

# It depends on how you look at it: Scanpath comparison in multiple dimensions with MultiMatch, a vector-based approach

Richard Dewhurst · Marcus Nyström ·  
Halszka Jarodzka · Tom Foulsham · Roger Johansson ·  
Kenneth Holmqvist

Published online: 31 May 2012  
© Psychonomic Society, Inc. 2012

**Abstract** Eye movement sequences—or scanpaths—vary depending on the stimulus characteristics and the task (Foulsham & Underwood *Journal of Vision*, 8(2), 6:1–17, 2008; Land, Mennie, & Rusted, *Perception*, 28, 1311–1328, 1999). Common methods for comparing scanpaths, however, are limited in their ability to capture both the spatial and temporal properties of which a scanpath consists. Here, we validated a new method for scanpath comparison based on geometric vectors, which compares scanpaths over multiple dimensions while retaining positional and sequential information (Jarodzka, Holmqvist, & Nyström, *Symposium on Eye-Tracking Research and Applications* (pp. 211–218), 2010). “MultiMatch” was tested in two experiments and pitted against ScanMatch (Cristino, Mathôt, Theeuwes, & Gilchrist, *Behavior Research Methods*, 42, 692–700, 2010), the most comprehensive adaptation of the popular Levenshtein method. In *Experiment 1*, we used synthetic data, demonstrating the

greater sensitivity of MultiMatch to variations in spatial position. In *Experiment 2*, real eye movement recordings were taken from participants viewing sequences of dots, designed to elicit scanpath pairs with commonalities known to be problematic for algorithms (e.g., when one scanpath is shifted in locus or when fixations fall on either side of an AOI boundary). The results illustrate the advantages of a multidimensional approach, revealing how two scanpaths differ. For instance, if one scanpath is the reverse copy of another, the difference is in the direction but not the positions of fixations; or if a scanpath is scaled down, the difference is in the length of the saccadic vectors but not in the overall shape. As well as having enormous potential for any task in which consistency in eye movements is important (e.g., learning), MultiMatch is particularly relevant for “eye movements to nothing” in mental imagery and embodiment-of-cognition research, where satisfactory scanpath comparison algorithms are lacking.

---

R. Dewhurst (✉) · M. Nyström · K. Holmqvist  
Humanistlaboratoriet, Lund University,  
Helgonabacken 12, P.O. Box 201, 22100 Lund, Sweden  
e-mail: richard.dewhurst@humlab.lu.se

H. Jarodzka  
Centre for Learning Sciences and Technologies,  
Open Universiteit Nederland,  
Valkenburgerweg 177, P.O. Box 2960, 6401 DL Heerlen,  
The Netherlands

T. Foulsham  
Department of Psychology, University of Essex,  
Colchester CO4 3SQ, UK

R. Johansson  
Department of Cognitive Science, Lund University,  
Helgonabacken 12,  
222 22 Lund, Sweden

**Keywords** Scanpath · Similarity · Eye movement ·  
Sequence · Dimensional · Comparison

Whether passively viewing a picture or engaged in a higher-level cognitive task, we make sequences of eye movements commonly known in the literature as *scanpaths*. Recent years have seen rapidly growing interest in scanpaths (Burmester & Mast, 2010; Day, 2010; Mannan, Kennard, & Husain, 2009; Underwood, Humphrey, & Foulsham, 2008) because, unlike unitary eye movement events (such as fixations and saccades, as well as the measures that can be derived from them), scanpaths encompass a whole range of oculomotor data into one construct, which reveals visual processing over time and in space. However, precisely because scanpaths comprise both spatial and temporal information, this has led to major computational challenges in comparing one scanpath to another

(e.g., Coco, 2009; Foulsham & Underwood, 2008). In this article, we evaluate the strength of a new method for scanpath comparison, MultiMatch, that incorporates spatial and temporal information together, using geometric vectors to represent saccades and fixations. The basic principles of MultiMatch have been previously outlined in Jarodzka, Holmqvist, and Nyström (2010). Here, we expand on our previous discussion, testing the performance of the algorithm with real and simulated data of the type commonly encountered in research on visual perception and cognition.

### Scanpath comparison: Methods that rely on areas of interest (AOIs)

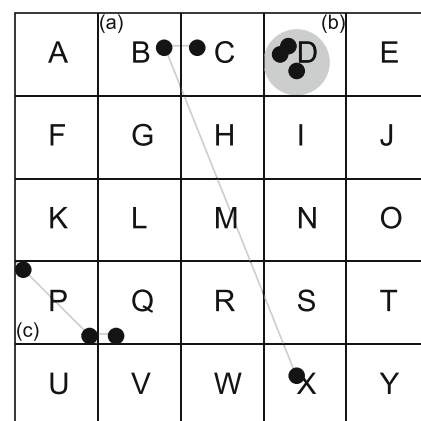
*Levenshtein distance* Interest in scanpaths dates back to the early studies of Noton and Stark (1971), who recorded eye movements during pattern perception and found a striking resemblance between the sequence of fixations produced during initial inspection (encoding) and subsequent presentation (recognition) of the same image. Quantifying this “resemblance,” Brandt and Stark (1997) went on to utilize a measure of scanpath similarity that had originally been implemented for eye movement research by Hacısalihzade, Stark, and Allen (1992): the Levenshtein distance principle, used to identify commonalities in strings of symbols for DNA sequence matching (Levenshtein, 1966). The basic principle of Levenshtein distance, or *string-edit*, as it is also known, is that fixation sequences are first represented by overlaying discrete AOIs onto stimulus space so that the locations of fixations can be replaced with characters corresponding to the AOI positions occupied. Scanpaths are thus reduced to strings of symbols that retain information about the sequential order of the fixations and provide an approximation of their spatial position. The similarity between two such strings can then be calculated as the “editing cost” of transforming one string into the other via *insertion*, *deletion*, and *substitution* of constituent characters. The Levenshtein *distance* is then the least number of editing steps required, while *similarity* is often expressed as 1 minus this number (normalized over scanpath length or the number of elements in the string).

While the Levenshtein technique is relatively powerful from the point of view of capturing information pertaining to the *order of fixations* within a scanpath, it suffers from a number of problems that have been documented in the literature but to date not adequately solved.

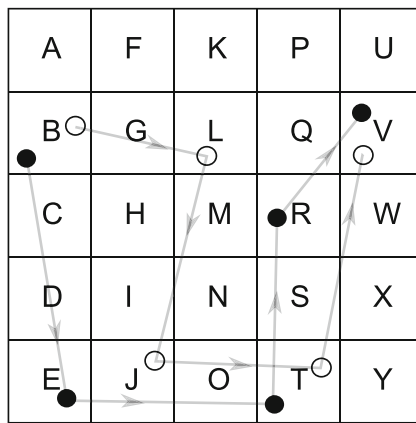
These problems revolve around the division of stimulus space into AOIs. Owing to this, the original *positions of fixations are not well represented*. This lack of spatial resolution can have drastic consequences for the similarity metric produced. For example, two fixations close to one

another in proximity but that fall on either side of an AOI border will be judged to be as dissimilar as two fixations in completely different regions of the stimulus. Likewise, fixations occupying the same AOI are often grouped such that a cluster of fixations in one spatial region become represented by only one letter. This has the effect of ignoring local differences in eye movement control and ongoing cognitive processing, in favor of retaining a better representation of the global characteristics of the scanpath. Moreover, depending on the sizes of the AOIs and how they are chosen, situations could easily arise in which two fixations, maximally distant within one area, are calculated to be more similar to each other than they are to other fixations that are closer in Euclidean distance but that fall outside of the AOI’s strict perimeter. The problems associated with using AOIs are common to many of the approaches to scanpath comparison (see Fig. 1).

*ScanMatch* ScanMatch (Cristino, Mathôt, Theeuwes, & Gilchrist, 2010) has emerged recently as a more advanced adaptation of Levenshtein distance, addressing some of the problems to do with AOIs. The strength of ScanMatch lies in its use of the Needleman–Wunsch algorithm to align and compare eye movement sequences. If we take two letter strings that represent the sequences of fixations within AOIs, BETRV and BLJTV (Fig. 2), the first step is to align one string with the another. The Needleman–Wunsch algorithm does this by using a *substitution matrix* containing all potential pairings of elements in the two strings. Aligning B with B, for instance, would be a perfect match obtaining the highest possible score in the matrix, whereas the cost of aligning mismatches such as E and L can be adjusted according to the relationship between the AOIs that these



**Fig. 1** Problems with AOIs and with the string-edit method. **a** Fixations in regions B and C are close in space, but are judged to be as different as those in B and X. **b** All three fixations in D are represented by just one letter. **c** The fixations in P span the farthest distance within the AOI, and though the fixation in Q is closer to one of them, it will be regarded as more different than the other fixation in P, according to the Levenshtein similarity metric



**Fig. 2** Two scanpaths represented by the strings BETRV (filled circles) and BLJTV (unfilled circles)

letters represent. For example, they might be close to each other in spatial proximity, have some basic, low-level visual property like luminance in common, or contain similar semantic content. A *comparison matrix* can then be created in which the sequence of letters in one scanpath forms the column divisions, and the sequence of letters in the other forms the row divisions. The comparison matrix uses the substitution matrix as a lookup table to find the entries (costs) for each of its cells. The Needleman–Wunsch algorithm then searches for the optimal route through this matrix (i.e., yielding the highest score from the top left corner to the bottom right). This score, normalized for the length of the sequence, is the similarity result between 0 and 1, where 1 indicates identical scanpaths according to AOI visits and the substitution matrix.

If we refer back to our letter-string example (Fig. 2), Levenshtein distance treats all differences between strings equally. The editing cost of replacing E with L, inserting J, and deleting R in the first sequence, to align it with the second, is the same in all cases, and this would translate to the same value in the substitution matrix for all cells except the diagonal, where there is no cost associated with identical letters. ScanMatch, on the other hand, can take both physical and cognitive characteristics of the stimulus into account when creating the AOI-based substitution matrix. For instance, E and L may be the same color, and the substitution matrix could be designed accordingly, attributing less cost to alignments where AOIs containing the same color are visited. This is a large advantage in comparison with Levenshtein distance, which is effectively blind to the stimulus being viewed, because no relationship between the AOIs is specified. However, care must be taken in how the scoring of the substitution matrix is devised. The weightings cannot be subjective, but should quantitatively reflect the connection (visual or semantic) between AOIs in the stimulus.

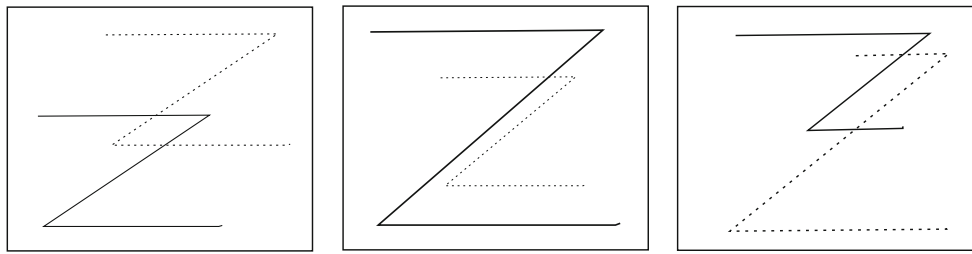
However, whilst ScanMatch is a much needed improvement to the Levenshtein method, because it rests on the

same principles of carving up stimulus space into AOIs, similar criticisms may be leveled at this new technique for comparing scanpaths. While allowing the researcher to specify the relationship between AOIs is a worthwhile advancement, one may still question this way of segmenting the recorded eye movement data. As with Levenshtein, two maximally distant fixations within one AOI will still be treated as more similar to each other than they are to fixations that are closer in Euclidean distance but that occupy a different AOI (even if distance is taken into account in the substitution matrix; see Fig. 1c). This is because ScanMatch does not consider *where* in an AOI a fixation lands.

