frontiers

Vision Toolkit Part 4. Area of Interest and Associated Algorithms: A Review

Quentin Laborde[1,2,*], **Axel Roques**[1,3], **Allan Armougum**[2], **Nicolas Vayatis**[1], **Ioannis Bargiotas**[3], **Laurent Oudre**[1]

[1] *Université Paris Saclay, Université Paris Cité, ENS Paris Saclay, CNRS, SSA, INSERM, Centre Borelli, F-91190, Gif-sur-Yvette, France*
[2] *SNCF, Technologies Department, Innovation & Research, F-93210, Saint Denis, France*
[3] *Thales AVS France, Training & Simulation, F-95520, Osny, France*
[4] *Université Paris-Saclay, Inria, CIAMS, F-91190, Gif-sur-Yvette, France*

Correspondence*:
Corresponding Author
quentin.laborde@ens-paris-saclay.fr

ABSTRACT

Eye-tracking has become a key methodology for studying human attention, perception, and cognition. A central step in many gaze analysis pipelines consists in transforming raw eye-movement data into sequences of Areas of Interest (AoIs), enabling the use of symbolic and sequence-based models to characterize visual behavior. While this abstraction is essential for quantitative analysis, it also introduces methodological choices that critically shape how gaze dynamics are interpreted. This article reviews major approaches for analyzing AoI sequences, with an emphasis on how gaze dynamics can be modeled, summarized, and compared. We discuss probabilistic and information-theoretic frameworks for characterizing transition structure and uncertainty, as well as pattern-based and sequence-oriented methods designed to reveal recurrent or higher-order viewing strategies. Approaches for extracting representative gaze sequences and for visually exploring AoI dynamics are also examined, highlighting the complementary roles of quantitative modeling and qualitative interpretation. Throughout the review, we emphasize the impact of abstraction choices, particularly AoI definition, on downstream analyses, and discuss key challenges related to scalability, parameter sensitivity, reproducibility, and the analysis of complex or dynamic stimuli. By integrating visualization and modeling perspectives, this work provides a coherent framework for understanding AoI-based gaze analysis and outlines methodological directions for advancing the study of visual behavior.

Keywords: Eye-tracking, areas of interest, gaze dynamics, sequence analysis, visual attention

1 INTRODUCTION

The use of eye-tracking technology has increased dramatically over the past decades, driven by major advances in recording devices that are now more accessible, portable, and capable of delivering high-quality data. Modern eye trackers enable long-duration recordings and support increasingly complex experimental paradigms, both in controlled laboratory environments and in naturalistic settings. While these technological developments have substantially improved the acquisition of raw gaze data, a critical gap remains between the volume of data collected and our ability to interpret the visual strategies and cognitive processes that underlie observed eye movements.

Understanding how observers explore visual scenes and how gaze behavior reflects perception, attention, and cognition is of central importance across disciplines such as psychology, neuroscience, human–computer interaction, and ergonomics. However, the growing abundance and complexity of eye-tracking data make intuitive, low-level analysis increasingly impractical. This challenge has motivated the development of *high-level* representations that reduce data complexity while preserving meaningful structure, allowing researchers to efficiently analyze and compare visual strategies at both individual and group levels.

Within this context, *areas of interest* (AoIs) have emerged as a key abstraction. Rather than analyzing gaze as a sequence of precise fixation coordinates, AoI-based representations describe gaze behavior in terms of visits to semantically meaningful regions of the visual scene. This abstraction not only simplifies raw gaze data, but also embeds it with contextual information, enabling researchers to reason about *what* is being observed and *how* attention is distributed across relevant visual elements.

The use of AoIs is conceptually aligned with principles originating from Gestalt theory Wagemans et al. (2012), which emphasizes that visual perception is structured around coherent wholes rather than isolated sensory elements. From this perspective, individual fixations acquire meaning only when interpreted in relation to larger visual entities. Regardless of the theoretical scope attributed to Gestalt principles, the pursuit of semantic interpretation in eye-tracking data requires that fixations be contextualized within coherent visual regions. AoIs thus provide a natural framework for interpreting scanpaths in terms of the visual elements they traverse, rather than as mere geometric trajectories.

AoI sequences can be viewed as a specialized, symbolic representation derived from traditional scanpaths, as introduced in the *Scanpaths and Derived Representations* part of this review series (Laborde et al., 2024c). In this representation, a gaze trajectory is encoded as a sequence of symbols—typically letters or numbers—each corresponding to a specific area of interest, optionally enriched with temporal information such as fixation duration. While this abstraction is conceptually simple, it gives rise to a rich class of analytical methods that differ substantially from those traditionally applied to spatial scanpaths, and that form a distinct and specialized body of literature.

More specifically, AoI sequences are constructed by assigning each fixation to the semantic region it falls within, with each AoI associated with a unique symbol. The literature commonly distinguishes between *a priori* AoIs, defined in advance based on stimulus content, and *post hoc* AoIs, inferred directly from gaze data (Huang et al., 2024). In addition, it is often proposed to simplify AoI sequences by collapsing successive fixations within the same AoI, thereby emphasizing transition dynamics between regions rather than fine-grained temporal structure. An overview of this symbolization process is illustrated in Figure 1.

This article is the fourth contribution in an ongoing series of methodological reviews dedicated to the analysis of oculomotor signals and gaze trajectories. The first article (Laborde et al., 2024a) synthesized current knowledge on canonical eye movements, with particular emphasis on differences between controlled laboratory conditions and naturalistic viewing. The second article (**?**) reviewed segmentation algorithms and oculomotor features used to identify and characterize fixations, saccades, and smooth pursuits. The third article (Laborde et al., 2024c) focused on scanpath representations and comparison methods grounded in spatial trajectories. The present work builds on these foundations by concentrating on *AoI-based sequence* *representations* and on the methods specifically designed to analyze the dynamics, structure, and semantic content of these symbolic gaze sequences.

The objective of this review is to introduce and organize the diverse methodologies used to characterize and analyze *high-level* AoI sequence representations. Section 2 briefly reviews approaches for defining areas of interest, which constitute the basis of this representation. Section 3 examines Markov-based models commonly used to describe transitions between AoIs. Sections 4 and 5 address pattern mining techniques and methods for identifying common or representative AoI sequences across observers. Finally, Section 6 discusses visualization techniques aimed at facilitating the intuitive exploration and interpretation of AoI sequence dynamics. The principal metrics and algorithms reviewed in Sections 3, 4, and 5 are summarized in Table 1.

Before proceeding, two clarifications are necessary. First, although AoI sequences can be treated as symbolic inputs to a variety of sequence analysis methods—many of which were introduced in the *Scanpaths and Derived Representations* review—this article focuses specifically on approaches that analyze AoI sequences in terms of their semantic content and transition structure, rather than their spatial geometry. Consequently, basic AoI-level metrics such as fixation counts or dwell times, which naturally extend the features discussed in earlier parts of this series, are not covered here.

Second, our goal is not to provide an exhaustive technical treatment of each approach, but rather to propose a unified conceptual framework that organizes the diversity of existing methods and clarifies their underlying assumptions, required inputs, and interpretability. Throughout this review, we therefore emphasize conceptual relationships between methods and provide references to formal mathematical descriptions and implementation details, enabling interested readers to pursue more technical developments as needed.

2 AOI DEFINITION

The analysis of gaze behavior through areas of interest (AoIs) first requires a principled definition and identification of these regions within the visual field. To address this foundational step, a wide range of methods and algorithms have been proposed in the literature. These approaches can be broadly grouped into two complementary families: *stimulus-driven* and *data-driven* methods, also commonly referred to as *a priori* and *post hoc* AoI definitions, respectively.

In *stimulus-driven* approaches, areas of interest are defined based on properties intrinsic to the visual stimulus itself. Such properties may include spatial layout, geometric structure, color, contrast, or semantic content that are hypothesized to attract visual attention. These methods typically rely on predefined criteria, often grounded in theoretical or computational models of visual perception, to delineate regions that are expected to be behaviorally or cognitively relevant. By contrast, *data-driven* approaches infer areas of interest directly from experimentally recorded gaze data. In this case, AoIs emerge from the spatial distribution of gaze points and reflect the observer's actual visual exploration behavior. This *post hoc* strategy enables adaptive and empirically grounded AoI definitions, making it particularly well suited to exploratory paradigms, complex scenes, and naturalistic viewing conditions in which relevant visual elements are not known *a priori*.

The following sections provide an overview of these two families of methods, outlining their conceptual foundations, typical implementations, and inherent limitations, and highlighting the implications of AoI definition choices for subsequent analyses of gaze behavior.

2.1 Stimulus-Driven

Historically, areas of interest in eye-tracking studies have most often been defined through manual annotation of the visual stimulus by human analysts. While conceptually simple and intuitive, this approach suffers from several well-documented limitations. Manual annotation is inherently time-consuming, particularly for large datasets or complex stimuli, and often requires the involvement of multiple annotators. This reliance on subjective judgments introduces inter- and intra-annotator variability, which can compromise the robustness, reproducibility, and comparability of AoI-based analyses across studies.

Despite these limitations, manual annotation has long remained a prevalent practice in psychological and behavioral research. Its popularity can be attributed to the simplicity of implementation and to experimental traditions that favor controlled and interpretable stimulus manipulations. As a result, user-defined AoIs have been widely employed in diverse application domains, including facial emotion recognition Vassallo and Douglas (2021), the assessment of task-related expertise Wang et al. (2022), and the analysis of eye movement strategies in scientific problem solving Tang et al. (2016).

A common strategy in manual AoI annotation consists of overlaying simple geometric shapes — such as rectangles, squares, or polygons — onto the stimulus in order to delineate regions of interest. These shape-based AoIs are favored for their ease of specification and computational simplicity. However, they often fail to align with semantically meaningful objects or perceptual units in the visual scene, which can complicate interpretation and, in some cases, lead to misleading conclusions. Nonetheless, shape-based annotation remains effective in highly structured contexts, such as reading, where discrete visual units — *e.g.* words, lines, or sentences — can be naturally approximated by simple geometric regions Just and Carpenter (1980).

To alleviate the subjectivity and scalability issues inherent to manual annotation, a range of automated stimulus-driven approaches has been proposed. One of the earliest such methods consists of partitioning the visual field into a regular grid Mackworth and Morandi (1967), thereby providing a uniform and reproducible segmentation of the stimulus. While straightforward, grid-based approaches ignore the spatial structure and semantic boundaries of visual content, limiting their ability to capture meaningful areas of interest in complex scenes.

More refined stimulus-driven approaches aim to define AoIs based on intrinsic properties of the visual stimulus itself, independently of the observer's gaze behavior. Among these, *saliency-based* methods have emerged as a practical compromise between automation and content awareness. These approaches seek to identify visually prominent regions based on low-level image features, such as contrast, orientation, color, or spatial frequency, and are particularly valuable in situations where data-driven AoI definition is impractical—for example, when gaze data are sparse, noisy, or when multiple potential AoIs overlap spatially. Early work by Privitera and Stark (1998, 2000) advocated segmenting images into coherent regions using feature extraction and clustering techniques, laying the foundations for automated, content-aware AoI definition without manual intervention.

Several stimulus-driven AoI generation methods are rooted in classical image segmentation techniques Shi and Malik (2000); Arbelaez et al. (2010), which aim to partition an image into perceptually homogeneous regions based on color, texture, or edge information. These methods are particularly effective when clear object boundaries are present in the stimulus, enabling precise delineation of visually meaningful regions. Complementary approaches leverage frequency-domain analyses, such as frequency-tuned salient region detection Achanta et al. (2009) or spatio-temporal saliency estimation Guo et al. (2008). By analyzing the amplitude and phase spectra of images or videos, these techniques highlight regions that stand out perceptually, offering computational approximations of bottom-up attentional mechanisms in human vision.

More recently, Fuhl et al. (2018a) proposed an integrated framework for stimulus-driven AoI generation that combines biologically inspired saliency models Itti et al. (1998); Hou and Zhang (2007) with gradient-based image segmentation Fuhl et al. (2015). In this approach, a saliency map is first computed to approximate attentional relevance, and is subsequently segmented to yield precise AoI boundaries adapted to the content of the visual stimulus. While such methods significantly advance the automation and objectivity of AoI definition, they primarily fall within the domain of computer vision. As a result, a comprehensive treatment of these techniques lies beyond the scope of the present review. Broader questions related to the prediction of attention allocation and saliency modeling are discussed in greater detail in the *Scanpaths and Derived Representations* part of this review series Laborde et al. (2024c).

The definition of stimulus-driven AoIs becomes particularly challenging in the context of dynamic stimuli. While manual annotation already poses substantial challenges for static images, these difficulties are amplified for videos, interactive environments, or wearable eye-tracking recordings acquired in unconstrained settings. In such cases, the visual content evolves continuously over time, and spatially similar regions may correspond to different semantic entities at different moments. As discussed in the *Neurophysiology and Experimental Paradigms* part of this review series Laborde et al. (2024a), dynamic gaze mapping can be simplified by projecting gaze points onto a fixed reference frame. However, this strategy does not fully resolve the complexities associated with defining temporally consistent AoIs in changing scenes. Recent advances in computer vision have begun to address these challenges through automatic or semi-automatic annotation tools designed for dynamic stimuli Kurzhals et al. (2016); Kopácsi et al. (2023); Barz et al. (2023). Although these approaches show considerable promise for improving annotation efficiency and consistency, defining reliable and semantically meaningful stimulus-driven AoIs in dynamic environments remains an open and active area of research.

2.2 Data-Driven

Data-driven approaches define areas of interest directly from experimentally recorded gaze data by exploiting the spatial distribution of fixation points. In contrast to stimulus-driven methods, they do not rely on predefined hypotheses about the structure or semantic organization of the visual stimulus. This makes them particularly well suited to exploratory paradigms, naturalistic stimuli, and situations in which the relevant visual elements are unknown or difficult to annotate *a priori*. A key advantage of data-driven AoI definitions lies in their ability to jointly localize regions of interest and quantify the attentional engagement they elicit, purely from observed gaze behavior.

Importantly, data-driven AoIs are not restricted to explicitly fixated visual elements. As emphasized by Boisvert and Bruce (2016), the absence of a direct fixation on a given object — for instance, a face — does not necessarily imply the absence of attentional processing. Peripheral vision, covert attention, and anticipatory gaze behavior may all contribute to visual processing without being reflected in fixation locations alone. By modeling gaze distributions rather than individual fixation targets, data-driven approaches provide a more flexible framework for capturing these broader attentional mechanisms.

