. Such cross-subject scanpath similarity may bear witness to the interplay of design features and individual scanpath preferences.

The eye path sequences for the news stimulus (Figure 7) are more similar to each other than are the most similar sequences for the portal or advertising stimuli, perhaps because the text region tended to keep the fixations once it had captured them. Given this "pull," remaining within-subjects differences in the scanpath for the news stimulus are interesting, since they may suggest either (1) lack of

stability, over time, of the person's scanpath (e.g., drift and subsequent stabilizing) or (2) a confounding memory effect wherein familiar textual material is scanned differently from new material (e.g., attention decrement).

In a similar fashion, the strong within-subjects resemblances between sequences for the portal and the advertising stimuli, which in design are less governed by left–right/top–down conventions, are notable, since they suggest either (1) scanpath stability or (2) particular page features tending to capture attention in the same general sequence. Note, however, that when different subjects show high within-subjects scanpath resemblance, but nevertheless with marked difference between subjects, we can conclude personal scanpath preferences do indeed have some explanatory power.

CONCLUSION

Our results are mixed. Some individuals show scanpaths that resemble each other over time. However, we also found many instances in which the most similar sequences were from different subjects, rather than from the same subject, suggesting strong stimulus influences. On the other hand, the clusters tend to include pairs of sequences from the same subject. The fact that clusters of sequences also typically contain paths from multiple subjects suggests that other forces may be important, such as features of the Web page or memory. These could be tested in future research with carefully manipulated page versions.

This study is descriptive and interpretive in nature. To have significance tests, we would need to determine a critical value for how much Levenshtein distance is needed before we consider two sequences to be significantly different. A significance test could be generated from Monte Carlo simulations in future research. Erickson and Sellers (1983) demonstrated how such a test operates: A large sam-

ple of random permutations of each sequence in a given pair is generated, and the distances between those permuted sequences are calculated. The mean and standard deviation of the resulting distance indexes allow characterization, at a given confidence level, that an achieved distance index represents differences in sequence order, rather than in sequence composition.

The linear trend to increased similarity across viewings suggests that paths appear to drift more from the first to the second viewing than from the second to the third viewing for a given subject. The following questions present themselves. Is there an increased tendency with the passing of time to ignore material on the top of Web pages? Do the paths become simpler on subsequent viewings? Do viewers dwell longer on selected regions of particular interest? Does task fatigue result in shifts in visual attention? Does a scanpath become increasingly habituated?

In summary, on the World-Wide Web, with somewhat complex digital images, some viewers' eye movements appear to follow a "habitually preferred path" across the visual stimulus, as asserted in the scanpath theory of Noton and Stark (1971b). Given the still-considerable variation in paths between and across subjects and the differences found across stimuli, much more research is needed to explore the influence of scanpaths, content, and form on sequences of eye movement.

REFERENCES

ABBOTT, A., & FORREST, J. (1986). Optimal matching sequences for historical sequences. *Journal of Interdisciplinary History,* **16**, 471-494.

ABBOTT, A., & HRYCAK, A. (1990). Measuring resemblance in sequence data. *American Journal of Sociology,* **16**, 144-185.

ANTES, J. R. (1974). The time course of picture viewing. *Journal of Experimental Psychology,* **103**, 62-70.

BARON, L. (1980). Interaction between television and child-related characteristics as demonstrated by eye movement research. *Education, Communication & Technology: A Journal of Theory, Research & Development,* **28**, 267-281.

BORGATTI, E. F., EVERETT, M., & FREEMAN, L. C. (1992). *UCINET IV* (Version 1.0) [Computer software]. Columbia, SC: Analytic Technologies.

BRADLEY. D. W., & BRADLEY, R. A. (1983). Application of sequence comparison to the study of bird songs. In D. Sankoff & J. B. Kruskal (Eds.), *Time warps, string edits, and macromolecules: The theory and practice of sequence comparison* (pp. 189-207). Reading, MA: Addison-Wesley.

BRANDT, H. F. (1940). Ocular patterns and their psychological implications. *American Journal of Psychology,* **53**, 260-268.

BRANDT, S. A., & STARK, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience,* **9**, 27-38.

BUSWELL, G. T. (1935). *How people look at pictures.* Chicago: University of Chicago Press.

ELLIS, S. R., & SMITH, J. D. (1985). Patterns of statistical dependency in visual scanning. In R. Groner, G. McConkie, & C. Menz (Eds.), *Eye movements and human information processing* (pp. 221-238). Amsterdam: Elsevier.

ELLIS, S. R., & STARK, L. W. (1978). Eye movements during the viewing of Necker cubes. *Perception,* **7**, 575-581.