Likewise, the *shapes of the scanpaths* being compared are neglected in the similarity calculations of both the Levenshtein method and ScanMatch: One scanpath could produce exactly the same shape as another in terms of the angular distances between saccades and the saccadic amplitudes, but because the fixations fall outside AOI limits, the scanpaths would not be recognized as similar (Fig. 3).<sup>1</sup> Shape has been clearly shown to be a crucial attribute of scanpaths in the mental imagery research (Johansson, Holsanova, & Holmqvist, 2006, 2011; Laeng & Teodorescu, 2002; Zangemeister & Liman, 2007). Johansson et al. (2006) found that when orally describing a picture while looking at a blank screen, participants' eye movements matched those made during the original encoding of the picture. Moreover, this effect was equally strong for visual encoding (looking at a picture's spatial layout) and aural encoding (listening to a description of a picture's spatial layout). Critically, however, the correspondence between the scanpaths at encoding and the subsequent imagery was often in terms of shape but not absolute spatial position. Scanpaths could be scaled (shrunk or enlarged) during imagery but otherwise remain consistent. Similarly, scanpaths could assume the same shape during imagery, but be spatially offset (recentered or shifted relative to a new locus). Johansson and colleagues (2006, 2011; see also Johansson, Holsanova, Dewhurst, & Holmqvist, 2011) used a coarse measure of relative saccade direction and fixation distribution to tackle this issue, but it is evident that a more stringent measure of scanpath shape is desirable.

A caveat is necessary here, however. While there are advantages to studying scanpath shape in the context of mental imagery research, and in light of aspects of neurological visual disorders, it would not be appropriate to let similarities in scanpath shape dominate in the absence of a clear hypothesis. For example, if comparing monitors of different sizes or big-screen projections of images, it would

<sup>1</sup> The example gives extreme cases of scanpath shape similarity in order to exemplify the point. In reality, such occurrences would be rare, and in the majority of cases one should question similarity on the basis of shape *alone*, because this would mean that different regions in space were foveated.



**Fig. 3** Scanpaths can form similar shapes with completely different fixation positions. The panels show a scanpath (solid lines) and its comparison pair (dotted lines). The locus can be shifted (left panel), a scanpath can be scaled (center panel), or the two scanpaths can comprise

different saccadic amplitudes (right panel). From *Eye Tracking: A Comprehensive Guide to Methods and Measures*, by K. Holmqvist et al., 2011, p. 351; Oxford, U.K.: Oxford University Press. Copyright 2011 Oxford University Press. Reprinted with permission

be appropriate to predict similarity in shape instead of absolute space, but if comparing eye movements on different images of constant size, the absolute spatial positions of fixations (and their order) become more important.

Nevertheless, when imposing restricted AOIs onto the stimulus, it is inevitable that basic spatial information about the scanpath will be compromised (see Jarodzka, Holmqvist, & Nyström, 2010): The locations of individual fixations are generalized to extended areas, restricting the spatial resolution of the final similarity calculation. This also has the effect of distorting scanpath shape to the extent that this representation is lost in the final similarity metric.

### Scanpath comparison: Methods that do not rely on AOIs

**Mannan linear distance** The Mannan linear-distance approach to scanpath comparison (Mannan, Ruddock, & Wooding, 1995, 1996) directly compares fixation positions without the use of AOIs, thereby not quantizing space, as is the case with both Levenshtein distance (e.g., Brandt & Stark, 1997) and ScanMatch (Cristino et al., 2010). It takes as input two sets of position data in  $x, y$  space:  $A$  and  $B$ , each having specified  $M$  and  $N$  locations. The Euclidean distance ( $d$ ) between each point ( $i$ ), of a possible  $M$  in  $A$ , and its nearest neighboring point ( $j$ ), of a possible  $N$  in  $B$ , is then computed. The resulting output is a “similarity index” that can be expressed as the mean of the  $d$ s squared:

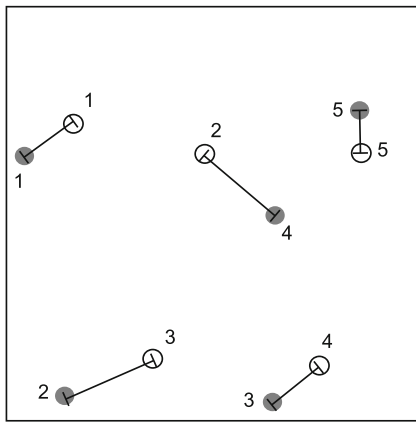
$$D_{Mannan} = \frac{1}{M+N} \left( \sum_{i=1}^M \min d^2_{i,j} + \sum_{j=1}^N \min d^2_{i,j} \right) \quad (1)$$

Basically, the linear-distance method simply quantifies how close positions are to each other. It has the advantage of comparing absolute points in space, so the similarity between the veridical positions of fixations can be directly calculated (though, equally, raw sample data from the eye-tracking system could in theory also be used). However, what is gained in spatial resolution is lost in sensitivity to the

more fundamental dimensions of a scanpath. Because each position is associated with its nearest neighbor, the impact of individual fixations can be disproportionate. For example, a cluster of several local fixations in Scanpath Sequence 1 can be mapped onto just one fixation in Scanpath Sequence 2. A high similarity score could then be returned, despite an overall impression of two very different scanpaths. Such concerns have been raised in the literature (see, e.g., Henderson, Brockmole, Castelano, & Mack, 2007; Tatler, Baddeley, & Gilchrist, 2005; Underwood et al., 2008), such that the original formulation of Mannan’s linear distance is now hardly used. Henderson et al. developed a “unique assignment” version of the similarity index to address these concerns, where each fixation in set  $A$  is linked to a single mate in set  $B$ —the pairings producing the lowest variance being selected for Eq. 1 (Fig. 4). Nevertheless, while order *might* be assumed for scanpaths containing few fixations, as the sequence grows in length, it is not improbable that fixations will be matched irrespective of temporal order (the 4th fixation in Scanpath 1 of the figure is paired with the 2nd fixation in Scanpath 2). Moreover, unique assignment requires that the two scanpaths be of equal length (i.e., contain the same number of fixations), which they commonly would not be.

Recently, however, Mathôt, Cristino, Gilchrist, and Theeuwes (2012) tried to rectify some of these shortcomings. Instead of mapping one single point to one other point, they advised double mappings, to provide the lowest overall position variability. This overcomes the problem of the necessary pruning of sequences with unique assignments when the two scanpaths are not of equal length (but may still lead to several points in one scanpath being mapped to just one in the other). Moreover, Mathôt et al.’s “Eyenalysis” procedure allows for some dimensionality in scanpaths to be considered: the duration of the fixation points, for instance, or their timestamps, which—although they are not exact measures of order—do allow the temporal characteristics of the sequence to be taken into account.

Taking the actual coordinates of the recorded eye movement data is clearly an advantage of Mannan’s linear-distance

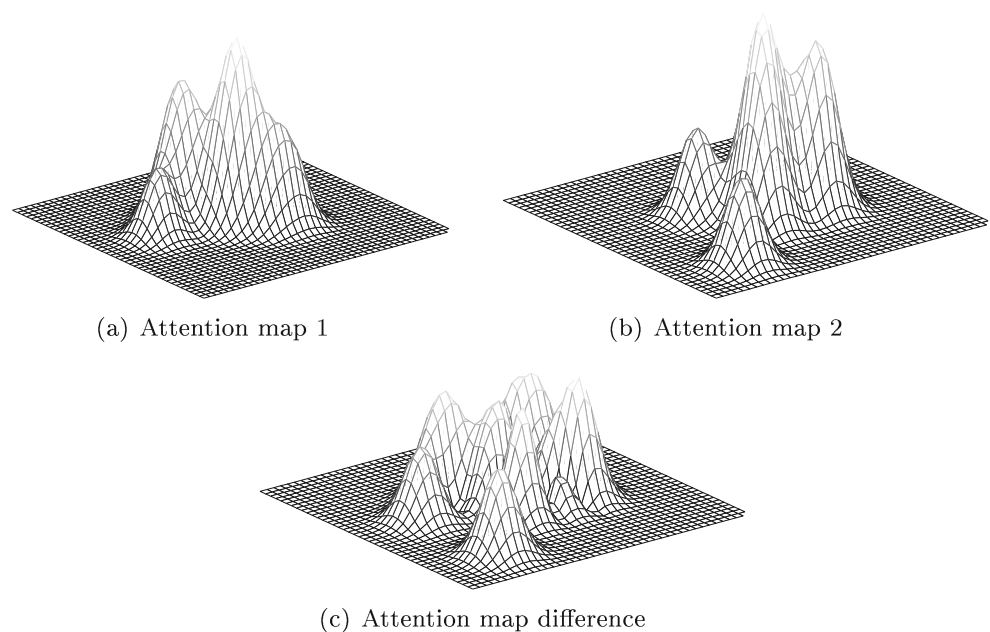


**Fig. 4** Mannan similarity, with unique assignment. Of the same two scanpaths from Fig. 2, the numbers indicate the orders of the fixations from the first scanpath (filled circles) and the second (open circles). The lines show the minimum linear distance between the matched fixations of each scanpath. Even with just five fixations, the sequence comparison is compromised

method and its derivatives. However, what is gained in spatial resolution is lost in sensitivity when matching fixations, and it becomes very difficult to retain sequential information from the scanpaths, which is one of their defining characteristics.

*Attention/heat maps* Some have argued that attention maps, and the family of comparison measures that can be derived from them, capture scanpath similarity (e.g., Caldara & Miellet, 2011; Grindinger, Duchowski, & Sawyer, 2010). With attention maps, there is not an issue of which fixation point to map onto another for position comparisons, since a smooth Gaussian landscape is produced from the fixations that make up each scanpath (Fig. 5), and from here one may take the attention map difference (Wooding, 2002) as a

**Fig. 5** Two attention maps comprising fixations (a and b) and the difference between them (c)



measure of scanpath similarity (though, equally, more complex measures such as the correlation coefficient between attention maps [Ouerhani, von Wartburg, Hügli, & Müri, 2004], the Kullback–Leibler distance [Rajasekar, Cormack, & Bovik, 2004], or the earth mover distance [Dempere-Marco, Hu, Ellis, Hansell, & Yang, 2006] may also be used). However, despite their computational strength, attention maps, like the Mannan similarity index (e.g., Mannan et al., 1996), struggle to effectively capture the sequence of fixations inherent to a scanpath. There are ways that this issue can be tackled—for example, by weighting the attention map by first entry time, so that peaks in the attentional landscape reflect regions in the stimulus that were visited earlier (see Holmqvist et al., 2011, p. 233). Trying to preserve ordinal information in this way, however, necessitates the use of AOIs, because the researcher must specify the time that it took to reach a spatially restricted area. As we have seen, this introduces quantization noise. Moreover, neither attention maps nor the Mannan similarity index tackle the limitation of existing methods already pointed out—that of representing scanpath shape.

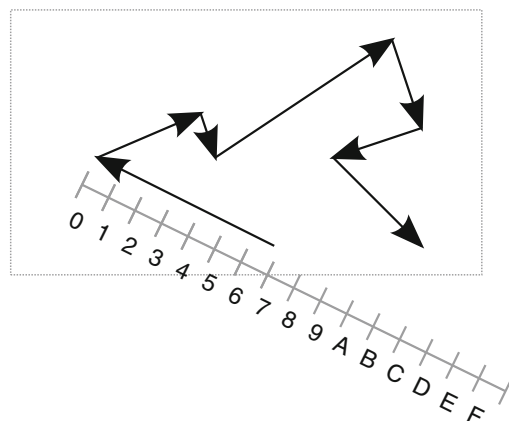
### What's missing in scanpath comparison?

It is evident that more advanced, or rather, more finely tuned, scanpath comparison measures are needed. One feature of scanpaths that this article has not yet covered is the duration of fixations that comprise a scanpath. We know from many empirical examples that the amount of time spent foveating a particular point greatly influences (and is influenced by) ongoing cognitive processing (cf. Henderson & Pierce, 2008). It is strange, then, that this avenue is

remarkably unexplored in the context of scanpath analysis. ScanMatch (Cristino et al., 2010) does attempt to address this, approximating fixation duration by representing AOIs that were inspected for longer with more letters (e.g., each occurrence of a letter could denote, say, 50 ms, so that if a region labeled A is dwelt upon for 150 ms it would be represented by AAA in the string). Note, however, that dwell times (usually defined as the sum of fixation durations from entering an AOI until leaving it) will rarely fit into the boundaries imposed by binning in this way. If the bins are 50 ms and an AOI is inspected for a total of 130 ms in one dwell, should this be rounded down to XX or rounded up to XXX for the particular AOI in question? Again, a kind of *quantization of time* applies here, in the same way that quantization of space applies to the use of AOIs. However, one should point out that there is no actual physical constraint on temporal bin size in ScanMatch, and in principle this could be set to the sampling frequency of the eyetracker.