To account for uncertainty in gaze direction and measurement noise, many data-driven algorithms implicitly tolerate inaccuracies arising from calibration errors, tracker precision limits, or oculomotor overshoots and undershoots. This tolerance distinguishes them from stimulus-driven approaches, which typically assume precise correspondence between fixation locations and visual targets. Readers interested

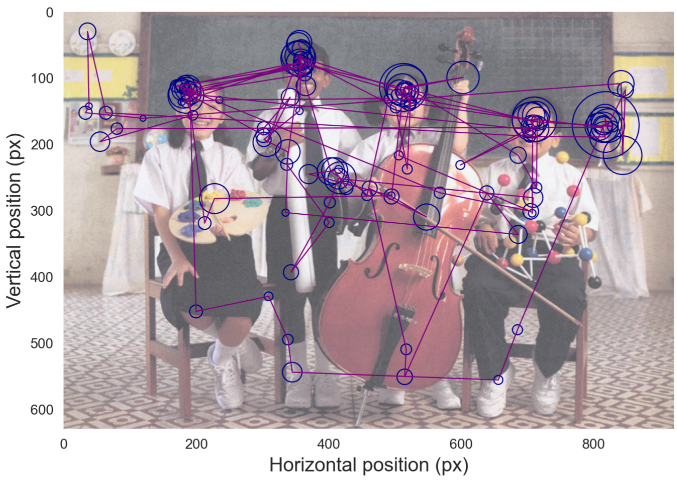

Figure 1a. Input scanpath

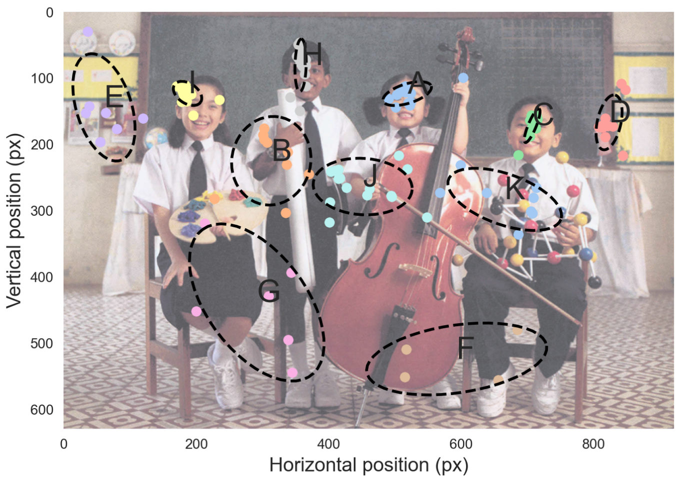

Figure 1b. Areas of interest inferred from *affinity propagation*

Figure 1. Symbolization Process. We illustrate the process of computing areas of interest with a simple example: in Figure 1a, we present a scanpath, which is conventionally used as input for the AoI inference process. Each fixation in the scanpath is utilized to form clusters, here using the *affinity propagation* method. Once formed, these clusters are treated as AoIs. Analyzing the clustering results reveals well-defined AoIs — such as clusters A, C, D, and H — while others appear less relevant, such as clusters F and G.

in theoretical discussions of covert attention and inverse inference in gaze modeling are referred to Boisvert and Bruce (2016) and Haji-Abolhassani and Clark (2014).

A seminal contribution to the methodological foundations of data-driven AoI generation was provided by Santella and DeCarlo (2004), who outlined three desirable properties for clustering-based AoI algorithms: (i) robustness to initialization, (ii) independence from a predefined number of clusters, and (iii) resilience to outliers. These criteria have since served as reference benchmarks for evaluating the suitability of clustering techniques applied to eye movement data. Most data-driven AoI generation methods rely on standard clustering algorithms, adapted to the specific characteristics of fixation distributions. In this framework, fixations belonging to the same spatial cluster are interpreted as instances of a common area of interest. A schematic illustration of this *symbolization* process is provided in Figure 1.

One commonly employed data-driven approach is *k-means* clustering, which is often used either directly (Nguyen et al. (2006); Privitera and Stark (2000)) or in combination with other methods. For instance, Latimer (1988) proposed partitioning gaze data by creating histograms of fixation durations across the visual field, followed by k-means clustering of these histograms to identify well-separated and compact fixation clusters. While k-means is effective in cases where fixation groups are clearly delineated, its performance deteriorates in more complex scenarios, particularly when the gaze data is noisy. In such cases, the algorithm fails to precisely characterize complex AoIs, and its ability to handle more intricate patterns of visual attention is limited. Additionally, k-means clustering requires the user to specify the number of AoIs a priori, which can be problematic in situations where the number of areas of interest is unknown or variable. In practice, this limitation is sometimes mitigated by evaluating several candidate values of k using internal validity criteria such as the silhouette score, which quantifies the compactness and separation of the resulting clusters — note that this procedure increases computational cost. The algorithm is also sensitive to outliers, which can distort the results. For these reasons, the use of k-means clustering may be constrained, leading researchers to seek alternative approaches that offer greater flexibility and robustness in AoI identification.

A more refined alternative to traditional clustering methods is the *mean-shift* approach for defining areas of interest, originally proposed by Fukunaga and Hostetler (1975) and later adapted to eye movement analysis by Santella and DeCarlo (2004). Mean-shift is a non-parametric, density-based technique that operates in two successive stages. In the first stage, known as the *mean-shift procedure*, fixation points are iteratively displaced toward regions of higher local density. Each fixation is shifted according to the weighted mean of its neighbors, where the weights are defined by a kernel function. In practice, a multivariate Gaussian kernel with zero mean and isotropic covariance is commonly employed, with the kernel bandwidth acting as a scale parameter that can be adjusted to reflect the approximate foveal extent of human vision, typically around 5° of visual angle. In the second stage, a distance-based clustering step groups the converged fixation points into distinct clusters, which are subsequently interpreted as areas of interest. A key advantage of the mean-shift approach lies in its robustness to extreme outliers, achieved by limiting the support of the kernel—for example, by assigning zero weight to points located beyond a given multiple of the bandwidth. This mechanism prevents isolated fixations from exerting undue influence on cluster formation. The resulting clusters provide a structured, noise-resistant, and reproducible representation of visual attention patterns (Santella and DeCarlo (2004); Duchowski et al. (2010)). Owing to these properties, mean-shift has become a widely adopted method for AoI generation, particularly in the analysis of noisy or unconstrained eye-tracking data Drusch and Bastien (2012).

Another prominent approach for defining areas of interest is the *density-based spatial clustering of applications with noise* (DBSCAN) method, originally proposed by Ester et al. (1996). DBSCAN has inspired a wide range of density-based clustering techniques and remains one of the most widely adopted algorithms for data-driven AoI identification, owing to its conceptual simplicity and robustness to noise (McGuire and Chakraborty (2016); Marchal et al. (2016); Reani et al. (2018)). The core principle underlying DBSCAN is that clusters correspond to contiguous regions of high point density, separated by regions of lower density. The algorithm operates using two key parameters: a neighborhood radius and a minimum number of points required to define a dense region. Based on these parameters, DBSCAN introduces the notion of reachability: a fixation is considered directly reachable from another if it lies within the neighborhood radius, and indirectly reachable if it can be connected through a chain of directly reachable fixations. As a result, fixation points are iteratively classified into three categories: (i) *core points*, which possess a sufficient number of neighbors within the specified radius; (ii) *border points*, which are reachable from core points but do not themselves meet the density criterion; and (iii) *outliers*, which are not reachable from any core point. All fixations that are mutually reachable are grouped into the same cluster, which is subsequently interpreted as an area of interest. By explicitly identifying noise and eliminating the need for an *a priori* specification of the number of clusters, DBSCAN is particularly well suited for gaze data characterized by irregular cluster shapes and variable fixation densities (Ma and Angryk (2017)). A notable extension of DBSCAN is the *ordering points to identify the clustering structure* (OPTICS) algorithm Naqshbandi et al. (2016). Rather than relying on a single density threshold, OPTICS produces an ordering of fixation points based on their reachability distance, enabling the exploration of clustering structures across multiple density levels. This property makes OPTICS especially valuable for eye-tracking datasets in which fixation density varies substantially across regions or observers.

Another density-based approach for defining areas of interest is the *density peak* clustering method, originally introduced by Rodriguez and Laio (2014) and later adapted to AoI identification in eye-tracking studies by Li and Chen (2018). This method relies on the assumption that cluster centers correspond to fixation points characterized by both high local density and a large distance from other fixation points of higher density. More specifically, the algorithm associates each fixation with two quantities: (i) a local density value, typically defined as the number of neighboring fixation points within a user-defined spatial radius, and (ii) a distance measure corresponding to the minimum distance between the fixation and any other fixation exhibiting higher local density. Fixations that simultaneously exhibit high values for both measures are identified as cluster exemplars, which serve as the centers of the resulting areas of interest. To determine the number of AoIs, the algorithm computes the distance–density product for each fixation and sorts these values in descending order. A threshold is then applied to this product, with fixations exceeding the threshold being assigned as AoI centers. The threshold can be empirically defined using either the arithmetic mean or the geometric mean of the distance–density products. The choice of thresholding strategy directly influences the resulting AoI configuration: arithmetic mean thresholds tend to produce fewer, more compact clusters, whereas geometric mean thresholds typically yield a larger number of clusters with increased overlap. While the density peak method offers an intuitive mechanism for automatically determining the number of clusters, its outcomes are sensitive to parameter selection, particularly the choice of neighborhood radius and thresholding strategy. As a result, careful parameter tuning is required to ensure that the resulting AoIs meaningfully reflect the underlying structure of the gaze data.

An alternative family of data-driven approaches for defining areas of interest is based on *exemplar-based* clustering, among which the *affinity propagation* method, originally introduced by Frey and Dueck (2007), has received particular attention in eye-tracking research (Huang et al. (2009); Li et al. (2017)). Unlike variance-minimization or purely density-based techniques, affinity propagation identifies a subset of representative data points, referred to as *exemplars*, which collectively summarize the structure of the dataset. All fixation points associated with the same exemplar are subsequently grouped into a single cluster, interpreted as an area of interest. Affinity propagation operates through an iterative message-passing procedure between all pairs of data points, based on two complementary quantities: *responsibility* and *availability*. Responsibility reflects how well suited a candidate exemplar point j is to represent a given fixation point i, taking into account the relative distances between points and the current availability values. Conversely, availability measures the degree to which a fixation point i is appropriate to act as an exemplar, based on the accumulated evidence provided by other fixation points in the dataset. During the iterative optimization process, responsibility and availability values are updated until convergence. At convergence,

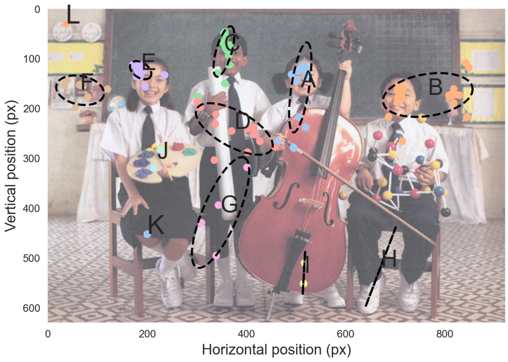

Figure 2a. Areas of interest inferred from *mean-shift*

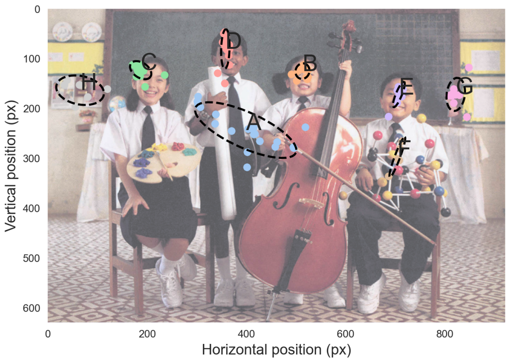

Figure 2b. Areas of interest inferred from *DBSCAN*

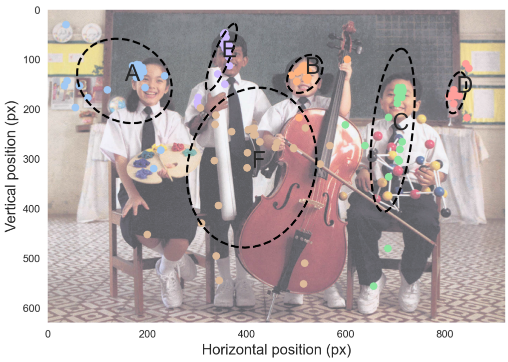

Figure 2c. Areas of interest inferred from *density peak*

Figure 2. Symbolization Process. We illustrate the data-driven AoIs computed using the various algorithms discussed in Section 2.2. A qualitative analysis of the clusters — representing areas of interest — generated by these methods reveals significant variability in results across algorithms. This variability is further amplified by the dependency of many algorithms on one or more parameters, which increases the sensitivity and variability of the outcomes. Therefore, researchers must exercise caution when selecting their approaches and consider adapting their methodology based on the visual context and the specific tasks performed by the viewers.

the set of exemplars emerges naturally from the data, eliminating the need for an *a priori* specification of the number of clusters. Fixations are then assigned to the exemplar with which they exhibit the highest affinity, yielding a partition of the dataset into distinct AoIs. Affinity propagation satisfies several key criteria for robust AoI generation outlined by Santella and DeCarlo (2004), notably independence from initialization and from predefined cluster numbers. In addition, the method exhibits a degree of robustness to outliers by forming small auxiliary clusters that isolate atypical fixations, which can be readily identified and excluded from further analysis (Huang et al. (2009); Li et al. (2017)). These properties make affinity propagation a compelling option for AoI identification in gaze datasets characterized by complex spatial structures or heterogeneous fixation distributions.

Despite their methodological diversity, a comparative study by Fuhl et al. (2018b) has shown that applying different data-driven AoI generation algorithms to the same gaze dataset can result in markedly different AoI configurations. In particular, approaches based on local maxima thresholding, heatmap gradient analysis, or overlap-based clustering—each grounded in plausible density-driven heuristics—were shown to produce substantially different scene partitions. This observation highlights a fundamental limitation of data-driven AoI definition: the problem is inherently ill-posed, and no single algorithm can be regarded as universally optimal. To illustrate this variability within a unified methodological framework, Figure 2 presents AoI configurations obtained using the principal data-driven clustering algorithms detailed in this section. Once defined, these data-driven AoIs constitute the symbolic alphabet from which AoI sequences are constructed, making the properties of the AoI definition stage critical for all subsequent sequence-based analyses.