ERICKSON, B. W., & SELLERS, P. H. (1983). Recognition of patterns in genetic sequences. In D. Sankoff & J. B. Kruskal (Eds.), *Time warps, string edits, and macromolecules: The theory and practice of sequence comparison* (pp. 55-91). Reading, MA: Addison-Wesley.

FISCHER, P. M., RICHARDS, J. W., BERMAN, E. F., & KRUGMAN, D. M. (1989). Recall and eye tracking study of adolescents viewing tobacco ads. *Journal of the American Medical Association,* **261**, 84-89.

GBADAMOSI, J., OECHSNER, U., & ZANGEMEISTER, W. H. (1997). Quantitative analysis of gaze movements during visual imagery in hemianopic patients and control subjects. *Journal of Neurological Rehabilitation,* **3**, 165-172.

HOLMES, M. E. (1996). *WinPhaser user's manual* (Version 1.0c) [computer software]. Muncie, IN: Author.

HOLMES, M. E. (1997). Optimal matching analysis of negotiation phase sequences in simulated and authentic hostage negotiations. *Communication Reports,* **10**, 1-8.

ISCAN, INC. (1998, January). *RK-726 PCI Pupil/Corneal Reflection Tracking System (PCI Card Version) Operating Instructions.* Burlington, MA: Author.

KRUSKAL, J. B. (1983). An overview of sequence comparison. In D. Sankoff & J. B. Kruskal (Eds.), *Time warps, string edits, and macromolecules: The theory and practice of sequence comparison* (pp. 1-44). Reading, MA: Addison-Wesley.

LLEWELLYN-THOMAS, E. (1968, August). Movements of the eye. *Scientific American,* **219**, 88-95.

LOFTUS, G. R. (1976). A framework for a theory of picture recognition. In R. A. Monty & J. W. Senders (Eds.), *Eye movements and psychological processes* (pp. 499-513). New York: Wiley.

LOHSE, G. L., & JOHNSON, E. J. (1996). A comparison of two process tracing methods for choice tests. *Organizational Behavior & Human Decision Processes,* **68**, 28-43.

MACKWORTH, N. H., & MORANDI, A. J. (1967). The gaze selects informative details within pictures. *Perception & Psychophysics,* **11**, 547-551.

NOTON, D., & STARK. L. W. (1971a). Scanpaths in eye movements during pattern perception. *Science,* **171**, 308-311.

NOTON, D., & STARK, L. W. (1971b). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research,* **11**, 929-942.

PARKER, R. E. (1978). Picture processing during recognition. *Journal of Experimental Psychology: Human Perception & Performances,* **4**, 284-292.

PIETERS, R., ROSBERGEN, E., & WEDEL, M. (1999). Visual attention to repeated print advertising: A test of scanpath theory. *Journal of Marketing Research,* **36**, 424-438.

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

RAYNER, K. (1995). Eye movements and cognitive processes in reading, visual search, and scene perception. In J. M. Findlay, R. Walker, & R. W. Kentridge (Eds.), *Eye movement research: Mechanisms, processes and applications* (pp. 3-22). Amsterdam: Elsevier.

SALVUCCI, D. D., & ANDERSON, J. R. (2001). Automated eye-movement protocol analysis. *Human–Computer Interaction,* **16**, 39-86.

SANKOFF, D., & KRUSKAL, J. B. (Eds.) (1983). *Time warps, string edits, and macromolecules: The theory and practice of sequence comparison.* Reading, MA: Addison-Wesley.

STARK, L. W. (1994). Sequences of fixations and saccades in reading. In J. Ygge & G. Lennerstrand (Eds.), *Eye movements in reading* (pp. 135-161), Tarrytown, NY: Pergamon.

STARK, L. W., & ELLIS, S. R. (1981). Scanpaths revisited: Cognitive models direct active looking. In D. F. Fisher, R. A. Monty, & J. W. Senders (Eds.), *Eye movements: Cognition and visual perception* (pp. 193-226). Hillsdale, NJ: Erlbaum.

YARBUS, A. L. (1967). *Eye movements and vision.* New York: Plenum.

ZANGEMEISTER, W. H., OECHSNER, U., & FREKSA, C. (1995). Short-term adaptation of eye movements in patients with visual hemifield defects indicates high level control of human scan path. *Optometry & Vision Science,* **72**, 467-478.

ZHU, J. J., & UNGAR, L. H. (2000). *String edit analysis for merging databases.* Retrieved August 10, 2002 from http://www.cis.upenn.edu/datamining/Publications/KDDzhu.pdf.