To summarize what is missing, ideally, a measure that has the *minimum requirements* of incorporating the following dimensions into scanpath comparison is desirable:

1. *Order*: To be a “path,” the ordinal sequence of fixations and saccades must be reflected in the comparison, otherwise two scanpaths traversing exactly the same spatial positions but in the reverse order would be judged as identical.
2. *Position*: The scanpath representations being compared must reflect, as accurately as possible, the locations of fixations and saccades in  $x, y$  space. This avoids the quantization noise associated with the use of AOIs.



**Fig. 6** Representing a scanpath as a hexadecimal sequence. The amplitudes of saccades are measured on a 16-unit hexadecimal ruler (left); a 16-region segmentation of circular space is used to represent saccadic direction (right). The first saccade of the example sequence is measured with each representation type, giving D7. The whole scanpath is represented by the string D7 23 71 28 73 B3 54. From here, it is possible to perform editing operations between two strings, adopting the same principles as in the Levenshtein method. As the hexadecimal

3. *Shape*: An intuitive property if asked to subjectively judge the similarity between two scanpath visualizations would be shape. The proportional relationships between the geometry of two scanpaths can have important implications in vision research.
4. *Fixation duration*: When the point of highest visual acuity—the fovea—pauses en route, this is arguably the most important aspect of a scanpath from a behavioral and perceptual perspective, because this is when visual information is extracted. Fixation duration should therefore be a prerequisite in scanpath comparison.

These criteria for scanpath comparison go some way to meeting the definition of a scanpath set out in Holmqvist et al. (2011): “the route of oculomotor events through space within a certain timespan” (p. 254). It should be noted, however, that these criteria might be contradictory in some circumstances. For example, in Fig. 3, if we are interested in shape similarity, then by definition we are not interested in position similarity. This is not to say that the dimensions of our approach—outlined below—are mutually exclusive, however; it is quite possible for two scanpaths to be similar in shape *and* position.

### Scanpath shape and MultiMatch

We have not yet touched on methods that take shape into account. There are ways in which this can be done, and one of the lesser-known methods for scanpath comparison implemented by Gbadamosi and Zangemeister (2001; Gbadamosi, 2000) treats scanpaths as a sequence of vectors,

sequence is independent of the stimulus (AOIs are not used), the costs associated with insertions, deletions, and substitutions can be drawn directly from the vector differences (i.e., the Euclidean distance between the endpoints of vector pairs). From *Eye Tracking: A Comprehensive Guide to Methods and Measures*, by K. Holmqvist et al., 2011, p. 271; Oxford, U.K.: Oxford University Press. Copyright 2011 Oxford University Press. Reprinted with permission

retaining shape information (Fig. 6). We have already discussed that shape can be important in mental imagery research—for example, when one scanpath is spatially scaled down relative to the other, or shifted relative to a new locus (Johansson et al., 2006). Indeed, Gbadamosi and Zangemeister used vector string-editing to study visual imagery in homonymous hemianopia, revealing that, despite their perceptual deficit, patients with hemianopia present consistency in their scanpaths during visual imagery of a picture.

It is important to point out, however, that while shape *might* be important in isolation for other reasons (e.g., comparing between experiments with different  $x$ ,  $y$  coordinates, comparing rotated stimuli, or validating phenomena such as the well-known F-shape scanning pattern on websites; Nielsen, 2006), usually the position of fixations and their order take precedence outside of the field of mental imagery research.

In neural terms, it makes sense to consider shape, though: It has been proposed that “buildup” cells in the superior colliculus exclusively code the landing position of an upcoming saccade, whilst “burst cells” independently execute the eye movement with the required metrics to reach this location (Findlay & Walker, 1999). This is analogous to treating saccades as vectors, since the superior colliculus carries separate neural representations of saccadic direction (in buildup cells) and amplitude (in burst cells; Munoz & Wurtz, 1995a, 1995b). In this view, scanpath comparison should always consider the shape that the scanpath forms, since this is how the saccade programming system handles the complex job of directing the fovea.

However, as it stands, vector-based string-editing regionalizes angles between vectors into discrete bins, such that 360 deg can be split into 16 equal segments of 22.5 deg. Representation of vector length is done similarly, with lengths 0– $n$  being divided on a ruler that measures the vectors, in some fairly liberal unit like centimeters.

Despite its appeal, therefore, vector-based string-editing gives a similarity value that is entirely blind to the veridical positions of fixations in the recorded data. Moreover, quantization noise (not of stimulus space itself, but of the scanpath representation) remains a problem, because saccades are segregated into discrete bins. How can we incorporate shape satisfactorily, as well as the other desirable criteria for scanpath comparison outlined in the “What’s Missing in Scanpath Comparison?” section?

Our principle for scanpath comparison (described in Jarodzka, Holmqvist, & Nyström, 2010) considers all the prerequisites outlined, shape included. It uses vector mathematics to represent scanpaths as ideal saccades connecting fixation points via the shortest route between them. Each saccade vector  $[u = (x, y)]$  has direction and length, and because fixations are joined, both the position and duration of fixations remains unchanged. This technique makes several dimensions available to the user in a way that addresses the vast majority

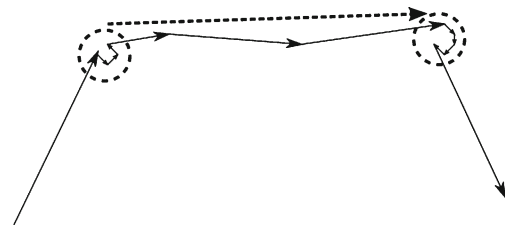
of the concerns raised above. It uses a representation of eye movements closer to the recorded oculomotor data, whilst retaining the temporal sequence and spatial structure inherent to scanpaths. In the previous article, we described this technique and highlighted its strengths with some basic artificial data. Here, we put it to the test with simulations and real empirical data, to evaluate its utility for scanpath research into visual and cognitive processing. As our approach has the flexibility to compare multiple scanpath dimensions, we here name it “MultiMatch.”

### MultiMatch: Implementation details

As our approach has been fully outlined in a previous article (Jarodzka, Holmqvist, & Nyström, 2010), we limit the coverage here to information that will assist the reader in the context of the present experiments. Several consecutive steps must be followed in MultiMatch to meet the conditions for scanpath comparison set out above.

#### Simplification

Scanpaths form rich data that can be too complex to compare without a prior stage of simplification: Small, irrelevant differences can (invalidly) lower, or sometimes inflate, the similarity result. The goal is to condense the scanpath enough so that small saccades and fixations in local areas do not overly bias the final similarity result, but not so much that the properties of the original scanpath are lost in the simplification process. As we have seen, this is at the heart of the challenge in scanpath comparison. MultiMatch achieves this via thresholding that groups small, locally contained saccades together. Thus, if successive saccade vectors ( $u_1, u_2, u_3, \dots, u_m$ ) have amplitudes smaller than the threshold set ( $T_{amp}$ ), they are grouped to form one new vector ( $u' = u_1 + u_2 + u_3 + \dots + u_m$ ), as illustrated in Fig. 7 (dashed circles). Likewise, if successive saccades follow the same general direction between two points in the stimulus, these vectors are merged, because they do not, in of



**Fig. 7** MultiMatch Step 1: Simplification—Amplitude-based (dashed circles) and direction-based (dashed arrow) clustering. From *Eye Tracking: A Comprehensive Guide to Methods and Measures*, by K. Holmqvist et al., 2011, p. 275; Oxford, U.K.: Oxford University Press. Copyright 2011 Oxford University Press. Reprinted with permission

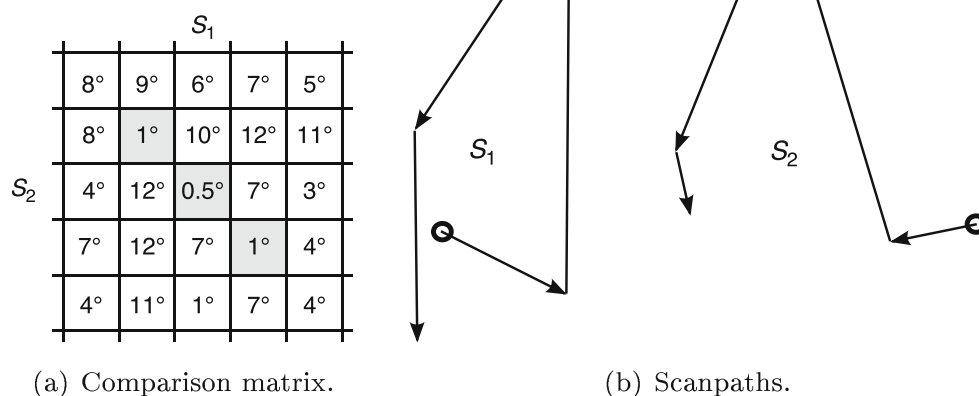
themselves, impart any extra meaning to the generic shape of the scanpath. Consequently, when the angular deviation between two saccades is smaller than the threshold level ( $T_{\Phi}$  radians), the vectors are combined into one larger saccade vector as in Fig. 7 (dashed arrow). The simplification process is iterative, with direction- and then amplitude-based clustering operating in a loop until no further simplification can be carried out. Throughout this article, the threshold level for amplitude-based simplification was set to 10 % of the screen diagonal, while the threshold for direction was 45 deg.

It is crucial to note here that direction-based clustering can be dangerous, and in many cases it would be necessary to set a condition for the removal of fixations. If the example in Fig. 7 were taken from scene viewing, we might not want to remove fixations along the path of the dashed arrow, since they could correspond to information acquisition from the scene. This is clearly important, and in such cases it would be advisable to set a minimum duration above which fixations would never be excluded (e.g., the average fixation duration within the trial).

### Alignment

Once we have two simplified scanpaths, the next step is to *temporally align* them, so that we know which vector

pairings to compare. This is done on the basis of scanpath shape, taking the new, simplified scanpaths ( $S_1 = u_1, u_2, u_3, \dots, u_m$  and  $S_2 = v_1, v_2, v_3, \dots, v_n$ ) and matching the sequence of vectors using a comparison matrix in which costs are drawn from vector differences between potential pairings. Note that throughout this article the similarity metric used for alignment is vector difference—that is, shape—but it is equally possible to align scanpaths on the basis of other dimensions—for instance, the positions of fixations or fixation duration. MultiMatch uses the Dijkstra (1959) algorithm to find the shortest path through the comparison matrix, from the top left corner to the bottom right. This algorithm works on a graph representation of the matrix, where matrix elements are called *nodes* and vector differences are the costs of links between nodes. The algorithm finds the shortest path from the first node to the last, taking all possible routes into account. This approach is analogous to way-finding, where nodes are cities and the costs are the driving distances between them; the goal is to find the shortest route. Now the vector sequences can be aligned according to this path. The process of temporal alignment is indicated in Fig. 8 and is explained in full in Jarodzka, Holmqvist, and Nyström (2010).



**Fig. 8** Pairwise vector comparison of all saccades in two vector representations ( $S_1 = u_1, u_2, u_3, u_4, u_5$  and  $S_2 = v_1, v_2, v_3, v_4, v_5$ ) of scanpaths. Circles denote the onsets of scanpaths. **a** The comparison matrix shows, for each pairwise vector comparison, the length of the differential vector ( $u_i - v_j$ ) in degrees of visual angle. If the two vectors are similar in amplitude and direction, this value is low. **b** Scanpaths

used in the comparison. Gray matrix cells in panel (a) indicate consecutive mappings that would produce a good alignment for a subset of the saccades. From *Eye Tracking: A Comprehensive Guide to Methods and Measures*, by K. Holmqvist et al., 2011, p. 277; Oxford, U.K.: Oxford University Press. Copyright 2011 Oxford University Press. Reprinted with permission



## Comparison

Finally, MultiMatch can compare the aligned scanpaths. This is a simple process of subtracting dimensions between the vectors (see Table 1) and taking an average. Each pair of simplified saccade vectors, and their accompanying fixations, can be compared on the basis of five dimensions, yielding similarity values between 0 and 1 (inverted so that 1 equates to *identical*).

**Table 1** Dimension subtraction

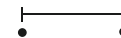
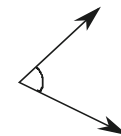
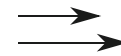
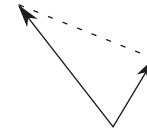
**Shape:** This is the vector difference between aligned saccade pairs,  $(u_i - v_j)$ —a measure of similarity in scanpath shape. The resulting similarity value is normalised by  $2 \times$  the screen diagonal (the maximum theoretical value).