At this point, a few remarks are in order. Firstly, a number of Markov-based methods have been introduced in the literature to define AoIs. These approaches however will be discussed in the Section 3, as they combine the tasks of defining areas of interest and analyzing the attention transition dynamics between these regions. Secondly, by implementing the various methods mentioned above, we can assign the fixation points to distinct areas of interest. Nevertheless, in order to preserve the spatial information of the gaze trajectories, it may be necessary to take into account the locations of the AoIs, and in particular their centers. A first approach, which despite its simplicity remains very popular, simply consists of averaging the coordinates of the different fixations assigned to the same cluster — or AoI. However, such a basic treatment could ignore the spatial relationship among fixations and be sensitive to noise. Alternative approaches have therefore been developed to refine this process. For instance, Pillalamarri et al. (1993) early included several methods to calculate the center of a cluster as the weighted — by duration — mean of clustered fixations and the center of the convex polygon encompassing fixations belonging to the same cluster (Špakov and Miniotas (2007)). On the other hand, Wang et al. (2016) proposed to compute a density value — the number of fixations within a given distance threshold — around each fixation of a given cluster and then determine the maximum density value which points to the AoI center. We can also mention the approach suggested by Chen and Chen (2017) who proposed a method based on random walks to identify AoI centers, highlighting the robustness and low susceptibility to outliers of such an approach (**?**.

Finally, we conclude this section by pointing out that, in general, the problem of defining areas of interest has been little explored in the specific case of dynamic stimuli. Indeed, in such a context, spatially identical or close but temporally different fixations may originate from visual elements with totally distinct semantic meanings. Consequently, current data-driven approaches may fail to define precise areas of interest and lead to erroneous conclusions when used as input data for the examination of dynamic eye movements. Recent approaches are beginning to address this issue. We can mention the *AOI bounding boxes estimation* method that uses image segmentation technique to estimate AoI bounding boxes in dynamic stimuli (Lagmay et al.

Method name	Input	Description	Reference
Transition matrix	AoI sequence	Computes transition probabilities between areas of interest as a row-normalized stochastic matrix, from which graph-based metrics — *e.g.* density, centrality, transitivity, small-worldness and global efficiency — can be derived.	Hsiao et al. (2021)
Transition entropy	Transition matrix	Computes entropy-based metrics characterizing gaze dynamics, including stationary entropy, joint entropy, conditional entropy, and mutual information.	Krejtz et al. (2015)
Hidden Markov model (HMM)	Fixation sequence	Learns a probabilistic generative model of gaze behavior by estimating hidden states and transition probabilities using the Baum–Welch algorithm.	Bilmes et al. (1998)
Fisher vector (HMM-based)	Hidden Markov model	Encodes gaze sequences by computing gradients of the HMM likelihood with respect to model parameters, yielding a fixed-length vector representation.	Kanan et al. (2014)
VHEM clustering	Set of HMMs	Clusters hidden Markov models into groups of similar gaze dynamics and estimates representative HMMs for each cluster using variational inference.	Coviello et al. (2012)
Lempel–Ziv complexity	AoI sequence	Computes the complexity of visual exploration by incrementally identifying novel subsequences while scanning the AoI sequence from left to right.	Lounis et al. (2020)
Common n-gram analysis	AoI sequence	Computes frequency distributions of fixed-length AoI subsequences to capture higher-order transition patterns and local sequential structure.	Ayres et al. (2002)
Sequential pattern mining	Set of AoI sequences	Identifies frequent sequential patterns shared across multiple AoI sequences using data mining algorithms such as Apriori or SPAM.	Ayres et al. (2002)
T-pattern detection	AoI sequence with timestamps	Identifies statistically recurrent temporal patterns in AoI sequences by detecting event structures with consistent inter-event intervals.	Salah et al. (2010)
Longest common subsequence (LCS)	Pair of AoI sequences	Computes the longest subsequence shared by two AoI sequences using dynamic programming, allowing gaps but preserving order.	Bergroth et al. (2000)
Smith–Waterman alignment	Pair of AoI sequences	Performs local sequence alignment using a substitution matrix and gap penalties to identify similar AoI sub-sequences.	West et al. (2006)
eMINE	Set of AoI sequences	Iteratively derives a consensus AoI sequence by combining global similarity measures and longest common subsequence extraction.	Eraslan et al. (2014)
Scanpath trend analysis (STA)	Set of AoI sequences	Identifies commonly visited and temporally significant AoIs and constructs a representative trend sequence based on fixation frequency and dwell time.	Eraslan et al. (2016)
Candidate-Constrained DTW Barycenter Averaging	Set of AoI sequences	Aggregates multiple AoI sequences into a representative sequence by iteratively minimizing the average dynamic time warping distance.	Li et al. (2017)
Dotplot hierarchical clustering	Set of AoI sequences	Extracts common AoI sub-sequences using dotplot analysis and hierarchical clustering based on sequential matching distances.	Goldberg and Helfman (2010)

Table 1. Methods for AoI sequence analysis and their required input representations.

(2022)). Conversely, recent publications have proposed leveraging deep learning techniques to identify AoIs within dynamic visual stimuli (Jayawardena and Jayarathna (2021); Barz et al. (2023); Trajkovska et al. (2024)). While these approaches appear promising, defining areas of interest in the context of dynamic stimuli remains a critical research avenue to be further explored in the coming years.

3 MARKOV-BASED MODELS

Markov-based analysis provides a probabilistic framework for modeling sequential data by describing system dynamics in terms of state-to-state transition probabilities. A process is said to satisfy the Markov property when the probability of transitioning to a future state depends solely on the current state, independently of the sequence of states that preceded it. While this assumption deliberately abstracts away long-range dependencies, it yields models that are theoretically well grounded, interpretable, and computationally efficient, making them particularly attractive for the analysis of sequential behaviors.

In the context of eye-tracking, Markov models are naturally suited to the analysis of gaze dynamics expressed as sequences of areas of interest. Once gaze trajectories have been discretized into AoI sequences, each AoI can be interpreted as a discrete state, and eye movements correspond to probabilistic transitions between these states. Modeling these transitions using Markov-based approaches provides a compact description of visual exploration behavior, capturing systematic patterns in attentional shifts while offering a flexible foundation for both descriptive and inferential analyses.

3.1 Transition Matrix

Markov-based approaches aim to model transitions between visual elements by estimating transition probabilities, often represented in the form of *transition matrices*. A transition matrix — also referred to as a stochastic or probability matrix — is a square matrix in which each entry corresponds to the probability of transitioning from one current state — *e.g.* an AoI — to another. In an early application of this concept, Ponsoda et al. (1995) developed transition matrices to categorize successive saccade directions. Rather than focusing on AoIs — as is common in modern applications — their matrices represented transitions between eight cardinal saccade directions. Building on this, Bednarik et al. (2005) constructed transition matrices to characterize visual switching behavior across predefined AoIs, providing a framework for analyzing gaze transitions. Fischer and Peinsipp-Byma (2007) further extended the use of transition matrices by combining them with descriptive statistics to compare perceptual responses to the same stimulus. Transition matrices thus offer a powerful tool for summarizing and analyzing gaze patterns, supporting insights into both individual and group-level visual processing behaviors.

From a computational standpoint, the elements of a first-order transition matrix are derived from the observed frequencies of transitions between areas of interest. Each matrix entry represents the count of transitions from a specific source AoI to a corresponding destination AoI. To compute empirical probabilities, the matrix is row-normalized, ensuring that the sum of the values in each row equals 1. Consequently, the element in row i and column j reflects the conditional probability of transitioning to the j-th AoI, given that the current fixation is located in the i-th AoI.

Although transition matrices offer valuable insights into gaze dynamics, their direct application and analysis remain relatively underexplored in the literature. Nonetheless, several studies have leveraged transition probabilities to infer patterns and dynamics of visual behavior directly (Bednarik et al. (2005); Fischer and Peinsipp-Byma (2007); Vandeberg et al. (2013)). Furthermore, Privitera and Stark (1998) proposed a method for comparing transition matrices by introducing a distance metric designed to quantify similarities or differences in gaze behavior patterns, paving the way for more nuanced analyses of visual attention.

Beyond their direct probabilistic interpretation, transition matrices can also be reformulated into alternative representations that emphasize the structural organization of gaze behavior. Another element to consider is the directed graph, which can be directly derived from the transition matrix, as illustrated in Figure 3b. While directed graphs are often introduced primarily as visualization tools, they also constitute a powerful analytical representation of gaze dynamics. In this framework, areas of interest are modeled as nodes, and gaze transitions between them are represented as directed edges whose weights correspond to transition frequencies or probabilities. This graph-based abstraction makes it possible to analyze eye-movement behavior using concepts and metrics drawn from network theory, thereby shifting the interpretation of gaze from a purely sequential process toward a structured visual network.

Recent work has demonstrated the relevance of this perspective for characterizing individual differences in visual strategies. For instance, Ma et al. (2023) leveraged graph-theoretic measures computed on gaze transition networks to reveal systematic differences between low-ability and high-ability readers. Their results illustrate how the organization of gaze transitions — rather than isolated fixation metrics — can capture meaningful aspects of cognitive processing during reading.

Taken together, graph-theoretic measures provide a compact yet expressive description of gaze dynamics by characterizing both local transition patterns and global network organization. Network *density* reflects the overall richness of visual exploration by quantifying how many transitions between areas of interest are effectively used relative to all possible transitions. Measures of node *centrality* identify areas of interest that play a dominant role in the visual strategy, either by attracting a large proportion of incoming transitions or by acting as frequent intermediaries between other regions. Furthermore *transitivity* captures the tendency of gaze transitions to form locally clustered structures, revealing whether visual exploration favors tightly interconnected subsets of areas, while *small-worldness* describes the coexistence of such local clustering with short paths between distant regions, indicating an efficient balance between focused inspection and broader scene integration. Finally, *global efficiency* summarizes how easily information can propagate across the entire network, providing an aggregate measure of how coherently different areas of interest are connected through gaze behavior.

By embedding gaze transitions within a network representation, directed graphs enable the analysis of eye movements at a level that goes beyond pairwise transitions. They allow gaze behavior to be interpreted as an organized system whose structure reflects underlying perceptual, attentional, and cognitive strategies, complementing both transition-matrix-based analyses and more complex probabilistic models introduced in subsequent sections.

As previously discussed, a critical characteristic of Markov processes is that the next state — the target AoI — depends solely on the current state, defined as the AoI currently being fixated upon by the observer. This property allows for a mathematically tractable representation of AoI transitions but introduces a significant limitation. Specifically, the assumption that gaze transitions depend only on the current AoI disregards the prior sequence of fixations, effectively ignoring the historical context of gaze behavior. While this simplification may be adequate for straightforward perceptual tasks, it poses challenges for analyzing tasks involving higher-order cognitive processes, where sequential dependencies likely play a critical role.

In principle, this limitation can be mitigated by constructing higher-order Markov models that incorporate previously visited AoIs into the transition framework. However, this approach leads to a rapid increase in the number of transition probabilities that must be estimated. Such complexity can make higher-order Markov processes impractical for studying visual dynamics, especially in scenarios where the available gaze data are insufficient to reliably compute all entries in the extended transition matrix (Burmester and Mast (2010); von der Malsburg et al. (2015)).

To capture longer-term dependencies in gaze behavior, the *successor representation* (SR) matrix, originally introduced by Dayan (1993), has been adapted for use in the visual domain (Hayes et al. (2011)). The central idea of this approach is to encode not only direct transitions between areas of interest but also the expected sequence of future AoIs inferred from prior observations. In practice, when a transition occurs between AoIs, the SR matrix is updated to associate the originating AoI not only with the immediate target but also with the set of AoIs expected to follow. These associations are weighted by a temporal discount factor, which accounts for the diminishing influence of more distant future states. This approach allows the SR method to provide a richer, temporally extended representation of gaze behavior, accommodating sequential dependencies more effectively.

Finally, transition matrices play a pivotal role as foundational components in advanced methodologies, including *hidden Markov models* and *entropy-based analyses*. These approaches leverage the probabilistic structure of transition matrices to model and quantify gaze behavior with greater precision and depth. By capturing the underlying dynamics of visual attention, these methods facilitate a more comprehensive understanding of complex gaze patterns. The intricacies of these advanced applications and their broader implications will be examined in detail in subsequent sections.

3.2 Gaze Structure Entropy

Entropy, as originally formalized by Shannon (1948), provides a principled measure of uncertainty or unpredictability within a system. When gaze behavior is modeled as a Markov process through an AoI transition matrix, entropy-based measures offer a compact and interpretable way to quantify the structure and complexity of visual exploration. Several entropy metrics have been proposed in this context, each capturing a distinct aspect of gaze organization. Among these, the *stationary entropy* constitutes a foundational descriptor of gaze distribution.

Under standard assumptions that the Markov chain is irreducible and aperiodic, it admits a unique stationary distribution, defined as the left eigenvector of the transition probability matrix associated with the eigenvalue one (Meyer (2000)). This stationary distribution characterizes the long-term proportion of time the gaze is expected to spend in each area of interest, independently of the initial fixation state. Applying Shannon entropy to this distribution therefore quantifies the uncertainty associated with gaze position across AoIs (Krejtz et al. (2014)). Low stationary entropy values indicate that visual attention is preferentially concentrated on a limited subset of regions, whereas higher values reflect a more uniform allocation of gaze across the visual scene.

While stationary entropy captures the spatial dispersion of gaze over time, it does not account for the dynamics of transitions between areas of interest. To address this limitation, gaze behavior has also been conceptualized as a discrete information channel, referred to as the *gaze information channel* (Hao et al. (2019, 2020)). Within this information-theoretic framework, the *conditional entropy* is defined as the average entropy of the conditional transition probabilities associated with each AoI (Krejtz et al. (2015); Hao et al. (2020)). This metric quantifies the degree of uncertainty in gaze transitions themselves, capturing how predictable or stochastic the shifts of attention are from one region to another. Higher conditional entropy values correspond to more exploratory and less predictable transition patterns, whereas lower values indicate more structured, repetitive, or goal-directed viewing behavior.