**Length:** This is the difference in length between the endpoints of saccade vectors—a measure of similarity in saccadic amplitude. The resulting similarity value is normalized by the screen diagonal.

**Direction:** This is the angular distance between saccade vectors—a measure of similarity in shape when saccadic amplitudes are different. The resulting similarity value is normalized by  $\pi$ .

**Position:** This is the difference in position between aligned fixations—a common measure of scanpath similarity in terms of Euclidean distance. The resulting similarity value is normalised by the screen diagonal.

**Duration:** This is the difference in fixation durations between aligned fixations—a measure of similarity in processing time. The resulting similarity value is normalized against the maximum duration of the two being compared.



For each of the five dimensions—shape (vector difference), length, direction, position, and fixation duration—a difference can be calculated. Arrows denote vector–saccade pairs. The sizes of the circles in the duration case indicate one fixation that is longer than the other.

MultiMatch thus gives a finer level of detail in scanpath comparison. AOIs are not used, so quantization noise is much reduced, and this is true of fixation durations also, where the actual durations are used. Moreover, scanpaths of different lengths can be compared, which has been a distinct problem in scanpath comparison to date. Often this problem is solved by cutting off the scanpaths after 7–10 fixations, which is obviously not optimal because—for difficult tasks especially—one participant could inspect a region much later than another. MultiMatch takes the order (sequence) *and* position into account, and also preserves scanpath shape, which is a fundamental characteristic largely ignored in the literature.

The overarching goal of this article is to test the limits of MultiMatch with the kinds of data acquired in the attention and perception experiments employed in eye movement research. To this end, two experiments are

presented, in which MultiMatch is assessed against the most popular and accessible new scanpath comparison algorithm, ScanMatch (Cristino et al., 2010; note that we used the default settings with ScanMatch throughout). The first experiment evaluated how sensitive each algorithm is to spatial noise using simulated data. The second was designed to provide data with known issues in scanpath comparison—for example, the problems with AOIs or scaling described above—and real eye movement data

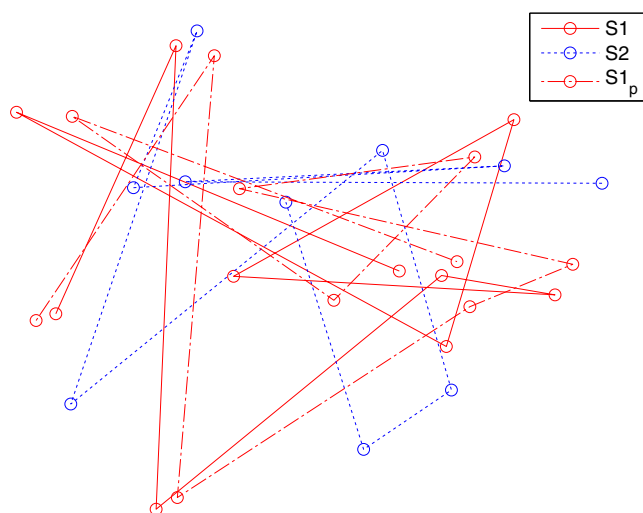
were collected to test how well MultiMatch and ScanMatch capture scanpath similarity in these dimensions.

### Experiment 1: Robustness to spatial noise

To estimate the sensitivity of MultiMatch, a procedure similar to that followed in Cristino et al. (2010, Exp. 1) was implemented. Whereas Cristino et al. pitted ScanMatch against the Levenshtein method, here we assessed MultiMatch versus ScanMatch.

#### Method and results

Two scanpaths ( $S_1$  and  $S_2$ ) with random positions were generated, each containing ten synthetic fixations. The fixation *positions* in scanpath  $S_1$  were perturbed with noise

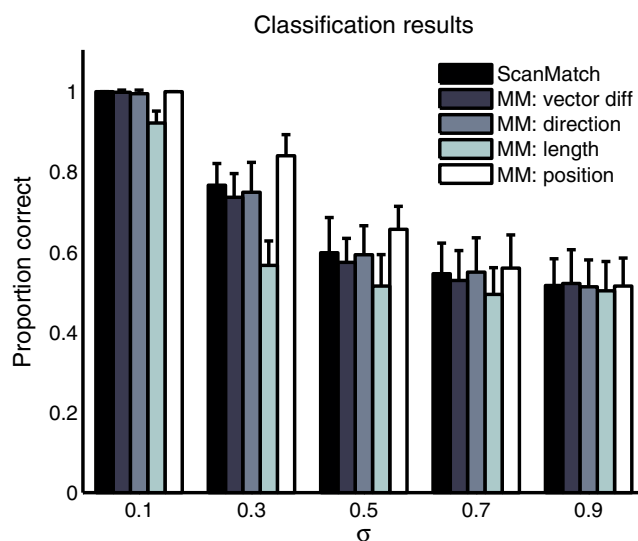


**Fig. 9** Example of two random scanpaths,  $S_1$  and  $S_2$ , and the variation  $S_{1p}$  that is created by adding spatial noise to  $S_1$  ( $\sigma = 0.1W$ ). A scanpath similarity algorithm should classify  $S_1$  and  $S_{1p}$  as more similar than  $S_1$  and  $S_2$

drawn from a Gaussian distribution, with a standard deviation ( $\sigma$ ) of 10 %–90 % of the screen width ( $W$ ). This created a new scanpath  $S_{1p}$ . Figure 9 shows examples of three such scanpaths.

Twenty-four sets of perturbed scanpaths were generated for each level of  $\sigma$ , which varied in five steps of 0.2 intervals; each set contained 50 new scanpaths. As  $\sigma$  increases, each new artificial fixation position in  $S_{1p}$  differs more and more from its origin in  $S_1$ , rendering the two scanpaths less similar, until their commonality is no longer distinguishable as compared to  $S_2$ . Thus, the similarity  $S$  between  $S(S_1, S_2)$  and  $S(S_1, S_{1p})$  was computed for each case using four of the dimensions in MultiMatch (duration was omitted because in this experiment, we were only concerned with the spatial characteristics of similarity classification). When  $S(S_1, S_2) < S(S_1, S_{1p})$ —that is, when the similarity between  $S_1$  and its noise-perturbed variation is larger than the similarity between the two random scanpaths—MultiMatch was considered to have correctly classified the perturbed scanpath. As  $S_{1p}$  varied in terms of spatial position, it was expected that the position dimension of MultiMatch would provide the best classification. Scanpaths were also classified according to the same procedure with ScanMatch using the default settings and a  $12 \times 8$  AOI grid. This gives a total of 30,000 classifications: 24,000 for the four dimensions of MultiMatch, and 6,000 for ScanMatch.

Figure 10 shows the results of Experiment 1. Adding noise with an increasing standard deviation to the fixation positions, the proportions of correctly classified trials decreased, until the algorithms could no longer detect that  $S_1$  was the base of  $S_{1p}$ , at around  $\sigma = 0.7$ . The decrease was somewhat steep for both MultiMatch and ScanMatch already at small noise levels (a drop in



**Fig. 10** Influence of adding Gaussian noise with standard deviation  $\sigma$  on the abilities of ScanMatch and MultiMatch to correctly classify a perturbed scanpath, such as  $S_{1p}$  in Fig. 9, as more similar to  $S_1$  than  $S_1$  is to  $S_2$ . The levels of  $\sigma$  reflect percentages of screen width, which in this case was 1,280 pixels; thus, perturbations of 0.1 translate to 128 pixels, up to 1,152 pixels for perturbations of 0.9. Error bars represent standard deviations of the means

classification accuracy of roughly 20 % on average between perturbation levels of 0.1 and 0.3). It is reassuring, however, that both the position dimension of MultiMatch and ScanMatch retained a high proportion of correctly classified perturbed scanpaths at higher noise levels, each still performing approximately 10 % above chance at noise level 0.5. The position dimension of MultiMatch did a little better than ScanMatch at a noise level of 0.3, but this was not a huge difference; otherwise, MultiMatch and ScanMatch performed comparably.

Because the scanpaths were perturbed spatially (i.e., in terms of position), the results are in line with what we hypothesized. But it is also important to note that the shape and direction dimensions performed comparably to ScanMatch throughout. The worst classification was on the length dimension, most likely because vector pairings are likely to differ least in length between  $S_{1p}$  and a randomly drawn  $S_2$ .

## Discussion

Taken together, these results demonstrate that it is possible to compare scanpaths on multiple dimensions in a way that makes sense: Position was varied, and the position dimension fared best in simulated classifications. Both MultiMatch and ScanMatch performed well here, with some noteworthy trends in the data. When subtracting between the positions of synthetic fixations in  $S_1$ ,  $S_{1p}$ , and  $S_2$ ,

MultiMatch takes the veridical difference in  $x$ ,  $y$  coordinates, whereas ScanMatch uses generalized AOIs. This may explain the approximately linear reduction in classification accuracy for the position dimension of MultiMatch, while the remaining dimensions, and ScanMatch, demonstrated a more exponential fall-off. It should also be noted that each dimension of MultiMatch (length excepted), and ScanMatch as well, still performed above chance up to high levels of noise ( $\sigma = 0.7$ ). This illustrates that both methods are robust, and the advantage of the distance criterion in the substitution matrix of ScanMatch over the basic Levenshtein distance.

## Experiment 2: Assessing performance for sequences with known issues in scanpath comparison

The first experiment demonstrated the favorable performance of MultiMatch to variations in spatial noise. How well does it cope with some of the other known issues in scanpath comparison outlined in the introduction? For example when scanpaths are spatially scaled down, or shifted in locus, as in mental imagery research, or when both position and order is integral, as when one scanpath is an exact reverse copy of the other. In the second experiment, we addressed the generality of MultiMatch to detect different kinds of scanpath similarity, and again contrasted its performance against that of ScanMatch. Real eye movement data were collected from human observers while they viewed sequences of dots, presented one at a time. To the participants, the dots appeared much as when the eyetracker was calibrated, but in reality each dot sequence was paired with another, randomly interleaved, that retained a degree of intuitive similarity difficult to capture and quantify by scanpath comparison algorithms at present (e.g., similarity

in scaling, spatial offset, or fixation duration; see Fig. 11). The logic is that by using multiple dimensions, MultiMatch should be able to identify similarity (or the lack of it) between the matched scanpaths that these dot pairings would produce, whereas ScanMatch should have difficulty.

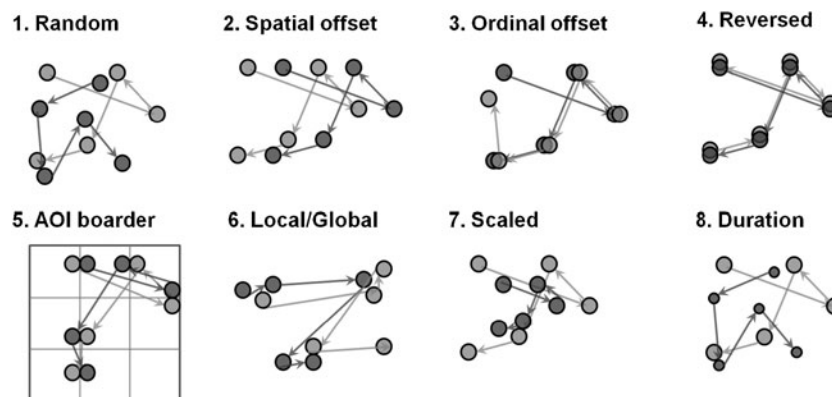
## Method

**Participants and apparatus** A group of 20 participants (9 females, 11 males;  $26.9 \pm 5.3$  years old) generated scanpaths by looking at stimuli in an eyetracker. All participant had normal or corrected-to-normal vision. In addition, ideal scanpaths were synthesized by assuming that a single participant perfectly performed the task of looking at points in the given order.

The stimuli were displayed on a Samsung Syncmaster 931c TFT LCD 19-in. ( $380 \times 380$  mm) screen running at 60 Hz, with a resolution of  $1,280 \times 1,024$  pixels. The stimuli were presented using MATLAB R2009b and the Psychophysics Toolbox (Brainard, 1997).

Binocular eye movements were recorded at 500 Hz with the SMI HiSpeed system and iView X 2.5, using default settings.

**Stimuli and design** Dot sequence stimuli were produced in order to elicit scanpaths with varying degrees of similarity between them. Random pairs were constructed to form a baseline for scanpath similarity against which scanpath similarity values for other pairs could be compared. The other seven types of scanpath pairs each represented a particular aspect of similarity (Fig. 11). A good scanpath similarity algorithm should be able to recognize each of these types of similarity by reporting a high similarity score for that particular aspect.



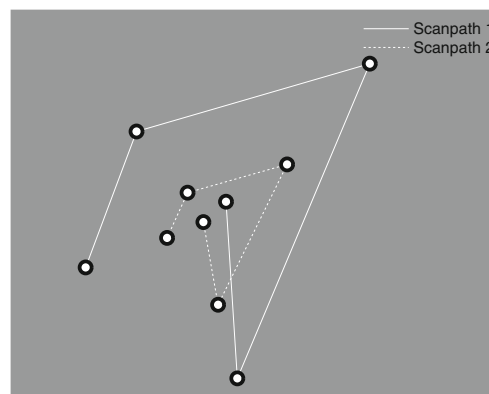
**Fig. 11** Eight different examples of known issues in scanpath comparison. The dark and light dot sequences share types of similarity that are often missed in scanpath comparison research. Each sequence of dots that participants viewed was matched with another, according to

one of the eight examples shown. The dot size represents presentation time. (This figure is presented for clarification only, and these exact dot sequence pairings were not necessarily used.)

Eight types of sequence pairs were constructed as follows. Each type contained ten pairs.

1. *Random* sequences contained dots with  $(x, y)$  coordinates drawn from a uniform distribution  $U(0.05, 0.95)$ . Random scanpaths served to produce baseline scanpath similarity values, the similarity expected by chance.
2. *Spatial offset* means that one dot sequence was translated, with an arbitrarily large spatial offset, in relation to the other dot sequence in the same pair.
3. *Ordinal offset* means that one sequence was shifted (with an ordinal offset) in relation to the other sequence in the same pair; that is, the dot at position  $i$  in one sequence corresponded to the dot at position  $i + 1$  in the other sequence, and so forth.
4. *Reversed* pairs contained sequences with the same positions, but in reversed order; that is, a dot at ordinal position  $i$  in one sequence would be at position  $l - i + 1$  in the other sequence, where  $l$  was the length (number of dots) of the sequence.
5. *AOI border* refers to the case in which two sequences differed only by the fact that the dots were located on each side of an AOI border, after stimulus space had been divided into a  $5 \times 5$  grid with equally sized sectors. Each AOI spanned a grid sector.
6. *Local/global* sequences were constructed such that two or more dots in a sequence formed local clusters, whereas the other dots were farther apart. The global shapes of two paired sequences were similar, but the local clusters could differ.
7. *Scaled* pairs of sequences differed in the degrees to which they covered the stimulus space. The sequence with large coverage was an upscaled, but otherwise identical, version of the one with small coverage.
8. *Duration* sequence pairs again had random positions, but the durations of each dot and its pair were unmatched, being drawn at random.

Each sequence contained five dots ( $l = 5$ ). The dots were black (intensity 0) with white centers (255), and presented on a midgray background (128). The diameter of a dot spanned half a degree of visual angle. Ten versions of each sequence pair were generated, giving in total 160 sequences (80 pairs). Figure 12 shows approximately what a sequence of dots and its pair looked like to the participants. All sequences except the random and duration ones were created manually. The durations for which dots remained visible were randomly selected from the interval 800–1,500 ms. With the exception of duration, each dot in one sequence was matched with the corresponding dot in its pair in terms of presentation time. In the case of duration, however, the presentation times for dots were not matched between sequences, but remained random



**Fig. 12** Example pair of dot sequences in which Scanpath 1 is an enlarged version of Scanpath 2 (equivalent to “Scaling”). The legend and the lines connecting dots were not used in the actual experiment, where each sequence was shown one dot at a time

(cf. the sizes of the circles for the random and duration patterns in Fig. 11).

These spatial and temporal restrictions allowed different properties of the scanpaths to be assessed. For example, fixation durations should be *dissimilar* for the duration sequence pairs, but the duration dimension should return higher similarity for the other seven types (where, if Dot 3 in  $s_1$  was presented for 900 ms, Dot 3 in  $s_2$  would also be presented for 900 ms).

*Procedure* On their arrival, informed consent was obtained from the participants, in accordance with the ethics policy of the Humanities Lab at Lund University. Instructions were available on the computer screen explaining the task, and the experimenter was present throughout to clarify any uncertainties that the participants might have. After the participants had agreed to take part and confirmed that they understood the task, the experimenter used a 13-point calibration procedure to calibrate the eyetracker, followed immediately by validation with the four points oblique to the center. The validation accuracy across all participants was  $0.91 \pm 0.60$  deg ( $x$  offset) and  $0.60 \pm 1.02$  deg ( $y$  offset) ( $M \pm SD$ ).

Each trial (i.e., one complete sequence of five dots) began with a central fixation cross presented for 2,000 ms. Dots sequences were presented one at a time, with the participants having been expressly instructed to “look at each dot in turn as it appears, and remain looking at it while it is still on the screen, without moving the eyes in anticipation of the next dot before it has appeared.” The dot sequences were presented to participants randomly interleaved, so participants were unaware of the sequence pairings. Accuracy was stressed to the participants—to look at each dot as directly as possible—but they were also asked to perform the task at a natural pace. After half of the sequences had been shown (80 trials), the participants were given a break, during which

recalibration was carried out if required. A chinrest was used to maintain a stable viewing position of 670 mm throughout the experiment. Depending on the participants' speed at the task and on the time taken to calibrate and recalibrate the eyetracker, the experiment took approximately 30–40 min to complete for one participant.

All participants viewed every sequence. This allowed us to evaluate scanpath similarity within subjects—that is, the similarity of the two scanpaths produced when the same participant viewed a dot sequence and its matched pair.

**Data preprocessing** Fixations were detected with the adaptive velocity-based algorithm created by Nyström and Holmqvist (2010) using the default settings stated in their report. The first fixation in each trial was excluded, since it derived from the initial fixation cross.

As an initial check that participants had followed the instructions and that the recorded eyetracking data had sufficient accuracy (acceptable offset), the proportions of raw data samples located inside AOIs were calculated. The AOIs comprised squares with 4-deg-long sides, centered on each stimulus dot. If the average proportion over all trials was below 50 %, then that participant was excluded. This was the case for one participant only (38 %). In the average trial for the remaining participants, 75 % of the raw data samples were located inside an AOI. Furthermore, participants with a longest common sub-sequence smaller than three (i.e., poor behavioral performance in following the dots; see below for details) were also omitted from further analysis. This criterion excluded another two participants.

## Results

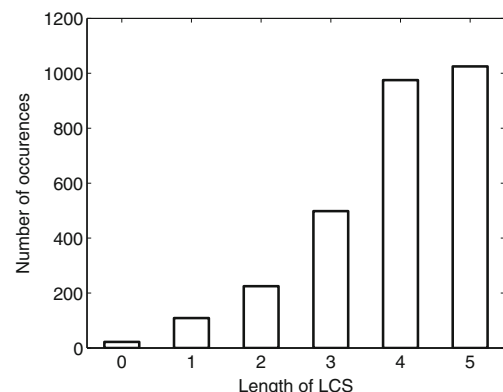
The findings of **Experiment 2** are organized as follows. Results that address participants' performance on the task are presented first. Then, the main body of the results is split into two main sections. The first section assesses how well MultiMatch and ScanMatch would handle the data if they were collected from ideal viewers—that is, viewers who would look exactly at the center of each dot every time, without making any mistakes. The second section carries out the same comparisons with the actual recorded eye movement data. Contrasting these two analyses is important, because it allowed us to gauge the objective similarity between the sequences with each dimension of MultiMatch and ScanMatch, as well as the similarity obtained when a degree of noise was introduced with real data, as would be the case in any eye movement experiment.

**Task performance** To determine how well participants performed the task, each recorded scanpath was first evaluated against the scanpath from an “ideal viewer”—that is, one

that would immediately take the correct path without making any mistakes. The commonality between ideal and observed scanpaths was assessed by computing their *longest common sub-sequence* (LCS; Hirschberg, 1977). This method compares two sequences—in this case, the numbers 1–5 in order, indicating the correct (ideal) sequencing of dots, versus whatever numbers were generated from the participants' actual data (observed). The LCS measure returns the maximum value if the same order is followed anywhere in the sequences—that is, irrespective of global differences and repetitions. In our case, an LCS of 5 would indicate that the participant successfully looked at the dots in the right order, regardless of deviations from the sequence along the way. Since the dots were shown one at the time on the computer screen and the task was very simple, an average LCS close to 5 could be expected. Figure 13 shows a histogram of the LCSs for all participants and trial types, and verifies high performance, giving an average LCS of  $3.88 \pm 1.14$  symbols.

However, it is possible to have an LCS that equals 5 but at the same time to look around a lot in the stimulus space. This would give an observed scanpath that was very different from the ideal one, even though all of the dots have been visited in the right order. To capture this aspect of task performance, we also report the numbers and durations of fixations. The average number of fixations per trial needed to complete the task was  $8.65 \pm 3.47$ , with an average fixation duration of  $422 \pm 158$  ms. Additional fixations beyond the ideal five are likely to reflect refixations and pauses with corrective saccades between dots that are far apart.

Given the data preprocessing steps taken and the behavioral performance measures, it is evident that the remaining data very well reflected participants who were completing the dots task properly, but that a certain amount of noise also remained. To quantify this source of natural variability, we first calculated the scanpath similarity for ideal viewers, and then contrasted this with the similarity results for the real participants' eye movement data.



**Fig. 13** Longest common sub-sequence histogram for **Experiment 2**, over all trial types and participants

**Table 2** Scanpath similarity for scanpaths from ideal observers

Type of Difference in Pair	Vector Difference	Direction	Length	Position	Duration	ScanMatch
Duration	.79	.73	.80	.63	.54	.40
Spatial offset	.87	.86	.81	.82	.75	.35
Ordinal offset	.78	.72	.79	.66	.55	.75
Reversed	.74	.64	.78	.66	.88	.47
AOI border	.89	.88	.87	.89	.73	.35
Local global	.78	.72	.76	.76	.51	.54
Scaling	.90	.88	.87	.84	.66	.34
Random	.69	.62	.70	.63	.58	.39

Each value should be compared to the baseline from the random scanpaths. For instance, a difference in spatial offset yields a higher (.87) than random (.69) similarity for the vector-difference dimension of MultiMatch.

*Ideal viewers* Table 2 presents the scanpath similarities between ideal pairs of scanpaths using the five dimensions of MultiMatch (cols. 1–5) and using ScanMatch (col. 6). These data are from hypothetical ideal viewers who would look exactly at the center of every dot in all sequences, without making any mistakes. What we then have are sets of matched scanpaths—perfectly positioned from a saccade-targeting point of view—according to the eight sequence types shown in Fig. 11. The input to MultiMatch consists of two fixation–saccade sequences generated from the ideal viewer, whereas the input to ScanMatch consists of two sequences of AOI labels created by dividing the stimulus space into a  $5 \times 5$  grid and labeling each “fixation” on the basis of the grid element (AOI) it landed on (Holmqvist et al., 2011, p. 193). ScanMatch was run using the default settings.

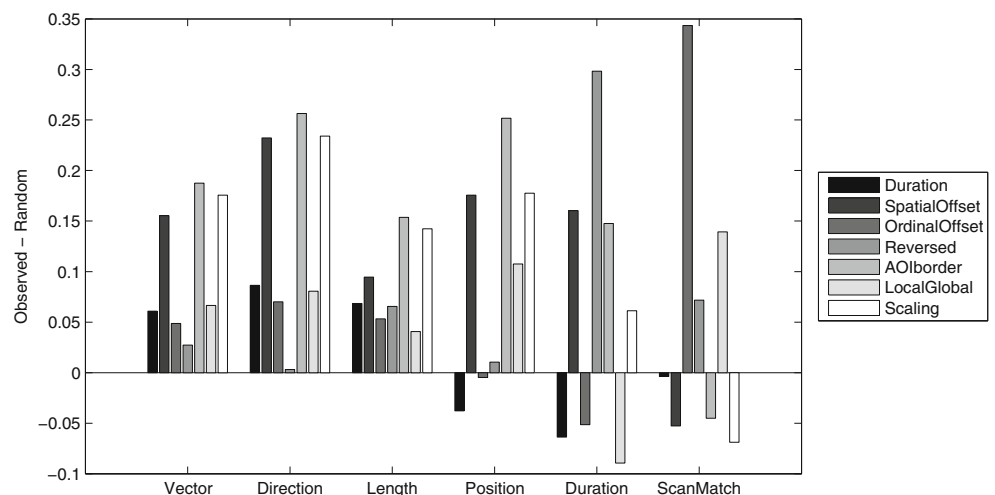
Values in the table indicate the average similarities for all of the sequence pairings of a particular type. The bottom row contains the similarities obtained when comparing two scanpaths with random positions. These values should be used as a baseline when interpreting the other similarity scores in the table. Thus, the numeric similarity scores are independent

between the dimensions of MultiMatch and ScanMatch. The data are then converted into difference scores in Fig. 14 so that it is easier to pick out which metrics are doing well, and when.