Beyond stationary and conditional entropy, the gaze information channel framework enables the definition of additional measures that further decompose visual uncertainty and information flow. These include metrics quantifying total uncertainty in gaze behavior, uncertainty conditioned on a specific AoI, as well as mutual information exchanged between pairs of areas of interest or between a given AoI and the rest of the

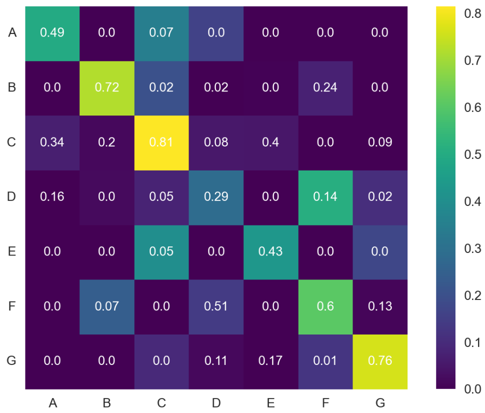

Figure 3a. Transition matrix

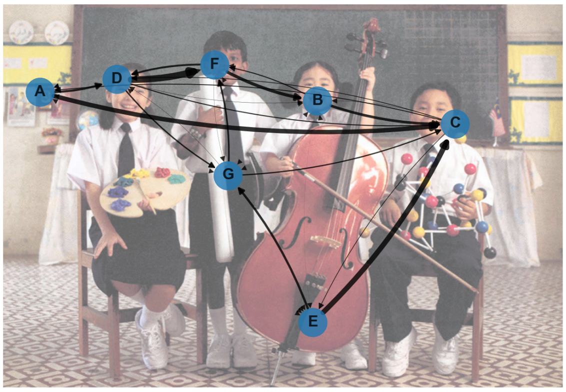

Figure 3b. Directed graph

Figure 3. Infered Markov model. Modeling gaze dynamics as a Markovian process is demonstrated in Figure 3a, which illustrates the transition matrix encoding the probabilities of shifting from one AoI to another. Additionally, a qualitative analysis of the directed graph associated with this transition matrix, presented in Figure 3b, offers valuable insights. This graph is overlaid on the visual stimulus presented during gaze recording, providing contextual information that enhances the interpretation of the observed transitions.

scene. Although a detailed exposition of these quantities lies beyond the scope of the present review, they provide complementary perspectives on how visual information is distributed and transferred during gaze exploration. Readers interested in the formal definitions and behavioral interpretations of these measures are referred to the comprehensive treatment by Hao et al. (2019).

When applied to eye-tracking data, entropy-based metrics offer several practical advantages. First, they yield scalar descriptors of gaze complexity at the individual level, facilitating aggregation across participants and experimental conditions. Second, they are well suited for inferential statistical analyses, such as group comparisons, correlation analyses, or regression modeling, enabling systematic investigation of individual differences in visual behavior. Entropy measures have, for instance, been used to distinguish eye-movement strategies between expert and novice teachers (McIntyre et al. (2017); Kosel et al. (2021)) and to quantify variations in visual search behavior as a function of task complexity or cognitive load (Krejtz et al. (2014)).

Taken together, gaze structure entropy metrics provide a theoretically grounded and computationally efficient means of summarizing both the spatial distribution and temporal organization of visual attention. By complementing transition-matrix-based and graph-based analyses, entropy measures contribute to a more comprehensive characterization of gaze dynamics, while naturally motivating the use of more expressive probabilistic models discussed in the following sections.

3.3 Hidden Markov Models

Recent studies have employed *Hidden markov models* (HMMs) to capture the temporal dynamics of gaze behavior. Like first-order Markov models, HMMs govern the transitions between states through a first-order Markov process. However, the key distinction in HMMs lies in the term *hidden*, which indicates that the states themselves are not directly observable. Instead, they must be inferred from the relationship between the hidden states — representing areas of interest — and the experimentally observed gaze positions (Chuk et al. (2014)).

Specifically, the emission densities, which describe the distribution of successive fixations within each AoI, are modeled using two-dimensional Gaussian distributions. The transitions between hidden states are governed by a probability distribution, which is encapsulated in the HMM's transition matrix. The inference process, which aims to estimate the hidden states, is typically performed using the Baum-Welch algorithm, also known as the forward-backward algorithm. This dynamic programming approach is a special case of the expectation-maximization algorithm (Bilmes et al. (1998)). Its purpose is to iteratively adjust the HMM parameters — namely, the state transition matrix, the Gaussian emission parameters, and the initial state distribution — to maximize the likelihood of the model fitting the observed gaze data.

HMM-based approaches have been proposed to model the visual attention process (Haji-Abolhassani and Clark (2014)). If early works assumed that the attentional targets could be defined in advance — for instance building on saliency approaches discussed in the *Scanpaths and Derived Representations* part of this review series (Laborde et al. (2024c)) — more recent works emphasize on the fact that during the inference of HMM-based models, the target selection and AoI centers is governed by a statistical process that is trained on natural eye trajectories measured during task execution. Indeed, if the assumption of *a priori* targets may be valid for simple tasks with objective results, it can appear limited in more abstract or complex tasks where defining the potential targets of attention is not straightforward and typically no prior information about the location of the targets is available to the model.

In addition, Gaussian distributions - commonly used as emission probabilities - also take into account overshoots and undershoots when directing the gaze at a given target. Consequently, the assumption of accurate targeting is relaxed in the HMM by using observational distributions over visual targets. In other words, this approach dispenses with the need for a preliminary data-driven definition of AoIs, directly integrated into the analysis of visual dynamics. As such, one popular proposed approach begins by using the k-means clustering technique Kaufman and Rousseeuw (2009) to locate — as a first approximation — potential targets in an image or scene and then uses the centroids of these clusters to initialize the HMM-based method for decoding eye dynamics and refining visual targets.

From this HMM modeling framework, Haji-Abolhassani and Clark (2014, 2013) proposed the computation of task-dependent HMMs, trained with task-specific eye trajectories under the constraint of a shared covariance matrix across all Gaussian emission densities. This approach allows for the creation of

HMMs tailored to specific visual tasks. Experimental gaze data can then be probabilistically assigned to a task using a Bayesian inference framework. In this context, the problem is conceptualized as the *inverse Yarbus process*, where the likelihood term is determined by the probability of the observed gaze sequence given the task-dependent HMMs previously trained. Similarly, Coutrot et al. (2018) proposed leveraging the HMM representation of gaze dynamics to infer both tasks and stimuli during eye-tracking experiments. In their approach, HMMs are first used to *encapsulate* the dynamics of individual gaze behavior. The resulting HMM parameters — namely the priors, transition matrix coefficients, and the center and covariance matrix coefficients of the Gaussian emissions — are then used as inputs for multiple machine learning-based classifiers. These approaches, however, must be used with care, as small modification of these parameters, can readily result in misclassification (Ellahi (2022)).

A significant limitation of the HMM approach is that the number of states must be defined prior to training. This requirement is feasible for images with a predefined number of areas of interest but becomes problematic for more complex or naturalistic stimuli. The *a priori* assumption of a fixed number of states holds primarily for simple visual stimuli, such as synthetic images. To ensure a more data-driven approach, the number of states should not be predetermined but instead optimized based on the recorded eye-tracking data. One possibility is to employ validation and robustness measures to determine the number of clusters, such as using the silhouette score (Elbattah et al. (2023)) during the k-means clustering step that initializes HMM inference or selecting the number of clusters that maximizes the *a posteriori* probability of the training data (Haji-Abolhassani and Clark (2014)).

Alternatively, several authors have proposed incorporating the selection of the number of clusters directly into the HMM inference process, thereby eliminating the need for separate initialization steps and enhancing the adaptability of the model. The *variational approach to Bayesian inference* offers a solution for simultaneously estimating model parameters and determining model complexity (McGrory and Titterington (2009)). Specifically, variational inference places prior distributions on the parameters of the emission densities, transition matrix, and initial state distribution, factoring them to compute the maximum *a posteriori* estimate (Bishop (2006)). This method inherently determines the optimal number of hidden states by pruning unnecessary ones, enabling an automatic selection of model complexity, including the determination of the number of AoIs (Chuk et al. (2014, 2017); Coutrot et al. (2018)).

On the other hand, the gaze representation offered by HMMs can be made more compact using Fisher vectors, which are a concatenation of normalized HMM parameters into a single vector. For instance, Kanan et al. (2014, 2015) proposed to use Fisher vectors as input features for common classifications algorithms — *e.g.* support vector machine classification algorithm — to infer an observer's task. To compute Fisher vectors an HMM — for instance with Gaussian emissions — is learned using maximum likelihood estimation. The parameters of the HMM — including transition probabilities and the probability of observing particular AoI fixation given the internal state — will then reflect the sequential information in the data. The idea of a Fisher kernel is to compute how a new sequence of gaze data would change the parameters of the model — *i.e.* the parameter gradients of the generative model when given a novel gaze trajectory as input. Subsequently, two gaze trajectories from the same category will likely change the model in the same way.

Finally, a variational approach (Coviello et al. (2014); Chuk et al. (2014)) has been proposed for clustering individuals HMM. The method, referred to as *variational hierarchical expectation maximization* (VHEM) algorithm, clusters HMMs based on the hierarchical EM (HEM) algorithm. This approach allows clustering a given collection of HMMs into groups of HMMs that are similar, in terms of the distributions they represent. Furthermore, it characterizes each group by a *cluster center*, that is, a novel HMM that is representative for the group. To cope with intractable inference in the E-step, the HEM algorithm is formulated as a variational optimization problem, and efficiently solved for the HMM case by leveraging an appropriate variational approximation.

One of the main advantages of the VHEM approach lies in the fact that, once the clusters have been learned, a gaze trajectory from a new observer can be compared with these models using a Viterbi decoder — *i.e.* classified into the most similar class, considering the estimated log-likelihood score. For each cluster, the VHEM algorithm also produces a representation HMM — cluster center — which summarizes the common AoIs and transitions in the cluster. Interestingly, if this approach does not produce a representative sequence — see Section 5 — it produces a representative dynamical model between gaze data similarly clustered.

4 PATTERN MINING AND CHARACTERIZATION

As discussed in the previous section, the Markov property provides a convenient and interpretable framework for modeling gaze transitions between areas of interest. However, by construction, first-order Markov models remain limited in their ability to capture long-range dependencies and higher-order structures within gaze sequences. This limitation becomes particularly salient in cognitive tasks that involve planning, reasoning, or strategy formation, where visual behavior often unfolds over extended temporal scales.

To address these limitations, a broad family of methods has been developed to identify and characterize recurrent patterns within AoI sequences. Rather than modeling gaze behavior as a sequence of memoryless transitions, these approaches explicitly seek structured motifs, subsequences, or regularities that recur across time or across observers. Such patterns may reflect stable visual strategies, task-specific routines, or shared modes of information extraction that are not readily captured by transition-based models.

Pattern detection techniques span a wide methodological spectrum, ranging from exhaustive searches for all subsequences of a given length to more flexible approaches that allow for temporal gaps, substitutions, or variations in pattern structure. Importantly, these methods originate from the broader field of pattern mining and sequential data analysis and are not specific to eye-tracking data. Their application to AoI sequences therefore represents a methodological transfer, enabling gaze behavior to be analyzed using tools originally developed for domains such as text processing, bioinformatics, or user interaction modeling.

In the context of gaze analysis, pattern mining approaches provide a complementary perspective to probabilistic and graph-based models. By focusing on the identification of meaningful and recurrent subsequences, they offer a direct way to uncover higher-level regularities in visual exploration, facilitating the comparison of gaze strategies across observers, tasks, or experimental conditions. The following sections review the main families of pattern mining methods that have been applied to AoI sequences, along with their underlying assumptions and interpretative implications.

4.1 Lempel–Ziv Complexity

The complexity of visual scanning behavior can be quantified using *Lempel–Ziv complexity* (LZC), a general-purpose measure originally introduced in the context of information theory to characterize the structural richness of symbolic sequences (Lempel and Ziv (1976)). In recent years, this metric has been adapted to the analysis of eye-tracking data by applying it to sequences of transitions between areas of interest, thereby providing a global, model-free descriptor of gaze dynamics (Lounis et al. (2020)). Within the broader landscape of pattern mining approaches, LZC occupies a distinctive position, as it quantifies the overall sequential complexity of gaze behavior without explicitly defining or extracting individual patterns.

Algorithmically, Lempel–Ziv complexity can be understood as an incremental parsing procedure applied to a symbolic sequence. The AoI sequence is scanned from left to right, and each time a novel symbol or combination of symbols is encountered — one that cannot be reconstructed from previously observed subsequences — it is added to a dynamically growing codebook. This codebook thus represents the minimal set of distinct subsequences required to encode the original sequence. Rather than identifying predefined motifs, the parsing process implicitly captures the emergence of new structures as the sequence unfolds, making LZC sensitive to both local novelty and global organization.

Subsequently, the Lempel–Ziv complexity of a sequence is defined as the size of this codebook, that is, the number of distinct subsequences identified during the parsing process. Sequences characterized by highly regular, repetitive, or stereotyped gaze transitions typically yield small codebooks and low complexity values. In contrast, sequences that exhibit diverse transition structures, frequent novelty, or less predictable ordering lead to larger codebooks and higher complexity scores. When applied to AoI sequences, LZC therefore provides a compact summary of the richness of visual exploration strategies, reflecting how constrained or diversified gaze behavior is over time.

One of the main advantages of LZC lies in its simplicity and computational efficiency. Because it does not require an explicit definition of what constitutes a *pattern*, nor does it rely on assumptions about the temporal scale, duration, or semantic relevance of gaze regularities, LZC can be readily applied to a wide range of experimental paradigms. This property makes it particularly attractive for the analysis of naturalistic or unconstrained viewing conditions, where gaze behavior may be highly variable across observers and tasks, and where meaningful patterns may not conform to simple or predefined templates.

At the same time, several considerations are important when interpreting LZC values in the context of gaze analysis. First, LZC is inherently sensitive to sequence length, as longer sequences provide more opportunities for novel subsequences to emerge. As a result, comparisons across observers or conditions typically require controlling for sequence length or applying appropriate normalization procedures. Second, LZC depends on the symbolization scheme used to generate AoI sequences, meaning that differences in AoI definition or granularity can directly influence complexity estimates. These dependencies highlight that LZC should be interpreted as a relative rather than absolute measure of gaze complexity, meaningful primarily within a given representational framework.

Finally, while LZC effectively captures the presence and degree of sequential structure, it does not explicitly identify which subsequences or motifs contribute to this complexity. Consequently, LZC is best viewed as a global descriptor that complements, rather than replaces, pattern mining methods aimed at extracting and characterizing specific gaze patterns. In the following sections, we review approaches that explicitly enumerate, compare, and summarize recurrent subsequences within AoI sequences, thereby providing a more detailed and interpretable account of visual scanning strategies.