It can be seen from the table and figure that when the fixation positions in one scanpath are close to the fixation positions in the other scanpath (e.g., ordinal offset and local/global), ScanMatch does very well in capturing the similarity. Even two scanpaths with reversed order are judged as more similar than in the random case. However, when other aspects of similarity, such as spatial offset and scaling, are considered, the dependency on an AOI division of space in ScanMatch means that it does not recognize the inherent similarity in terms of shape between the scanpath pairs. This is better captured in several of MultiMatch’s five dimensions—vector difference and direction, in particular.

This exercise with simulated data captures the fundamental differences between MultiMatch and ScanMatch. It can be seen in Fig. 14 that MultiMatch never really treats ordinal offset and local/global as containing as much similarity as some of the other sequence types, whereas these are the two where ScanMatch finds the *most* in common. This is where

**Fig. 14** The same data (for ideal observers) shown as difference scores. This illustrates the relative differences on each dimension, given the different random baselines



**Table 3** Scanpath similarity for scanpaths within the same participant

Type of Difference in Pair	Vector Difference	Direction	Length	Position	Duration	ScanMatch
Duration	.79	.62	.84	<b>.70</b>	.41	.31
Spatial offset	<b>.89</b>	<b>.81</b>	<b>.90</b>	<b>.78</b>	.44	.28
Ordinal offset	<b>.85</b>	<b>.73</b>	<b>.88</b>	<b>.80</b>	.42	<b>.46</b>
Reversed	.78	.61	.81	<b>.72</b>	.41	<b>.36</b>
AOI border	<b>.85</b>	<b>.80</b>	<b>.85</b>	<b>.82</b>	.44	.32
Local/global	<b>.85</b>	<b>.74</b>	<b>.87</b>	<b>.81</b>	.45	<b>.42</b>
Scaling	<b>.84</b>	<b>.79</b>	.76	<b>.79</b>	.44	.27
Random	.75	.58	.81	.63	.43	.30

Each value should be compared to the baseline from the random scanpaths. For instance, a difference in spatial offset yields a higher (.89) than random (.75) similarity for the vector-difference dimension of MultiMatch. Numbers in bold font represent similarities that are significantly *larger*—that is, more similar—than the random baseline.

your choice of comparison approach depends critically on how subtle you expect differences between scanpath pairs to be. MultiMatch is very sensitive to geometric differences following sequence alignment (i.e., at the micro level between saccade vector pairs) but may underrepresent overall spatial similarities (as with ordinal offset and local/global). ScanMatch, on the other hand, is a spatially more coarse measure, but this may be a good thing if we want to identify general differences between scanpaths.

Finally, it is also important to highlight in these data that MultiMatch tells us *what* differs between the scanpaths being compared. For instance, the difference between two scanpaths, one of which is a reversed copy of the other, is signaled as a difference in direction and position, but similarity in terms of vectors, length, and duration.

*Within-subjects scanpath similarity* We now show the same analysis for the real eye movement data. It is reasonable to expect that scanpaths recorded from the same participant over the paired dot sequences will highly resemble the ideal-viewer data, just with added spatial error (cf. Exp. 1). The data were thus compared *within subjects* using MultiMatch and ScanMatch, as before. Differences in similarity between the scanpaths recorded from random dots and the scanpaths recorded from the other types of dot sequences were tested using a two-sample Kolmogorov–Smirnov test with  $\alpha = .01$ .

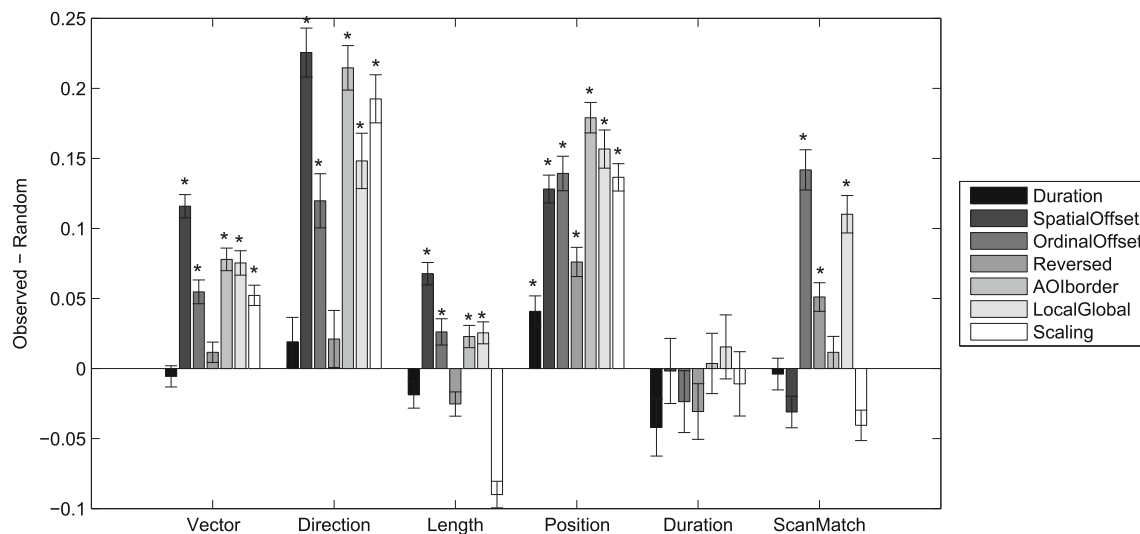
Table 3 and Fig. 15 show that every type of difference in the pairs of dot sequences is reflected in a significant difference in one or more of the dimensions of MultiMatch or ScanMatch. Again, ScanMatch is robust to sequence deviations when position similarity is high (e.g., the ordinal-offset and reversed sequences) and can cope with minor differences in shape (e.g., local/global). However, ScanMatch does not do so well when clear geometric similarities fall outside of AOI boundaries (e.g., spatial offset and scaling).

This is most clearly demonstrated in the AOI-border case, in which the only difference between the scanpaths is that fixations are directed to positions on either side of AOI boundaries, yet no similarity is detected by ScanMatch here, as compared to MultiMatch.

Attention should also be drawn to two particular cases of matched sequence pairs: spatial offset and scaling. Here, MultiMatch again achieved its objective to capture similarity in shape very well. Direction returns very large differences in similarity from the random baseline (.23) for the spatial-offset sequence types, whereas the difference from baseline in the position dimension is much smaller (.15). This is exactly what would be expected when the main difference between the two matched sequences is a shift in spatial position but an exact retention of scanpath shape. The ability of MultiMatch to capture shape similarity is also illustrated in the case of scaling, where all that has changed is the lengths of the saccades, and indeed, the length dimension no longer reaches statistical significance.

However, there are some downsides to the degree of dimensionality offered by MultiMatch in these analyses. First, the results can be difficult to interpret. One needs to keep in mind that the scanpaths are first temporally aligned before comparison (see the “Alignment” section above). Second, as the dimensions are independent,<sup>2</sup> it is problematic to compare between them. Third, with no real standard against which to gauge similarity, some results could be misleading. For example, MultiMatch still returns high (i.e., above chance) similarity on the position dimension between scaled scanpath pairs (Fig. 15). Perhaps it is more appropriate that ScanMatch finds even less position similarity in scaled scanpath pairs than between two randomly generated scanpaths (as indicated by the negative difference

<sup>2</sup> Note, however, that vector difference is essentially the combination of direction and length.



**Fig. 15** The same data (with real eye movements) shown as difference scores. This illustrates the relative difference on each dimension, given the different random baselines. The same results as in Table 3 are significant at the  $p < .01$  level with one-sample  $t$  tests; these are indicated with asterisks

score for the scaling sequences with ScanMatch). Again, this is a matter of measurement sensitivity and understanding the eye movement data that you collect.

Lastly, an interesting pattern emerges if we compare the results for the duration dimension between the ideal and real eye movement data. With ideal viewers (Table 2), we can see that duration similarity is high for all cases in which corresponding points in space are more likely to be aligned between the pairs (spatial offset, reversed, AOI border). As, apart from duration sequence types, dot presentation times were matched ordinally, it makes sense that when equivalent dots are compared, higher-duration similarities are returned.<sup>3</sup> However, this was not the case with the real eye movement data (Table 3). Here, the similarity in fixation durations never exceeded .45. This finding will be returned to in the Discussion section that follows.

## Discussion

Using real eye movement data, Experiment 2 revealed that MultiMatch is successful in detecting similarities in scanpath shape, as well as being sensitive to position, order, and fixation duration, the minimum requirements for scanpath comparisons set out in the “What’s Missing in Scanpath Comparison?” section. This goes some way to quantifying the intuitive sense of similarity that we have for certain scanpath visualizations.

While similarity was identified by MultiMatch in cases where ScanMatch fell short, it should be noted that the

default settings were used with ScanMatch. These could have been customized according to the task constraints (see “MultiMatch and ScanMatch” in the General Discussion below). Nevertheless, this introduces a certain subjectivity on the part of the experimenter in terms of where similarity is known or expected to occur, whereas MultiMatch returns a range of similarity metrics that are independent of stimulus features and spatial organization.

The fact that we found high duration similarities when equivalent points were more likely to be compared in the ideal, *but not* the real, eye movement data is noteworthy. It supports findings that fixation duration is idiosyncratic (Andrews & Coppola, 1999; Rayner, Li, Williams, Cave, & Well, 2007), but it is also likely to reflect the fact that real eye movement data contain instances of refixations on the target dots, as well as occurrences in which saccades are launched from farther away when the dots are not presented in close proximity to each other. This would create a variable time lag before the target fixation commences that is absent in the ideal data. From these findings, it seems that a push–pull relationship between maintaining fixation and initiating a saccade is more at play here than is the artificial task of looking at a dot for the whole length of time that it is shown (cf. Findlay & Walker, 1999). Thus, our data support previous findings that it is difficult to exert voluntary control over fixation durations (cf. Mosimann, Felblinger, Colloby, & Müri, 2004).

Experiment 2 was purposefully designed to demonstrate the strengths of MultiMatch, and in this regard it is not surprising that it outperformed ScanMatch. We can attribute the poorer performance of ScanMatch to its use of AOIs, not to the method itself, which, as we have seen in the introduction, provides very useful solutions to some of the main problems in scanpath research.

<sup>3</sup> Note that, although dot presentation times in the random case were matched between sequence pairs, because the positions were drawn at random, it is less likely that Dot 1 in  $s_1$  would be aligned to Dot 1 in  $s_2$  for comparison, and so on for Dots 2, 3, and so forth.



## General discussion

In this article, the application of a new method of scanpath comparison was assessed. Our method, MultiMatch, based on Jarodzka, Holmqvist, and Nyström (2010), performs comparably to ScanMatch (Cristino et al., 2010) with both simulated (Exp. 1) and real (Exp. 2) eye movement data. MultiMatch provides the flexibility to compare scanpaths in multiple dimensions, a feature that ScanMatch does not offer. The work presented here demonstrates the ability of MultiMatch to capture *how* two scanpaths differ, rather than simply returning a value that says that they are not the same but does not reveal why.

The converse of this is that MultiMatch is able to identify how scanpaths can be similar in one respect, but different in another. For example, when one scanpath is a scaled version of another, the difference is in length (cf. saccadic amplitude), but otherwise their spatial properties are very much alike (see Table 3).