4.2 Common n-gram sequences

The use of composite probabilities derived from first — or second — order Markov chains has proven to be a valuable approach for modeling visual perception in simple tasks, particularly when areas of interest are not predefined, as discussed in Section 3. Markov models rely on a limited-memory assumption, whereby the probability of transitioning to a given AoI depends solely on the current state. This property makes them well suited for tasks involving relatively straightforward perceptual processes. However, this memoryless assumption may become overly restrictive when tasks require higher-order cognitive functions or complex decision-making. To overcome these limitations, methods such as n-gram transition analysis have been introduced. These approaches enable a more expressive representation of gaze transitions by explicitly capturing higher-order dependencies across successive AoIs, thereby facilitating a richer characterization of visual dynamics in complex tasks.

Common n-gram sequences are a foundational concept in natural language processing (NLP), where they have long been used for language modeling and text prediction. They have also found wide application in bioinformatics, particularly for genome and transcriptome assembly, metagenomic sequencing, and error correction of sequence reads, where they are commonly referred to as k-mers (Melsted and Pritchard (2011)). In the context of AoI sequence analysis, an n-gram corresponds to a contiguous subsequence of n successive AoIs. Under this formulation, an n-gram model predicts the occurrence of a given AoI based on the $n-1$ preceding AoIs, thereby encoding local temporal dependencies that extend beyond first-order transitions.

To compute n-gram statistics, combinatorial methods are used to enumerate all possible subsequences of length n defined over a given AoI alphabet. The occurrences of these combinations within the AoI sequence are then counted and normalized by the total number of subsequences, yielding a distribution of n-gram frequencies. This procedure can be applied to simple transitions between two AoIs, referred to as *bi-grams*, or extended to longer n-grams that capture increasingly rich temporal context. In principle, there is no upper bound on the length of n-grams that can be analyzed. In practice, however, the number of unique n-grams grows exponentially with n, leading to increasingly sparse distributions that can undermine interpretability and statistical robustness (Kübler et al. (2017)). Consequently, the use of long n-grams is typically constrained by computational feasibility and diminishing returns in terms of informative content.

In the specific context of AoI sequence analysis, n-gram methods have been applied to a variety of cognitive studies to uncover underlying reasoning processes. For instance, Plummer et al. (2017) employed tri-grams of AoI fixations to investigate how individuals reason about rational numbers, comparing fixation sequences to determine whether observers relied on counting-based or comparison-based strategies. This work illustrates how n-gram analysis can be used not only to describe gaze patterns, but also to probe the temporal organization of information processing during reasoning tasks.

Beyond task-specific investigations, n-gram analysis has proven valuable in broader contexts aimed at characterizing perceptual efficiency and expertise. For instance, Lounis et al. (2021) examined the frequency of increasing-length n-grams to compare novice and expert pilots, while Wang et al. (2022) used n-gram statistics to quantify cognitive processes and learning progression across trials in an airplane assembly task. These studies demonstrate the utility of n-gram analysis for identifying patterns that differentiate expertise levels or cognitive stages during task performance.

Several authors have further investigated the influence of n-gram length on the ability to discriminate between observers or groups. In particular, Reani et al. (2018) reported that subsequences of four or more AoIs were often ineffective at distinguishing between groups in complex reasoning tasks, suggesting that relatively short n-grams may be more informative for capturing meaningful patterns in human gaze behavior. This finding emphasizes the importance of selecting an appropriate n-gram length that balances representational expressiveness with statistical robustness.

To compare n-gram frequency distributions across observers or experimental conditions, divergence measures such as the *Hellinger distance* (Hellinger (1909)) have been shown to be particularly effective (Reani et al. (2018)). This distance provides a stable, non-parametric measure of dissimilarity between discrete probability distributions and is less sensitive to zero-probability events than alternatives such as the Kullback–Leibler divergence (Van Erven and Harremos (2014)), making it well suited for the sparse distributions commonly encountered in n-gram analysis.

In a broader methodological context, it is worth noting that n-gram analysis constitutes a foundational component of more advanced sequence comparison and characterization approaches. For example, the *Subsmatch* algorithm — described in the *Scanpaths and Derived Representations* part of this review series — relies explicitly on n-gram representations to compare gaze sequences. More generally, n-grams serve as versatile building blocks across a wide range of algorithms in different domains, underscoring their broad applicability and methodological significance.

Despite their flexibility and interpretability, n-gram approaches remain limited by their reliance on fixed-length, contiguous subsequences. They do not account for temporal gaps, variability in pattern realization, or hierarchical organization of gaze behavior. As a result, while n-grams provide valuable descriptive insight into local sequential structure, they may fail to capture more complex or temporally flexible patterns of visual exploration. These limitations motivate the use of more expressive pattern mining techniques, which aim to identify recurrent gaze patterns while allowing for temporal variability and structural flexibility. Such approaches are reviewed in the following sections.

4.3 Sequential Pattern Mining

A distinct approach to analyzing AoI patterns of varying lengths relies on *sequential pattern mining*, a specialized subfield of frequent pattern mining originally developed in the context of market basket analysis (Agrawal et al. (1993)). Unlike classical frequent itemset mining, sequential pattern mining explicitly incorporates temporal ordering constraints, enabling the discovery of patterns that unfold over time (Agrawal and Srikant (1995)). More generally, sequential pattern mining algorithms are applicable whenever the objective is to identify frequent subsequences in datasets composed of temporally ordered elements, making them particularly well suited for the analysis of gaze trajectories expressed as AoI sequences.

In eye-movement research, one of the earliest and most widely adopted families of sequential pattern mining methods is based on the *Apriori* principle. Apriori-based sequential mining algorithms aim to discover frequent ordered subsequences by exploiting the anti-monotonicity property of pattern support, which allows infrequent candidates to be pruned early during the search process. These methods provide a natural extension of frequent itemset mining to temporally ordered data and have been extensively used in domains where interpretability and explicit pattern enumeration are desired. In the context of gaze analysis, several *Apriori*-based sequential mining algorithms have been applied to analyze AoI patterns. For instance, such approaches have been used to study how undergraduate computer science students read and comprehend pseudo-code (Obaidellah et al. (2020)).

Originally proposed by Agrawal et al. (1994), the Apriori algorithm identifies frequent itemsets through an iterative, level-wise search strategy, in which frequent n-itemsets are used to generate candidate $(n + 1)$-itemsets. This procedure is guided by the Apriori principle, which states that any superset of an infrequent pattern cannot itself be frequent, thereby substantially reducing the search space. When adapted to sequential data, Apriori-based methods enable the discovery of frequent ordered AoI subsequences. However, despite their conceptual simplicity and interpretability, these algorithms remain computationally demanding due to the need for repeated database scans and the combinatorial growth of candidate patterns.

An important improvement over classical Apriori-based approaches is the *Sequential Pattern Mining* (SPAM) algorithm, introduced by Ayres et al. (2002). SPAM adopts a depth-first search strategy combined with a vertical bitmap representation of the data, allowing for efficient support counting and substantial compression of the search space. This representation is particularly advantageous for mining long sequential patterns, as it enables compact storage and fast logical operations. In the context of eye-movement analysis, SPAM has been successfully applied to identify frequent AoI patterns across multiple scanpaths (Hejmady and Narayanan (2012)) and has also been used as a reference method for evaluating newly proposed pattern discovery techniques (Eraslan et al. (2016)).

Beyond Apriori-based and SPAM algorithms, other well-established mining techniques have also been employed to characterize AoI sequences. For instance, Geisler et al. (2020) proposed using the *MinHash* algorithm to efficiently assess similarity between AoI sequences. Their approach begins by extracting fixed-length subsequences from AoI sequences using a sliding-window procedure that allows for gaps between elements. The frequencies of these subsequences—conceptually similar to n-grams—are then used to construct frequency dictionaries for each sequence. To compare these dictionaries efficiently, the MinHash algorithm is applied to approximate the Jaccard similarity index between sequences. Although not a pattern mining method in the strict sense, this approach enables scalable and robust comparison of AoI sequences based on shared subsequence structure while maintaining computational efficiency.

In a broader methodological context, it is worth noting that sequential pattern mining occupies an intermediate position between simple transition-based analyses and more flexible pattern discovery frameworks. By allowing variable-length subsequences and temporal gaps, these methods overcome some of the limitations inherent to fixed-order Markov models and fixed-length n-gram representations. At the same time, their reliance on frequency thresholds and explicit pattern enumeration preserves interpretability, making them particularly attractive for exploratory analyses of gaze behavior. Moreover, n-gram analysis forms a foundational component of more complex comparison and characterization approaches, such as the *Subsmatch* algorithm—described in the *Scanpaths and Derived Representations* part of this review series—and more generally serves as a versatile building block across a wide range of algorithms in different domains. These characteristics naturally motivate the introduction of more expressive pattern detection techniques, discussed in the following sections, which further relax temporal constraints and aim to capture recurrent visual strategies in a more flexible manner.

4.4 T-Pattern Detection

T-pattern detection was originally developed by Magnusson (2000) to identify temporal and sequential structures in behavioral science, particularly in the context of social interactions. Since its inception, this approach has been applied across diverse fields (Mast and Burmester (2011)) and has been extensively utilized to uncover recurring patterns within viewer gaze dynamics in eye-tracking studies (Mast and Burmester (2011); Burmester and Mast (2010); Drusch and Bastien (2012)). In eye-tracking research, the primary objective of T-pattern analysis is to elucidate the temporal organization of visual behaviors that occur frequently and recurrently, thereby revealing structured attentional routines that are not apparent from fixation statistics alone.

As originally defined by Magnusson (2000), a T-pattern is a sequence of events that satisfies two essential criteria: (i) the sequence occurs in the data more frequently than would be expected by chance, and (ii) there exists a consistent temporal relationship between the events. Importantly, T-patterns do not require deterministic timing; that is, the inter-event intervals may vary across occurrences and are treated as stochastic rather than fixed quantities. The detection process therefore focuses on identifying a *critical* *interval*, which defines the characteristic temporal window within which successive events of the pattern are expected to occur and distinguishes meaningful temporal structure from random co-occurrence.

The algorithm for T-pattern detection operates under the null hypothesis that the events being analyzed are independent Bernoulli processes with constant probabilities of occurrence per time unit over the observation period. Based on this assumption, the algorithm computes the probability that a second event occurs within a specified time interval following a first event (Qazi et al. (2019)). Through an iterative and hierarchical process, detected T-patterns are progressively combined with additional events or with previously identified T-patterns, enabling the discovery of increasingly complex temporal structures. This procedure continues until no further statistically significant T-patterns can be identified.

It is important to note that many of the algorithms discussed in the preceding sections primarily focus on the sequential order of events, often disregarding the temporal intervals that separate them. Moreover, a large class of classical data-mining approaches implicitly assumes constant or discretized time steps between events. By explicitly incorporating temporal constraints, T-pattern detection introduces an additional layer of complexity, both conceptually and computationally. As a result, this approach is particularly well suited for AoI sequences of moderate size, for which variations in inter-event timing can be modeled without incurring prohibitive computational costs.

Additionally, the detection of T-patterns is highly sensitive to the choice of the significance level parameter, which directly influences the estimation of critical intervals as well as the number and length of the patterns discovered. A stringent significance threshold typically yields fewer and shorter T-patterns, whereas a more permissive threshold facilitates the detection of longer and more numerous patterns. T-patterns can also be filtered according to various additional criteria, such as minimum pattern length or minimum number of occurrences. Consequently, the results of T-pattern analysis are strongly dependent on parameter settings, which may hinder reproducibility and complicate comparisons across studies if not carefully reported.

Finally, it should be mentioned that while Magnusson's original T-pattern detection algorithm is not freely available, a number of subsequent studies have provided detailed descriptions of the underlying computational principles, along with several methodological improvements and extensions (Tavenard et al. (2008); Salah et al. (2010); Arora et al. (2017)). These contributions have clarified the statistical foundations of T-pattern detection, proposed alternative search strategies, and introduced several refinements aimed at improving robustness, computational efficiency, and interpretability. In doing so, they have broadened the applicability of T-pattern analysis beyond its original scope and facilitated its integration into contemporary analyses of gaze behavior and other forms of temporally structured behavioral data. Moreover, by making the core ideas and algorithms more accessible, this body of work has enabled researchers to adapt T-pattern detection to diverse experimental contexts and to combine it more readily with other sequence analysis frameworks.

5 COMMON SUBSEQUENCE IDENTIFICATION AND CLUSTERING

As outlined in the previous sections, a wide range of methods have been proposed to identify patterns within observer AoI sequences, either by detecting salient sub-sequences within a single AoI sequence or by extracting recurrent patterns across a set of sequences. In these approaches, the objective is primarily to characterize local or global regularities in gaze behavior without necessarily reducing the data to a single representative trajectory.

Conversely, an alternative family of techniques aims to summarize a collection of AoI sequences by identifying a single scanpath that is representative of the entire set, commonly referred to as a *representative* *sequence* or *consensus sequence*. When AoI sequences are used as abstractions of scanpaths, such consensus representations provide a compact description of the visual strategies shared among a group of observers, facilitating both interpretation and comparison across conditions or populations.

In this section, we introduce several algorithms that explicitly combine a common AoI subsequence identification step with clustering mechanisms to organize multiple AoI sequences according to shared visual patterns. By jointly addressing the problems of sequence comparison, grouping, and representation, these methods enable the identification of coherent subgroups of observers and the extraction of representative scanpaths for each group. As such, they constitute a natural continuation of the pattern mining approaches described earlier, while shifting the focus from local pattern discovery to global sequence-level synthesis.

5.1 Local Alignment Algorithms

Local alignment of sequences is a widely adopted technique, especially in bioinformatics, for uncovering shared patterns within data sequences. In this section, the term *local alignment* is used in a broad sense to refer to methods aimed at identifying common subsequences, independently of their absolute position within the full sequences. The identification of the *longest common subsequence* (LCS) plays a critical role in disciplines that analyze related or derived sequences. The goal of local alignment is distinct from that of global alignment methods, which are comprehensively explored in the *Scanpaths and Derived Representations* section of this review series (Laborde et al. (2024c)) in the context of scanpath comparisons. Specifically, local alignment emphasizes identifying highly similar subsequences by aligning only the segments that maximize the similarity score within two larger sequences. This approach enables the detection of analogous sub-patterns, even when their positions vary across the sequences under comparison. As a result, local alignment is particularly advantageous for isolating highly similar areas-of-interest subsequences from AoI sequence representations of scanpaths, facilitating deeper insights into shared patterns of visual exploration without necessitating global sequence alignment.