### MultiMatch and ScanMatch

An obvious question that arises is why MultiMatch returns such high similarity values in relation to ScanMatch, and whether this is appropriate. The reason is that, because MultiMatch is AOI-independent, it is more sensitive to small commonalities between scanpaths. This is particularly true of the position dimension, since we saw in both Experiments 1 and 2 that MultiMatch is a better similarity detector on this dimension (see Fig. 10 and Table 3). However, greater sensitivity may not always be appropriate; it will increase the probability of finding similarity by chance (notice that the baseline similarity, even for the randomly generated scanpaths, is already high on the vector-difference and length dimensions, in particular). It is therefore always advisable to compare results to random baselines when using MultiMatch; otherwise, it would be theoretically possible to identify similarity between any two sequences.

Another point often returned to in this article is the use of AOIs. We have frequently argued that the absence of AOIs is a major advantage of MultiMatch over ScanMatch. However, presumably it is technically possible to remove the need for AOIs from the implementation of ScanMatch? This would necessitate the use of absolute Euclidean differences in position in the substitution matrix. It would be interesting to see whether this is a computationally pragmatic solution, and how the results would compare if it were implemented.

Again, related to the substitution matrix in ScanMatch, one may also challenge our results on the grounds that we did not weight the matrix beyond the default settings, and therefore it would be possible to push ScanMatch much further with our data. This is a valid criticism. However, it is not obvious what the substitution matrix should be

weighted in terms of. Clearly, our stimuli here contained no semantic content, being simple dot sequences, and therefore weighting the substitution matrix by semantic relatedness would not be appropriate. Likewise, there are no low-level visual properties that some stimuli could share in common (e.g., color or luminance). ScanMatch could take into account the spatial distances between dots to greater effect. In fact, Cristino et al. (2010, p. 694) did recommend setting the distance-based substitution matrix to two standard deviations of all saccadic amplitudes in the experiment at hand. We fully acknowledge that to do so could have a big impact on the results. However, apart from this distance-setting recommendation, we argue that when choosing weightings based on stimulus attributes, the researcher has to have some prior knowledge of the similarity result expected. In the context of a real experiment not set up to study scanpath algorithms per se, this could introduce a level of subjective bias on the part of the experimenter.

Of course, our stimuli in Experiment 2 were purposefully designed to demonstrate MultiMatch's strengths, and in this regard it was an unfair test. Even if we had used a distance setting of two standard deviations of saccade size in the substitution matrix of ScanMatch, it is unlikely that the results would be dramatically different, since we purposefully manipulated shape in many of our conditions, and where we did not, ScanMatch did as well as or better than MultiMatch. Future research should test the relative strengths of both algorithms with the kinds of experiments employed in popular research areas (e.g., scene perception, mental imagery, or embodiment of cognition). It is likely that ScanMatch (with its better ability to define and predict scanpath similarity a priori) will be more suited for some research questions, while MultiMatch (with its better ability to identify different kinds of similarity independently of the stimuli) will be more suited to others.

### MultiMatch settings

We have repeatedly argued against quantization of scanpath representations prior to a comparison; however, there is inherent quantization in the simplification steps of MultiMatch. Amplitude- and direction-based clustering (Fig. 7) reduces information from the original scanpath, effectively contradicting our own position maxim from the introduction. The goal, though, is that only redundant information is dismissed—more so than is the case with previous scanpath comparison techniques, especially those relying on AOIs. The remaining, simplified scanpath is a truer representation than either letter strings (weighted or not) or positions in  $x, y$  space alone, the two predominant principles of ScanMatch and Mannan linear distance, respectively. However, this is only the case if the thresholds are chosen appropriately, a matter that we now turn to.

Why is it necessary to quantize at all? Scanpaths are intrinsically complex, hence the difficulty in developing algorithms that suitably convey their properties. Some compromise in data reduction must be made, otherwise the comparison will not be computationally viable and will be near impossible to interpret. We believe that MultiMatch provides a good solution. Moreover, it should be noted that it is possible to change the threshold settings for simplification in MultiMatch.<sup>4</sup> This would depend on the nature of your research. For example, in reading research, it would be inappropriate to use a high direction-based simplification setting, because this could compress all saccades across the text into a straight line (one vector) apart from regressions and return sweeps. The influence of thresholds on scanpath similarity for different tasks is a worthwhile avenue for further research with MultiMatch. In the present article, the threshold settings were effectively arbitrary (10 % of screen diagonal for amplitude, 45 deg for direction, as stated above in the “Simplification” section). It would be very useful to quantify nonarbitrary thresholds, perhaps on the basis of the distinction between ambient and focal scanning (cf. Unema, Pannasch, Joos, & Velichkovsky, 2005). But, as it stands, there is no such distinction; in fact, it is questionable whether distinct categories of saccades are associated with global and local subsaccades at all. The fact that the researcher has the option to effectively “tune the level of quantization” and measure its effects should, in this context, then be seen as an advantage. However, this is not an excuse for choosing threshold values that are not theoretically motivated. As quantization noise is unavoidable, it should at least be based on sound assumptions about the oculomotor system. Exactly what these would be for the optimal threshold settings in MultiMatch remains to be seen.

There is also the choice of setting which dimension to align scanpaths on the basis of. We have used vector difference, because one of the aims of this article was to highlight the importance of scanpath shape. However, nothing prohibits the user from aligning on the basis of other dimensions. For example, it may be highly relevant to align on the basis of fixation duration. As we have seen with the matched dot pairs in [Experiment 2](#), temporal alignment based on shape does not necessarily align corresponding dot pairs for comparison when their presentation in space is random ([Table 2](#)). This returns low duration similarity scores even though, sequentially, each dot was presented for the same duration as its mate. Alignment on the basis of duration would allow similarity in fixation durations to be detected

independently of space—however, possibly at the expense of finding genuine similarity in the other (spatial) dimensions. Deciding which threshold to use for alignment is another setting choice for the researcher. It may be appropriate to align scanpaths according to the dimension in which you expect to find a difference, providing that this choice is theoretically motivated.

*Which dimension is most relevant?* There is a danger with MultiMatch that in the absence of a clear hypothesis, similarity could be found without knowing why. In fact, because the number of possible ways in which two scanpaths can be similar is greater, so is the chance of making a Type I error. Therefore, some guidance is needed as to which dimension to choose from the output. Perhaps the best way to do this is with some example scenarios of experiments:

1. *Expertise:* Say that we have a between-groups design in which expert and novice radiographers each examine the same chest X-rays for lung nodules (cf. Donovan, Manning, & Crawford, 2008). Which dimensions are important here? First and foremost, we want to know whether the places containing tumors on the images are fixated, so *position* becomes our primary concern. Are expert radiographers a more homogeneous group than novices in terms of the ordinal sequence of the positions that they fixate? Conversely, do novices consistently fixate the wrong positions? If some spatial positions in the stimulus are identical between groups, and fixation of these locations is a prerequisite for correct task completion, then the position dimension should be the primary concern. Note, however, that by *position* we mean the exact position in  $x, y$  space, not spatially distributed or semantic AOIs.
2. *Visual search strategies:* We know that there is a tight relationship between saccade amplitude and fixation time in visual search (e.g., Hooge & Erkelens, 1996). If the search task is difficult—due to crowding, for example—we can expect longer fixation durations and shorter saccade amplitudes (Vlaskamp & Hooge, 2006). If we make the task easier—by increasing the target-distractor discriminability, for instance—fixations become shorter, and saccades longer and more direct. This is further compounded by the fact that fixation durations are highly idiosyncratic (Andrews & Coppola, 1999; Rayner et al., 2007): It is likely that duration similarity will be higher in MultiMatch when comparing the scanpaths of the same person on different images than when comparing different people on the same image. This demonstrates the importance of knowing your stimuli, and the potential search strategies that your participants will use, before making predictions on the basis of the *duration* and *length* dimensions. Fixation duration

<sup>4</sup> Using thresholds per se means that we introduce some of the quantization noise that we are trying to avoid, and in terms of the direction threshold, this means that some fixations will be excluded (unless steps are taken, as suggested at the end of the “Simplification” section).

should vary more between participants, whereas saccade amplitude (length) may be more sensitive to stimuli, depending on visual clutter.

3. *Gaze biases*: There is growing concern with biases in where we look, whether this be systematic tendencies to saccade toward the center of a scene (Tatler, 2007) or horizon biases along the scene meridian (Foulsham & Kingstone, 2010; Foulsham, Kingstone, & Underwood, 2008). As well as the obvious case of being able to differentiate between two scanpaths containing the same positions but fixated in the opposite order (e.g., the reversed scanpaths in Fig. 11), the instances of gaze bias mentioned above are times when we should be developing our hypotheses in terms of the direction dimension when using MultiMatch. If a rose-plot direction histogram is a suitable format for your results, or you are interested in saccadic curvature, it is likely that the sequence of direction similarity produced by MultiMatch would be revealing, too.
4. *Spatial transformations*: In several instances, it would be helpful to quantify shape information more comprehensively outside of the mental imagery domain and make predictions on the basis of the vector-difference dimension. Suppose that we were investigating communication in face-to-face interaction with mobile eyetracking devices mounted on two participants conversing (e.g., Gullberg & Holmqvist, 1999). Here, what is “left” for participant A will be “right” for participant B, and vice versa. The vector-difference dimension comes into its own here, because shape remains intact, despite the spatial translation. In other instances, however, related to the particular paradigm that you are working with, you may wish to prioritize predictions about vector similarity. The double-step paradigm in oculomotor planning research, for instance (Schlag & Schlag-Rey, 2002), raises a number of testable hypotheses about the deviations in saccade direction and amplitude (i.e., vector) from the planned saccade trajectory.

The note of caution here should be that, just because MultiMatch returns similarity on a particular dimension, does not necessarily mean that this similarity is meaningful, unless you know why. Furthermore, it can also be very revealing if dimensions are *not* showing a difference: The lack of similarity on the direction and length dimensions, respectively, differentiates the reversed and scaled scanpaths in Fig. 15.

#### Implications and future research

*Pairwise versus groupwise similarity* Something that MultiMatch cannot provide is *groupwise* scanpath similarity, in the way that attention maps do. We have stated from the

outset that this is a pairwise comparison algorithm (like Levenshtein distance, ScanMatch, and Mannan linear distance) for taking two vector strings and identifying commonalities between them. Given many participants and only one task (e.g., free web search), it is not immediately obvious which scanpaths should be compared without some justification for matching two participants together. Thus, for larger-scale, but ultimately more coarse, analysis, attention map difference measures (e.g., Dempere-Marco et al., 2006; Ouerhani et al., 2004; Rajashekar et al., 2004) or transition matrix measures (e.g., Holmqvist, Holsanova, Barthelson, & Lundqvist, 2003; Hyönä, Lorch, & Rinck, 2003) may be more satisfactory. Nevertheless, pairwise approaches like MultiMatch could be very revealing for such research purposes if the task and design were well enough controlled. The strength of MultiMatch for future research lies in its ability to more rigorously test hypotheses relating to scanpath theory: for example, the similarity in viewing strategies between participants for similar tasks, or the similarity within participants when viewing the same image. To this end, it may be particularly powerful for encoding–recognition paradigms in scene perception research.

*Different tasks, different similarity* We touched on the issues surrounding scanpath similarity for different tasks and stimuli when offering advice above about which dimensions are most relevant. It would be worthwhile to corroborate these suggestions and a number of other avenues that remain unexplored at present. Perhaps position, duration, and direction would be useful for identifying scanpath similarity in tasks such as learning strategies in educational software (e.g., Karemaker, Pitchford, & O’Malley, 2010) or for distinguishing between experts’ and novices’ eye movement profiles (e.g., Jarodzka, Scheiter, Gerjets, & Van Gog, 2010), so as to establish better eye movement training techniques (e.g., Dewhurst & Crundall, 2008; Van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009). This would make sense, since acquiring visual skill involves learning both the importance of fixed and variable positions and the ability to transition between them, as well as how long we look where, which is often argued to be a good indicator of processing. To tackle which combination of dimensions apply in these scenarios, a factor analysis could be conducted to see how similarity clusters along MultiMatch’s dimensions. This could be applied in a number of different task settings between different fields of research.

*How important is the fixation detection algorithm?* Something that is often overlooked in eye movement research is the importance of the fixation detection algorithm. Here we

used a saccadic-velocity-based method (Nyström & Holmqvist, 2010) that, as well as detecting saccades and fixations, also detects glissades (small “wobbles” of the eye appended to around 50 % of saccades). The very fact that analysis of glissadic eye movement events does not come as standard in commonly used event detection algorithms (e.g., SR Research, 2007) demonstrates the sensitivity of established measures, such as fixation duration and saccadic amplitude, to the algorithm used to parse raw sample data from the eyetracker. To this extent, the event detection algorithm used prior to scanpath comparison is likely to affect the similarity results obtained. This could be particularly true for MultiMatch because, as we have seen, it can be sensitive to small differences between two saccade vector strings (depending on the threshold settings for simplification described earlier). It would be interesting to see whether two different fixation detection algorithms produce different scanpath similarity results for the same data set (either using MultiMatch or one of the other scanpath comparison principles described herein). If this is the case, then this should caution eye movement researchers as to the root cause of scanpath similarity—whether this be the system and tools used to find it (undermining conclusions about human vision) or the participants under study (supporting conclusions about the etiology of eye movement sequences).

### Summary and conclusions

In this article, we have tested a new method for scanpath comparison previously described in Jarodzka, Holmqvist, and Nyström (2010). We showed that our method, which we here name “MultiMatch,” can detect scanpath similarity across several dimensions: saccadic amplitude (length), direction, fixation position, fixation duration, and scanpath shape (vector difference). As scanpaths are inherently complex entities in eye movement research, we showed, both with artificial and real eye movement data, that this is a worthwhile step in being able to more comprehensively analyze scanpaths in a way that simultaneously captures both their spatial and temporal characteristics. Are two scanpaths similar? With the use of multiple dimensions, this article shows that it really does depend on how you look at it.

### Using MultiMatch

If you would like to use MultiMatch, please contact one of the authors, who will be happy to supply code (in MATLAB).

**Author note** This work was supported by a postdoctoral research grant from the Swedish Research Council, awarded to R.D. (Grant No. 435-

2010-849), and by the Linnaeus Center for Thinking in Time: Cognition, Communication and Learning (CCL) at Lund University, which is also funded by the Swedish Research Council (Grant No. 349-2007-8695). The authors thank Sebastiaan Mathôt and one anonymous reviewer for their helpful feedback on an earlier draft of the manuscript.

### References

- Andrews, T. J., & Coppola, D. M. (1999). Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments. *Vision Research*, *39*, 2947–2953. doi:10.1016/S0042-6989(99)00019-X
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. doi:10.1163/156856897X00357
- 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. doi:10.1162/jocn.1997.9.1.27
- Burmester, M., & Mast, M. (2010). Repeated web page visits and the scanpath theory: A recurrent pattern detection approach. *Journal of Eye Movement Research*, *3*, 1–20.
- Caldara, R., & Mielle, S. (2011). iMap: A novel method for statistical fixation mapping of eye movement data. *Behavior Research Methods*, *43*, 864–878. doi:10.3758/s13428-011-0092-x
- Coco, M. (2009). The statistical challenge of scan-path analysis. In *Proceedings of the 2nd Conference on Human-System Interactions* (pp. 369–372). Piscataway, NJ: IEEE Press.
- Cristino, F., Mathôt, S., Theeuwes, J., & Gilchrist, I. D. (2010). ScanMatch: A novel method for comparing fixation sequences. *Behavior Research Methods*, *42*, 692–700. doi:10.3758/BRM.42.3.692
- Day, R. (2010). Examining the validity of the Needleman–Wunsch algorithm in identifying decision strategy with eye-movement data. *Decision Support Systems*, *49*, 396–403.
- Dempere-Marco, L., Hu, X., Ellis, S., Hansell, D., & Yang, G. (2006). Analysis of visual search patterns with EMD metric in normalized anatomical space. *IEEE Transactions on Medical Imaging*, *25*, 1011–1021.
- Dewhurst, R., & Crundall, D. (2008). Training eye movements: Can training people where to look hinder the processing of fixated objects? *Perception*, *37*, 1729–1744.
- Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, *1*, 269–271.
- Donovan, T., Manning, D. J., & Crawford, T. (2008). Performance changes in lung nodule detection following perceptual feedback of eye movements. In B. Sahiner & D. J. Manning (Eds.), *Proceedings of Medical Imaging 2008: Image perception, observer performance, and technology assessment*. New York, NY: International Society for Optical Engineering. doi:10.1117/12.768503
- Findlay, J., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *The Behavioral and Brain Sciences*, *22*, 661–674.
- Foulsham, T., & Kingstone, A. (2010). Asymmetries in the direction of saccades during perception of scenes and fractals: Effects of image type and image features. *Vision Research*, *50*, 779–795. doi:10.1016/j.visres.2010.01.019
- Foulsham, T., Kingstone, A., & Underwood, G. (2008). Turning the world around: Patterns in saccade direction vary with picture orientation. *Vision Research*, *48*, 1777–1790. doi:10.1016/j.visres.2008.05.018
- Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, *8* (2), 6:1–17. doi:10.1167/8.2.6

- Gbadamosi, J. (2000). *Visual Imagery und Mikro-Sakkaden bei Hemianopsie-Patienten*. Fachbereich Medizin der Universität Hamburg.
- Gbadamosi, J., & Zangemeister, W. (2001). Visual imagery in hemianopic patients. *Journal of Cognitive Neuroscience*, *13*, 855–866.
- Grinding, T., Duchowski, A., & Sawyer, M. (2010). Group-wise similarity and classification of aggregate scanpaths. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (pp. 101–104). New York, NY: ACM.
- Gullberg, M., & Holmqvist, K. (1999). Keeping an eye on gestures: Visual perception of gestures in face-to-face communication. *Pragmatics and Cognition*, *7*, 35–63.
- Hacisalihzade, S., Stark, L., & Allen, J. (1992). Visual perception and sequences of eye movement fixations: A stochastic modeling approach. *IEEE Transactions on Systems, Man, and Cybernetics*, *22*, 3474–3481.
- Henderson, J. M., Brockmole, J. R., Castelano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R. P. G. van Gompel, M. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537–562). Amsterdam, The Netherlands: Elsevier.
- Henderson, J. M., & Pierce, G. L. (2008). Eye movements during scene viewing: Evidence for mixed control of fixation durations. *Psychonomic Bulletin & Review*, *15*, 566–573. doi:10.3758/PBR.15.3.566
- Hirschberg, D. (1977). Algorithms for the longest common subsequence problem. *Journal of the ACM (JACM)*, *24*, 664–675.
- Holmqvist, K., Holsanova, J., Barthelsson, M., & Lundqvist, D. (2003). Reading or scanning? A study of newspaper and net paper reading. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 657–670). Oxford: Elsevier Science.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R. C., Jarodzka, H., & van der Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford, U.K.: Oxford University Press.
- Hooge, I. T. C., & Erkelens, C. J. (1996). Control of fixation duration in a simple search task. *Perception & Psychophysics*, *58*, 969–976. doi:10.3758/BF03206825
- Hyönä, J., Lorch, R. F., Jr., & Rinck, M. (2003). Eye movement measures to study global text processing. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 313–334). Amsterdam, The Netherlands: North-Holland/Elsevier.
- Jarodzka, H., Holmqvist, K., & Nyström, M. (2010). A vector-based, multidimensional scanpath similarity measure. In *Proceedings of the 2010 Symposium on Eye-Tracking Research and Applications* (pp. 211–218). New York, NY: ACM.
- Jarodzka, H., Scheiter, K., Gerjets, P., & Van Gog, T. (2010b). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction*, *20*, 146–154.
- Johansson, R., Holsanova, J., Dewhurst, R., & Holmqvist, K. (2011). Eye movements during scene recollection have a functional role, but they are not reinstatements of those produced during encoding. *Journal of Experimental Psychology: Human Perception and Performance*.
- Johansson, R., Holsanova, J., & Holmqvist, K. (2006). Pictures and spoken descriptions elicit similar eye movements during mental imagery, both in light and in complete darkness. *Cognitive Science*, *30*, 1053–1079. doi:10.1207/s15516709cog0000\_86
- Johansson, R., Holsanova, J., & Holmqvist, K. (2011). The dispersion of eye movements during visual imagery is related to individual differences in spatial imagery ability. In L. Carlson, C. Hölscher, & T. F. Shipley (Eds.), *Expanding the space of cognitive science: Proceedings of the 33rd Annual Meeting of the Cognitive Science Society* (pp. 1200–1205). Austin, TX: Cognitive Science Society.
- Karemaker, A., Pitchford, N., & O'Malley, C. (2010). Does whole-word multimedia software support literacy acquisition? *Reading and Writing*, *23*, 31–51.
- Laeng, B., & Teodorescu, D.-S. (2002). Eye scanpaths during visual imagery reenact those of perception of the same visual scene. *Cognitive Science*, *26*, 207–231. doi:10.1207/s15516709cog2602\_3
- Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*, 1311–1328. doi:10.1068/p2935
- Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics – Doklady*, *10*, 707–710.
- Mannan, S. K., Kennard, C., & Husain, M. (2009). The role of visual saliency in directing eye movements in visual object agnosia. *Current Biology*, *19*, R247–R248. doi:10.1016/j.cub.2009.02.020
- Mannan, S., Ruddock, K. H., & Wooding, D. S. (1995). Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-D images. *Spatial Vision*, *9*, 363–386. doi:10.1163/156856895X00052
- Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision*, *10*, 165–188. doi:10.1163/156856896X00123
- Mathôt, S., Cristino, F., Gilchrist, I., & Theeuwes, J. (2012). A simple way to estimate similarity between eye movement sequences. *Journal of Eye Movement Research*, *51*, 1–15.
- Mosimann, U., Felblinger, J., Colloby, S., & Müri, R. (2004). Verbal instructions and top-down saccade control. *Experimental Brain Research*, *159*, 263–267.
- Munoz, D., & Wurtz, R. (1995a). Saccade-related activity in monkey superior colliculus: I. Characteristics of burst and buildup cells. *Journal of Neurophysiology*, *73*, 2313.
- Munoz, D., & Wurtz, R. (1995b). Saccade-related activity in monkey superior colliculus: II. Spread of activity during saccades. *Journal of Neurophysiology*, *73*, 2334.
- Nielsen, J. (2006). *F-shaped pattern for reading web content*. Text available at the author's website ([www.useit.com/alertbox/reading\\_pattern.html](http://www.useit.com/alertbox/reading_pattern.html)), accessed May 2011.
- Noton, D., & Stark, L. (1971). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research*, *11*, 929–942.
- Nyström, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior Research Methods*, *42*, 188–204. doi:10.3758/BRM.42.1.188
- Ouerhani, N., von Wartburg, R., Hügli, H., & Müri, R. (2004). Empirical validation of the saliency-based model of visual attention. *Electronic Letters on Computer Vision and Image Analysis*, *3*, 13–24.
- Rajashekar, U., Cormack, L. K., & Bovik, A. C. (2004). Point of gaze analysis reveals visual search strategies. In *Proceedings of Human Vision and Electronic Imaging IX* (pp. 296–306). New York, NY: International Society for Optical Engineering. doi:10.1117/12.537118
- Rayner, K., Li, X., Williams, C. C., Cave, K. R., & Well, A. D. (2007). Eye movements during information processing tasks: Individual differences and cultural effects. *Vision Research*, *47*, 2714–2726.
- Schlag, J., & Schlag-Rey, M. (2002). Through the eye, slowly: Delays and localization errors in the visual system. *Nature Reviews Neuroscience*, *3*, 191–191.
- SR Research, Inc. (2007). *EyeLink user manual* (Version 1.3.0). Mississauga, ON: Author.
- Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor

- biases and image feature distributions. *Journal of Vision*, 7(14), 4:1–17. doi:10.1167/7.14.4
- Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research*, 45, 643–659. doi:10.1016/j.visres.2004.09.017
- Underwood, G., Humphrey, K., & Foulsham, T. (2008). *Knowledge-based patterns of remembering: Eye movement scanpaths reflect domain experience. HCI and usability for education and work, LNCS 5298* (pp. 125–144). Berlin, Germany: Springer.
- Unema, P. J. A., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition*, 12, 473–494.
- Van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior*, 25, 785–791.
- Vlaskamp, B. N. S., & Hooge, I. T. C. (2006). Crowding degrades saccadic search performance. *Vision Research*, 46, 417–425. doi:10.1016/j.visres.2005.04.006
- Wooding, D. (2002). Fixation maps: Quantifying eye-movement traces. In *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications* (pp. 31–36). New York, NY: ACM.
- Zangemeister, W., & Liman, T. (2007). Foveal versus parafoveal scanpaths of visual imagery in virtual hemianopic subjects. *Computers in Biology and Medicine*, 37, 975–982.