A common approach employed to solve the LCS problem between two string sequences leverages dynamic programming (Bergroth et al. (2000)), a method for addressing optimization problems by breaking them into smaller sub-problems and solving them iteratively in a *bottom-up* manner. Specifically, the algorithm constructs a comparison matrix — or table — where one sequence defines the column indices and the other defines the row indices. Each cell in the matrix represents the LCS at that stage, and the matrix is filled iteratively following this rule: if the AoIs corresponding to the current row and column match, the current cell is assigned a value equal to one plus the diagonal element; otherwise, the cell is assigned the maximum value of the previous row or column. The LCS itself can be reconstructed by backtracking from the matrix's final cell to the origin, following the recorded optimal paths. Meanwhile, the length of the LCS is directly obtained from the value of the last cell in the matrix.

Alternatively, the *Smith–Waterman algorithm* (Smith et al. (1981)), a highly regarded local alignment method extensively used in bioinformatics for DNA and protein sequence analysis (Setubal et al. (1997)), has been proposed as a robust approach for identifying similar patterns between pairs of observer AoI sequences (West et al. (2006)). This algorithm is particularly well-suited for analyzing localized correspondences between areas of interest, as it excels in identifying optimally aligned subsequences within larger sequences. Similar to the LCS approach, the Smith–Waterman algorithm utilizes a dynamic programming framework for pairwise sequence comparison; however, it introduces several critical enhancements, foremost of which is the incorporation of a substitution matrix that assigns similarity scores to pairs of AoIs. Matches typically receive positive scores, while mismatches are assigned lower or negative scores, allowing for a flexible similarity metric. This metric can be adapted to specific applications, for example, by reflecting spatial distances between the geometric centers of AoIs or, in more complex cases, semantic relationships between their contents. Another key feature is the algorithm's ability to handle gaps in alignment through a gap penalty function. This function assigns costs for opening and extending gaps, balancing the trade-off between alignment flexibility and sequence integrity. The parameters of the gap penalty function can be customized to suit specific contexts, enabling the algorithm to manage varying levels of sequence fragmentation or noise effectively.

Locally aligning sequences enables the evaluation of pattern similarity between pairs of AoI sequences, a capability particularly valuable in domains such as clinical reasoning. For example, locally aligned AoI sequences can be used to assess expertise by identifying overlap regions between the AoI sequence of a learner and a reference or optimal sequence (Jraidi et al. (2022)). This approach highlights areas where the learner's focus aligns with the reference, providing insights into their cognitive and visual processing strategies. Beyond pairwise comparisons, some studies have extended local alignment–based analyses to groups of observers in order to detect and classify distinctive visual strategies. For instance, Castner et al. (2020b) proposed a method to distinguish expert and novice dentists in their interpretation of dental radiographs. This approach uses hierarchical clustering based on a distance matrix derived from Smith–Waterman pairwise distances. However, this method does not apply local alignment directly to raw AoI sequences; instead, it analyzes features extracted from observers' scanpath trajectories. This adaptation allows the analysis to focus on high-level features of the visual exploration process, such as fixations and transitions, rather than the raw AoI sequence data, providing a more abstracted but still meaningful representation of visual attention.

A related approach is the *eMINE* algorithm introduced by Eraslan et al. (2014), which explicitly bridges pairwise sequence comparison and multi-sequence synthesis. This method combines both global and local alignment techniques to generate a consensus or abstracted string representation from a set of AoI sequences. The process begins by applying the Levenshtein distance, a global sequence similarity measure, to identify the two most similar sequences within the set (Laborde et al. (2024c)). Next, the LCS technique is applied to these sequences to extract their common subsequence, representing the shared visual attention patterns between them. Once the common subsequence is identified, the two original sequences are replaced by this shared subsequence within the set, and the process repeats iteratively until a single consensus AoI sequence remains. This iterative procedure facilitates hierarchical clustering, uncovering common visual strategies among observers and providing insights into shared patterns of visual attention.

While the eMINE algorithm provides valuable insights into overarching visual patterns, it also has certain limitations. Specifically, in scenarios involving numerous observers exhibiting slightly divergent visual dynamics, the intermediate steps of the process may lead to the loss of critical visual elements. The creation of short common subsequences at these stages may fail to capture the subtleties of individual observers' strategies, particularly when dealing with heterogeneous or highly variable AoI data. Although hierarchical clustering in eMINE offers the advantage of grouping similar visual strategies, its effectiveness is contingent on the degree of similarity between the sequences being analyzed. For groups of observers with considerable variability, this method may require further refinement or the incorporation of additional metrics to preserve the richness of the original sequences and prevent excessive abstraction during intermediate stages.

5.2 Trend Analysis

As previously discussed, common sequences identified through local alignment-based techniques often suffer from reductionism, resulting in scanpaths that are too short for meaningful analysis or further processing. This limitation arises because such algorithms typically retain only those visual AoIs that are shared across all individual AoI sequences, thereby discarding visual elements that appear in a large subset of observers but not universally. As a consequence, small yet behaviorally meaningful individual deviations are frequently overlooked. To address this limitation, Eraslan et al. (2016) introduced the *scanpath trend analysis* (STA) approach, which provides a more inclusive clustering strategy. STA not only identifies visual AoIs visited by all observers but also accounts for AoIs visited by the majority, independently of the order in which they are observed.

It is important to note that the term *scanpath* is somewhat misleading in this context. Indeed, the initial step of the STA algorithm consists in transforming each scanpath into a sequence of visual elements, making STA particularly relevant for the analysis of AoI sequences. Accordingly, the algorithm first represents individual inputs as sequences of visual elements, which effectively correspond to AoI sequences augmented with fixation durations associated with each visual region.

The STA algorithm operates in two main stages. In the first stage, successive fixations on the same visual element are concatenated, and each visit to a given visual element is defined as a *visual instance*. Each instance is characterized by the identity of the visited visual element and a value reflecting the duration of the visit relative to other instances associated with the same element. The algorithm then analyzes these visual instances to identify *trending* visual elements. If a visual instance is shared across all individual sequences—optionally allowing for a tolerance parameter (Eraslan et al. (2017))—it is considered trending. Visual instances that are not fully shared across all sequences are subsequently evaluated based on the amount of attention they receive, quantified in terms of fixation frequency and associated dwell time. Notably, STA does not rely on predefined thresholds. Instead, these criteria are derived directly from the data: a visual instance is considered trending if (i) its total number of fixations is greater than or equal to the minimum fixation count observed for fully shared instances, and (ii) its dwell time is greater than or equal to the minimum dwell time associated with fully shared visual instances.

In the second stage, the STA algorithm constructs a *trending sequence* from the set of identified trending visual instances. To determine the position of each instance within the final sequence, STA assigns a priority value reflecting its relative position within the individual input sequences. These priority values are then aggregated across all observers, and the resulting score is used to determine the ordering of visual elements in the consensus sequence. A visual instance that is not shared by all individual sequences—and whose aggregated priority value falls below the minimum priority of fully shared instances—is excluded from the final trending sequence.

Conceptually, STA occupies an intermediate position between strict consensus-sequence extraction and frequency-based pattern mining approaches, offering a compromise between robustness to inter-individual variability and the preservation of an interpretable ordered representation of gaze behavior. Originally developed and evaluated for the analysis of visual behavior on web pages, the STA algorithm has since been applied in a variety of contexts. For instance, it has been used in programming-related tasks to identify common code-reading strategies, with the aim of informing the design of pedagogical interventions and improving program comprehension skills (Ahsan and Obaidellah (2023); Tablatin and Rodrigo (2018)). STA has also been employed to investigate differences in AoI sequences between autistic and neurotypical individuals, providing insights into behavioral and cognitive variations in visual attention patterns (Eraslan et al. (2020)). These diverse applications highlight the versatility of STA as a method for capturing shared visual strategies while accommodating individual variability.

5.3 Candidate-Constrained Dynamic Time Warping Barycenter Averaging

Recently, Li and colleagues introduced the *candidate-constrained dynamic time warping barycenter averaging* (CDBA) method, an advanced algorithm for scanpath aggregation (Li et al. (2017); Li and Chen (2018)). The core idea behind CDBA is to aggregate multiple AoI sequences into a single, representative sequence by computing the barycenter of the experimentally recorded sequences. In this context, the barycenter can be understood as a sequence that is, on average, optimally aligned with all input sequences. More specifically, the algorithm seeks to determine a representative AoI sequence that minimizes the mean dynamic time warping distance to each individual sequence, thereby capturing a consensus sequence that is centrally positioned within the set of observed gaze behaviors. This formulation explicitly integrates both spatial proximity and temporal ordering into the construction of the representative scanpath.

The CDBA algorithm builds upon the *dynamic time warping barycenter averaging* (DTWBA) framework originally proposed by Petitjean et al. (2011) and operates through two main iterative steps. In the first step, the algorithm computes the *dynamic time warping* (DTW) distance between each individual AoI sequence and an initially defined reference AoI sequence, which is typically randomly initialized — for a detailed presentation of DTW and its properties, the reader is referred to the *Scanpaths and Derived Representations* part of this review series (Laborde et al. (2024c)). This alignment step establishes an optimal temporal correspondence between each input sequence and the current reference sequence, allowing for non-linear temporal deformations.

The second step, referred to as the *update step*, consists in refining the reference AoI sequence based on the alignments obtained in the first step. Each element of the reference sequence is updated using a constrained barycenter computed from the AoIs aligned to that position across all sequences. This constrained barycenter is defined as the AoI that minimizes the average distance to the aligned visual elements while satisfying two key constraints: (i) any two contiguous AoI components in the reference sequence must also appear contiguously in at least one of the individual input sequences, and (ii) the occurrence count of any AoI component in the reference sequence must not exceed its maximum occurrence count across the set of input sequences. These constraints ensure that the resulting consensus sequence remains consistent with the structural ordering and frequency characteristics of the original data.

The two steps are iteratively repeated until convergence, that is, until the reference AoI sequence remains unchanged across successive iterations, yielding a final representative AoI sequence. Notably, in their original implementation, Li et al. (2017) performed the DTW alignment at the level of individual fixations rather than directly on AoI centers. In this formulation, DTW is computed between fixation sequences and the reference AoI sequence, and the aggregation process is subsequently expressed at a higher level of abstraction during the update step. Specifically, each component of the reference scanpath is updated with the AoI that minimizes the average distance to all fixations aligned to that position. This hybrid formulation allows the algorithm to retain fine-grained spatial information during alignment while producing a symbolic AoI-based representation as the final output.

However, the constant switching between two levels of abstraction — fixations and AoIs — can be somewhat confusing, as it blends both detailed fixation-level information with higher-level AoI-based representations. Moreover, there are two important factors to consider when initializing the iterative refinement process: (i) the length of the initial reference sequence, and (ii) the specific elements that comprise the sequence (Petitjean et al. (2011)). Although these factors have been recognized in the DTW Barycenter Averaging literature, they have not been explicitly addressed by Li et al. (2017) in the context of the CDBA algorithm. Specifically, the length of the starting reference AoI sequence, when the input sequences vary in length, could significantly affect the process of generating a consensus sequence. Inadequate consideration of these factors may lead to suboptimal convergence or misalignment, impacting the accuracy and interpretability of the resulting aggregated scanpath. Compared to frequency-based or alignment-based consensus methods such as STA or eMINE, CDBA explicitly optimizes a geometric notion of centrality under temporal alignment constraints, making it particularly suitable when spatial information and sequence timing are critical.

5.4 Dotplot Clustering

Dotplots are a sequence comparison technique used to measure the similarity between two or more symbol sequences by highlighting common subsequences and shared symbolic patterns. This method provides a graphical representation for sequence comparison, offering an intuitive visualization of matching subsequences. It allows the comparison of sequences of different lengths, naturally accounting for gaps, while preserving the temporal aspect of the AoI sequence. Dot plots have been applied in various fields of string analysis, such as plagiarism detection in literature and informatics (Church and Helfman (1993)), and the study of biological sequences, also known as the diagram method (Gibbs and McIntyre (1970)), the enhanced graphic matrix procedure (Maizel Jr and Lenk (1981)), or dot-matrix plots (Huang and Zhang (2004)).

In brief, a dotplot can be considered a discrete variation of a cross recurrence plot — for a detailed discussion, refer to the *Scanpaths and Derived Representations* part of this review series (Laborde et al. (2024c)). The dotplot matrix visualizes the temporal coupling between two AoI sequences, with the vertical axis representing AoIs from the first sequence and the horizontal axis representing AoIs from the second sequence. A dot is placed when fixations from both sequences coincide within the same AoI, and the position is left empty otherwise.

Once the dotplot is computed, representative scanning strategies can be derived. These are reflected by diagonally aligned sequences of dots, representing common sub-sequences or recurring patterns of AoI sub-sequences across both sequences. Note that, while recent studies have proposed using *cross-recurrence quantification analysis* between AoI centers to quantify visual attention and workload variations (Atweh et al. (2023)), the dotplot approach leverages the semantic abstraction provided by AoIs more effectively. It focuses not on the geometric properties of the visualized regions, but on their categorical membership within the same AoI.

Additionally, in the context of eye movement analysis and AoI pattern assessment, some researchers have suggested accounting for the differential frequency of AoI occurrences — as some AoIs are more frequent than others — by weighting each dot (Goldberg and Helfman (2009)). Specifically, a weighting function introduced by Church and Helfman (1993) assigns to each dot the inverse of the frequency of that AoI in the concatenated sequence. This allows each dot to be classified as a significant match or not, based on whether its weight exceeds a predefined frequency threshold. This threshold may be manually set (Goldberg and Helfman (2010)) or computed experimentally, with significant matches corresponding to weights greater than one standard deviation from the mean.

Beyond simple qualitative analysis, Goldberg and Helfman (2009) proposed a two-step linear regression process to extract statistically significant patterns from dot plot pairs derived from a set of AoI sequences. In the first step, a linear fit is applied to the significant dots. If the coefficient of determination of the fit falls below a predetermined threshold, it is concluded that there is no common scanning strategy between the two AoI sequences, and another pair is considered. In a second step, if the regression line is deemed acceptable, a second linear fit is performed on the points within a small distance from the regression line.

The final common pattern in the two AoI sequences is defined as the sequence of dots closest to this new fit. The matching distance is quantified as the number of sequentially matching AoIs between the two sequences. This two-step procedure can be iteratively applied to each pairwise comparison, allowing for the identification of additional patterns in the input sequences

Subsequently, clustering algorithms, such as hierarchical clustering, have been proposed (Goldberg and Helfman (2010)) to identify strategies and sets of scanned regions that match progressively larger numbers of AoI sequences and individuals. To hierarchically cluster multiple AoI sequences, the two most similar sequences are first selected from the input set based on the dotplot matching distance defined above. These sequences are then replaced by their common regression sequence. This process is iteratively repeated until only a single sequence remains, which represents the common, or dominant (Albanesi et al. (2011)), AoI sequence.

It is important to note that hierarchical clustering can be used in conjunction with other comparative methods that enable the computation of distances between pairs of AoIs. As such West et al. (2006) employed hierarchical clustering techniques to directly analyze visual sequences using distance matrices. These distance matrices were computed using the Levenshtein distance or the Needleman-Wunsch distance between pairs of visual sequences. A detailed discussion of these distance measures can be found in the *Scanpath and Derived Representations* part of this review series (Laborde et al. (2024c)). This leads to several important observations. First, many of the techniques discussed in this work can be combined to refine our understanding of visual dynamics. Second, certain methods presented in the *Scanpaths Analysis* section can be adapted for the analysis of AoI sequences. In particular, methods derived from string edit distances, which take sequences of symbols as input, can be effectively applied in this context

Finally, dotplot-based approaches illustrate how symbolic sequence comparison, statistical filtering, and hierarchical clustering can be combined to uncover dominant visual strategies, reinforcing the idea that meaningful consensus scanpaths emerge from the interplay between local pattern alignment and global sequence organization.

6 VISUALIZATION METHODS

Quantitative methods have gained prominence in recent decades for their systematic approaches to identifying distinct gaze behaviors across populations or in response to various stimuli. However, visualization techniques offer a complementary, exploratory, and qualitative framework for analyzing gaze data. Beyond their descriptive role, such techniques are particularly valuable for examining the spatio-temporal organization of gaze behavior, revealing complex relationships that may remain difficult to capture through numerical summaries alone. By providing an intuitive understanding of gaze dynamics, visualizations facilitate hypothesis generation, interpretation, and communication of results, which can subsequently be tested or refined using quantitative methods.

While several reviews have addressed general approaches for visualizing eye-tracking data Blascheck et al. (2014, 2017); Claus et al. (2023), this section specifically focuses on techniques designed to visualize AoI sequences. These methods emphasize the dynamics of transitions between areas of interest and the temporal structure of gaze behavior, often building directly on the analytical frameworks introduced in the previous sections. By visually depicting how gaze progresses across AoIs and how these regions are interconnected over time, such visualizations provide an interpretable bridge between symbolic sequence representations and the underlying cognitive and perceptual processes driving visual attention.

6.1 Spatio-Temporal Based

Beyond their illustrative value, spatio-temporal visualizations of AoI sequences are often used in conjunction with quantitative models of gaze behavior. In particular, they provide an intuitive means of inspecting, validating, and interpreting temporal structures revealed by transition-based, entropy-based, or pattern-mining analyses. Among the earliest spatio-temporal visualizations of gaze data, the *time plot* Itoh et al. (2000); Räihä et al. (2005) represents discrete areas of interest along the vertical axis and time along the horizontal axis.

Alternatively, *gaze duration sequence diagrams* depict fixation durations within AoIs as continuous segments along the vertical axis, with horizontal transitions indicating shifts of attention between regions Raschke et al. (2012). Variants of these approaches appear frequently in the literature under different names, yet they share a common structure: AoIs are arranged along parallel timelines on one axis, while the other axis encodes temporal progression Crowe and Narayanan (2000); Holmqvist et al. (2011); Weibel et al. (2012). Figure 4 provides a representative illustration of these methods. These visualizations can be applied to individual participants, aggregated groups, or multiple observers displayed in parallel—often using color coding to differentiate them. However, when many participants are overlaid, the resulting visual clutter can substantially reduce interpretability.

To enable comparison between observers, the *scarf diagram* provides a simple yet effective method for spatio-temporal AoI visualization (Burch et al. (2021); Bakardzhiev et al. (2021); Richardson and Dale (2005); Kurzhals et al. (2017)) — see Figure 5. This technique represents areas of interest along the vertical axis, with time progressing on the horizontal axis. Each AoI is assigned a distinct color, and individual time bars are plotted for each observer. These bars are color-coded to indicate the AoI being observed at any given moment, enabling an intuitive depiction of gaze transitions over time.

Although scarf diagrams were not explicitly designed to visualize recurrent sequences, they facilitate the comparison of eye movements across multiple participants and support the identification of simple temporal patterns, such as sustained attention on a specific AoI or simultaneous focus on similar regions across observers (Claus et al. (2023)). However, their effectiveness decreases when applied to more complex datasets. In particular, their ability to convey long or intricate gaze patterns is limited when sequences span extended time periods or involve multiple overlapping transitions. As the number of AoIs increases, scarf diagrams also become prone to visual clutter, which can substantially reduce interpretability and hinder the extraction of meaningful insights.

To address these limitations, Yang and Wacharamanotham (2018) extended scarf plots by introducing Alpscarf, a compact visualization technique designed to facilitate the identification of visual patterns. Alpscarf incorporates three main extensions: (i) perceptually optimized color palette heuristics, (ii) three additional visual components — mountains, valleys, and creeks — and (iii) two alternative modes of scarf width, namely transition-focus and duration-focus. First, the color heuristics (Bianco et al. (2015)) assign each AoI group—implicitly reflecting a predefined AoI hierarchy—a specific color hue, thereby minimizing perceptual similarity between adjacent groups and improving visual discrimination. Secondly, Alpscarf introduces three visual components surrounding each scarf. The height of the mountains reflects AoI visits that follow a user-defined expected order (e.g., sequential paragraph reading resulting in prominent mountain structures). The depth of the valleys indicates instances of revisits, with frequent revisits appearing as wider or deeper valleys. Both mountains and valleys leverage the position channel to encode transition patterns, while the gaps between the scarf and these components—referred to as creeks—visually emphasize deviations from expected viewing behavior.

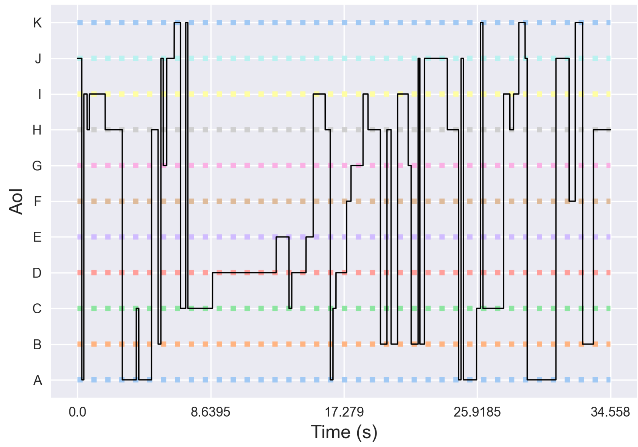

Figure 4. Time plot. A time plot is presented: each colored horizontal line represents an AoI, while the black line traces an individual's gaze over time — along the horizontal axis — as it transitions from one AoI to another. Despite its simplicity, this method is well-suited for examining the gaze dynamics of a single viewer between AoIs over time.

Finally, Alpscarf offers two modes of scarf width: duration-focus, in which bar width corresponds to the logarithmic fixation duration, and transition-focus, in which all bars are displayed with equal width to emphasize transitions rather than dwell times. More generally, scarf-based representations, and Alpscarf in particular, are frequently used as exploratory complements to sequence analysis methods. They enable researchers to visually assess group-level regularities, temporal alignment, or deviations in gaze behavior, which can subsequently be formalized and quantified using Markov models, entropy-based metrics, or common subsequence analyses.

Another visualization technique closely related to scarf plots is the *gaze stripe* (Kurzhals et al. (2015)). Similar to scarf plots, this method maps time along the horizontal axis and participants along the vertical axis. However, instead of relying solely on color coding, a thumbnail image capturing the local visual context of the stimulus is extracted for each gaze data point and aligned along the timeline. To facilitate the identification of patterns in gaze sequences based on their similarity, Kurzhals et al. (2015) introduced a hierarchical clustering method. In this approach, gaze sequences are represented as leaf nodes, and a cluster hierarchy is constructed using an agglomerative hierarchical clustering algorithm (Hastie et al. (2009)). Similarity between sequences is quantified using a distance metric based on the correlation of hue and saturation histograms derived from the thumbnails. The resulting dendrogram provides a powerful exploratory tool for analyzing sequence similarities and uncovering underlying structural patterns in gaze behavior.

The use of *3D graphics* has also been proposed to simultaneously represent the horizontal and vertical components of the visual field alongside the temporal dimension. However, the introduction of a third dimension requires careful consideration. Perspective effects can complicate the estimation of element sizes, while occlusion caused by overlapping elements may lead to information loss. In the specific context of eye movement analysis, the space-time cube (Kurzhals and Weiskopf (2013)) has been introduced as a promising visualization technique. In this representation, the horizontal and vertical dimensions describe spatial gaze distribution, while time is mapped to the Z-axis. The visual stimulus is displayed at the base of the cube, with gaze trajectories emerging from it and evolving as viewing time progresses. To enhance

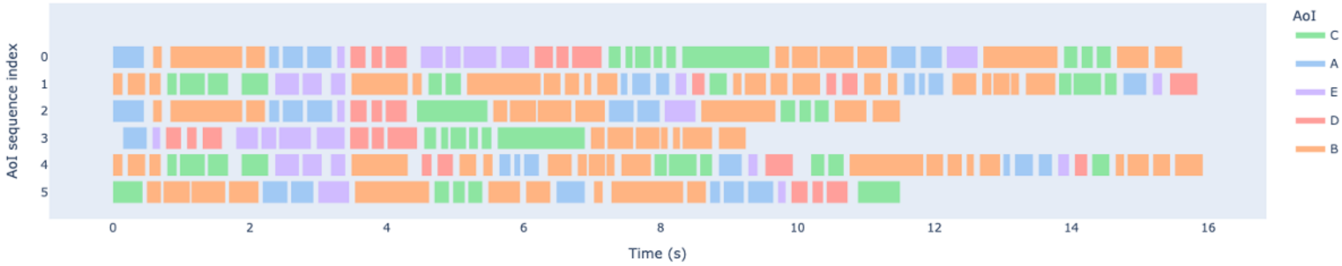

Figure 5. Scarf plot. An illustration of a scarfplot is presented, with six individuals — or six successive recordings from a single individual — displayed along the vertical axis. Time is represented on the horizontal axis, while transitions between AoIs for each individual are shown through changes in color, with each AoI assigned a specific color.

interpretability, segments of gaze traces corresponding to specific areas of interest can be highlighted using AoI-consistent color coding. This approach further supports interactivity and dynamism, allowing analysts to rotate the cube or animate it along the temporal axis (Kurzhals and Weiskopf (2013)). Despite their visual appeal, however, such 3D representations share the main limitations of scarf-based approaches: they provide limited support for exploring complex temporal models and become less practical when dealing with a large number of input–output relationships.

We conclude this section by introducing *AoI rivers* (Burch et al. (2013)), an interactive visualization method designed to explore time-varying fixation frequencies alongside transitions between areas of interest in data collected from a large number of participants. This method employs a *Sankey diagram* (Riehmann et al. (2005)), which provides an overview of frequently visited AoIs over time while highlighting transitions between them. Each AoI is visualized as a color-coded river, with its thickness at any given time corresponding to the frequency of visits to that AoI, normalized relative to the maximum AoI frequency sum across all time units. The river flows from left to right, dynamically changing width to reflect temporal variations in attention. More specifically, transitions between AoIs are depicted as color-coded sub-rivers connecting the main AoI rivers, representing transitions between areas over time. Note that this representation also accounts for gaze points that are not directly inside an AoI but leave or enter one at a later time, which are displayed as *effluents* and *influents*. An example Sankey diagram is provided in Figure 6 to illustrate this approach. While AoI rivers provide an effective overview of time-varying attention frequencies, they are less suited for revealing hierarchical or long-range sequential structures in gaze behavior. As noted by Burch and colleagues (Burch et al. (2013)), these representations are most effective when applied to datasets with a limited number of areas of interest. Consequently, AoI rivers are best viewed as a complementary tool that emphasizes aggregate temporal trends rather than detailed sequential dependencies, motivating the use of transition- and pattern-based visualizations introduced in the following subsection.

More generally, spatio-temporal AoI visualizations serve as a conceptual bridge between raw gaze recordings and abstract sequence-based representations. By explicitly encoding both the temporal evolution of attention and the transitions between areas of interest, these visualizations provide an interpretable intermediate layer that facilitates hypothesis generation and exploratory analysis. In practice, they enable researchers to visually inspect recurrent patterns, temporal alignments, or deviations across observers, which can subsequently be formalized and quantified using the computational models introduced in the

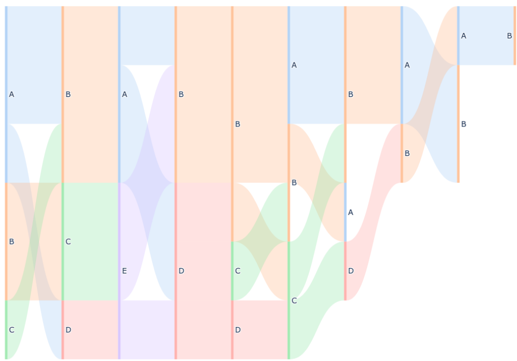

Figure 6. Sankey diagram. An illustration of a Sankey diagram is provided, where each AoI is represented by a distinct color-coded stream. The width of each stream reflects the frequency of visits to the corresponding AoI at any given time, with thicker sections indicating greater attention. The streams flow from left to right, dynamically adjusting in width to depict temporal variations in attention distribution. Transitions between AoIs are represented by interconnecting, color-coded sub-streams that bridge the primary streams, effectively visualizing the movement of attention between different areas over time.

preceding sections. As such, spatio-temporal visualizations play a critical role in validating, contextualizing, and complementing quantitative analyses, ensuring that higher-level sequence models remain grounded in the underlying gaze dynamics.

6.2 Transition Based

Visualizing the transitions of an observer, or, where applicable, a group of observers, presents a significant challenge due to issues such as visual clutter and the loss of explicit temporal information, which may obscure the underlying visual strategies employed by the observer. As a result, various visualization techniques have been proposed to improve the representation and interpretation of transitions between areas of interest (AoIs).

One of the simplest approaches consists in visualizing a previously computed AoI transition matrix (Goldberg and Kotval (1999); Krejtz et al. (2013)), as described in Section 3.1. In a transition matrix, AoIs are ordered along the horizontal axis — rows — and the vertical axis — columns — and each cell contains the count or proportion of transitions between the corresponding AoIs. A color scale is commonly applied to encode the frequency or intensity of gaze transitions, yielding a compact visual summary of pairwise transition probabilities — as illustrated in Figure 3a. This representation provides a direct visual counterpart to a Markov-based transition model and retains all pairwise transition statistics. However, despite this quantitative completeness, transition matrices remain limited from a perceptual and exploratory standpoint. In particular, they do not explicitly convey temporal ordering or higher-order sequential structure, and are often considered less expressive than graph-based visualizations, such as directed graphs, which offer a more intuitive depiction of gaze dynamics between AoIs.

A *directed graph* is a visualization that depicts transitions between areas of interest, with each node representing an AoI and each directed edge illustrating a transition between two AoIs. Although these diagrams are most commonly used to represent the eye movement dynamics of a single participant, they can also be constructed from aggregated or averaged data across multiple observers. In addition, nodes can be resized or color-coded to encode auxiliary metrics, such as fixation duration or visit frequency, while the thickness of the directed edges typically reflects the frequency or probability of transitions between the corresponding AoIs. While many visualization techniques present the graph independently of the stimulus (Itoh et al. (1998); Fitts et al. (2005); Tory et al. (2005)), *in-context* directed graphs embed the graph directly within the visual stimulus (Holmqvist et al. (2003); Hooge and Camps (2013); Opach et al. (2014)) — as illustrated in Figure 3b. This integration enhances interpretability by enabling a direct correspondence between AoI transitions and stimulus content (Blascheck et al. (2017)), thereby facilitating a more intuitive understanding of gaze behavior.

However, directed graphs also exhibit several inherent limitations. By construction, they are restricted to the representation of sequential pairs of fixations, *i.e.* first-order transitions between AoIs, which limits their ability to convey longer and more complex gaze patterns. Moreover, as the number of areas of interest increases, directed graphs tend to suffer from visual clutter, making large-scale transition structures increasingly difficult to interpret. Although a variety of graph visualization techniques have been proposed to mitigate these issues (Von Landesberger et al. (2011)), scalability remains a central challenge, motivating the development of alternative transition-based visualizations that better capture higher-order structure and global gaze dynamics.

An interesting alternative to directed graphs is the use of *chord diagrams*, a class of visualization techniques designed to represent pairwise relationships between nodes in a compact and symmetric manner. Their simplicity and intuitive radial layout — in which all nodes can be directly connected to one another — has led to their widespread adoption in affiliation and relational visualizations Rees et al. (2020). In the context of gaze analysis, chord diagrams have recently been proposed as effective tools for highlighting the global structure of eye movement transitions (Connor et al. (2020); Castner et al. (2020a); Wang (2021); Claus et al. (2023)). In a typical chord diagram, each AoI is represented as a segment along the outer circumference of a circular layout. Arcs, or chords, are drawn between AoIs, with the width of each arc proportional to the frequency of transitions between the corresponding visual elements. By construction, these arcs are symmetrical and therefore do not encode directionality, instead capturing the overall strength of association between pairs of areas of interest (Blascheck et al. (2013)).

To provide more specific information about gaze dynamics, color-coding schemes can be incorporated (Connor et al. (2020)) to encode the direction of transitions, for instance by indicating the AoI from which a transition originates. While this refinement enhances interpretability, it may also increase visual clutter, potentially reducing the clarity of the visualization when many AoIs or transitions are present. An illustration of this approach is shown in Figure 7. Although chord diagrams are most often used to visualize the gaze dynamics of individual observers, they have also been employed for group-level comparisons. For example, Connor et al. (2020) designed chord diagrams in which arc widths represented differences in transition frequencies between populations with varying levels of expertise, thereby emphasizing contrasts in visual strategies. Taken together, matrix-, graph-, and chord-based representations offer complementary perspectives on gaze transitions, trading temporal specificity and directionality for structural readability and global relational insight.

To refine transition visualization, Blascheck et al. (2016) proposed to introduce *hierarchical structure* between AoIs. Specifically, they distinguish between two types of hierarchy: (i) a *spatial* hierarchy with stimuli presenting nested AoIs, and (ii) a *semantic* hierarchy between AoIs. Once a hierarchy is established between areas of interest, Blascheck and colleagues proposed two complementary visual refinements: a hierarchical transition matrix and a graph-based representation grounded in an AoI tree structure. The proposed transition matrix visualization enhances a standard transition matrix by incorporating a hierarchical structure for the Areas of Interest, with the hierarchy displayed as a dendrogram positioned along the top and left sides of the matrix. Each AoI is represented both in the rows and columns of the

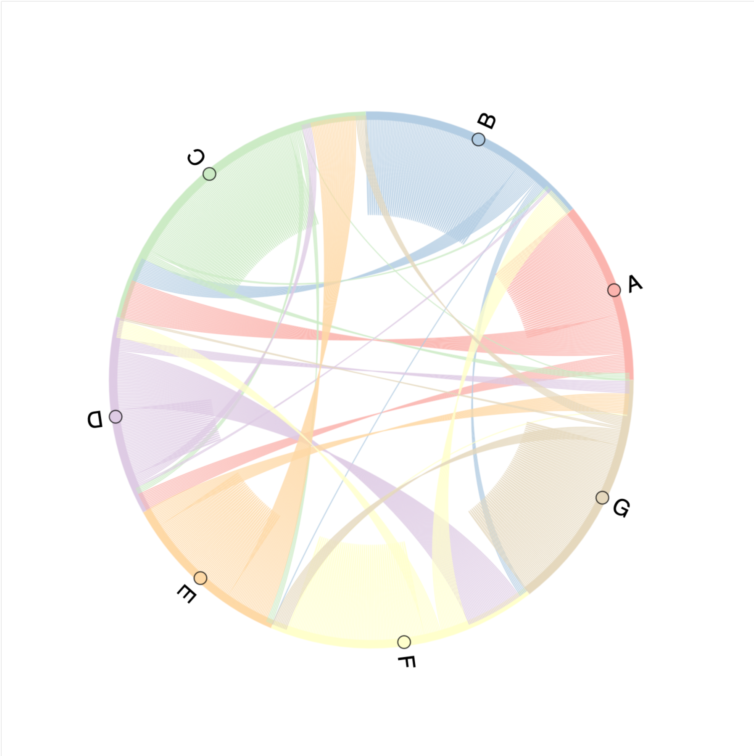

Figure 7. Chord diagram. An illustration of a chord diagram is presented, where each AoI is represented as a segment along the circumference of a circular layout. Arcs, or chords, link the AoIs, with the width of each arc proportional to the frequency of transitions between the respective visual elements. A color-coding scheme is employed to denote the direction of transitions, with arcs taking on the color of their source AoI.

matrix, and a color-coding scheme is applied to reflect the relative importance of transitions associated with each AoI. Specifically, the color of the row headers indicates the total number of outgoing transitions from a given AoI, while the color of the column headers represents the total number of incoming transitions to that AoI.

In addition to the matrix visualization, the AoI hierarchy can be further explored using an AoI tree representation. This tree structure utilizes indented pixel tree plots Burch et al. (2010), where each hierarchical element is represented by a rectangle, and the indentation of these rectangles corresponds to their hierarchical level. The root class and intermediate hierarchical classes are visually distinguished by vertically expanded rectangles, while the leaf nodes, which correspond to individual AoIs, are represented by smaller squares. Subsequently, transitions between AoIs are visualized in a manner similar to directed graphs, with arcs connecting the leaf nodes corresponding to individual AoIs. The width of each arc encodes the frequency of transitions, providing an intuitive visual cue for the intensity of gaze activity between regions.

The methods discussed thus far primarily focus on analyzing sequential pairs of fixations, *i.e.* transitions between two areas of interest. While informative, such pairwise representations provide only a limited view of gaze behavior, as they fail to capture the broader temporal organization and global structure of eye movement sequences. To overcome this limitation and represent the overall dynamics of gaze transitions, the *hierarchical flow of eye movements* approach introduced by Burch et al. (2018) proposes an innovative graph-based visualization.

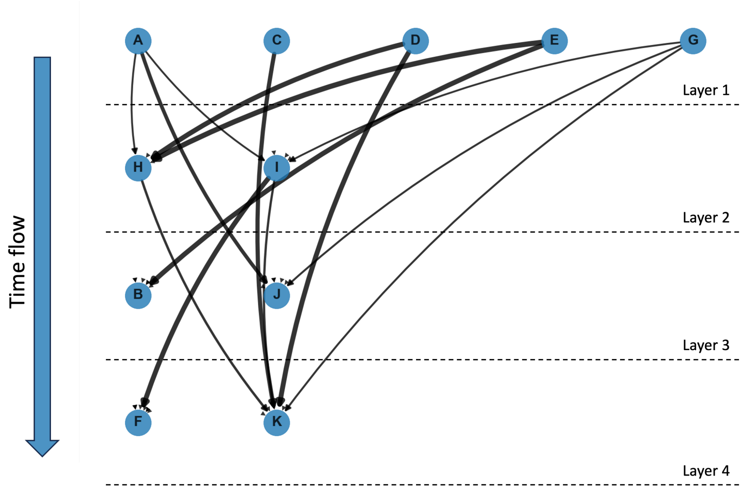

Figure 8. Flow diagram. An illustration of a flow diagram is provided, where AoIs are depicted as circular nodes arranged in vertically stacked layers according to the average order of their appearance in the analyzed sequences. Nodes corresponding to AoIs with earlier occurrences are positioned higher in the diagram, establishing a top-to-bottom flow. Transitions between AoIs are visualized as directed edges, with the thickness of each edge proportional to the frequency of the transitions, effectively highlighting the flow and magnitude of attention shifts.

In this approach, areas of interest are represented as vertices in a graph and visually depicted as circular boxes, as illustrated in Figure 8. These vertices are organized into vertically stacked layers according to the average order in which they appear within the analyzed AoI sequences. The resulting top-to-bottom layout reflects the typical progression of visual attention, with AoIs associated with earlier occurrences positioned higher in the diagram. Similar to directed graphs, transitions between AoIs are represented by oriented edges, whose thickness encodes transition frequency. The visualization thus builds upon a previously computed AoI transition matrix to highlight dominant transition patterns. Unlike standard directed graphs, however, the layered structure explicitly embeds temporal ordering into the spatial layout, allowing dominant gaze flows and global viewing strategies to be interpreted at a glance.

To highlight visual patterns at the sequence level, a *word tree* (Wattenberg and Viégas (2008)) visualization can be employed to compare AoI sequences across multiple observers. Tree structures are particularly well suited for illustrating the sequential organization of gaze patterns. In a word tree, each node represents an AoI, and branches display subsequent AoIs or subsequences of AoIs—many branches indicating a diversity of search strategies—thereby mapping the flow of gaze behavior. Each path from the root to a leaf node represents a unique observed AoI sequence, and, with the exception of the root, each node is connected to its parent by a curved link to ensure visual continuity. The size of a node's label reflects the frequency of sequences containing the AoI at that position; by construction, the frequency at a node is always greater than or equal to the sum of its children. Word trees may be either *left-rooted* or *right-rooted*, enabling analysts to examine which AoIs tend to precede or follow a given region of interest.

Word trees have been integrated into several visual analysis tools, including eSeeTrack (Tsang et al. (2010)) and EyeC (Ristovski et al. (2013)), often in combination with other visualization techniques. For example, eSeeTrack merges word trees with AoI clouds and suffix-tree-based representations to support exploratory analysis of AoI sequences. Building on this concept, AoI transition trees (Kurzhals and Weiskopf (2015)) have been proposed to visualize gaze patterns using thumbnail images to represent AoIs, with thumbnail size encoding frequency. While this compact representation improves visual efficiency, it lacks one of the key advantages of word trees: explicit textual labeling, which allows analysts to directly interpret the semantic content of gaze paths.

However, these approaches are generally limited to relatively short sequences. For long or highly variable sequences, inter-participant differences become too pronounced, effectively fragmenting viewers into distinct groups. To address this limitation, Kurzhals and Weiskopf (2015) proposed extending a single AoI transition tree into a *sequence of AoI transition trees*. This approach is particularly well suited to dynamic stimuli composed of successive scenes or semantically distinct segments. By exploiting natural temporal or semantic boundaries, a *divide-and-conquer* strategy can be applied to partition gaze data into coherent sections. Transition trees are then constructed for each segment, and corresponding AoIs across consecutive trees may be linked, yielding a sequence of smaller, interpretable transition trees rather than a single, visually overwhelming structure. As noted by Kurzhals and Weiskopf (2015), many dynamic stimuli can be naturally decomposed into semantically coherent units—such as task phases—allowing this approach to generalize across a wide range of experimental paradigms.

7 DISCUSSION

The study of areas of interest and gaze dynamics constitutes a cornerstone of eye-tracking research and, more broadly, of the investigation of human attention, perception, and behavior. By combining cognitive theory, computational modeling, and experimental observation, this interdisciplinary field has progressively established itself as a key methodological framework for understanding how visual information is sampled, structured, and exploited by observers. Recent advances in eye-tracking technology — including remote, mobile, and wearable devices — have dramatically increased both the ecological validity and the volume of gaze data that can be collected. These developments have enabled the analysis of gaze behavior across a wide range of contexts, from highly controlled laboratory experiments to naturalistic and real-world environments.

Despite these technological advances, the interpretation of gaze data remains intrinsically challenging. Visual trajectories are high-dimensional, noisy, and highly variable across individuals, tasks, and stimuli. Extracting meaningful visual strategies from such data therefore requires a careful balance between abstraction and fidelity to the original signal. As discussed throughout this review series (Laborde et al. (2024a,b,c)), progress in this domain depends not only on improvements in hardware precision, but also on the development of robust methodological frameworks capable of capturing both individual variability and group-level regularities.

A central theme emerging from this work is the role of abstraction in gaze analysis. Areas of interest, AoI sequences, transition models, and representative scanpaths all rely on transforming raw gaze recordings into symbolic or structured representations. While such abstraction is indispensable for quantitative analysis, it inevitably entails a loss of information, particularly with respect to fine-grained spatial detail, temporal micro-dynamics, and idiosyncratic viewing behaviors. The challenge, therefore, is not to eliminate abstraction, but to control it — ensuring that the chosen level of representation remains appropriate to the research question at hand. In particular, once areas of interest are defined, they constitute the symbolic alphabet from which AoI sequences are constructed. This makes the AoI definition stage a pivotal upstream decision, as it conditions the statistical and structural properties of all subsequent sequence-based analyses.

Within this context, visualization techniques play a crucial mediating role. Spatio-temporal visualizations, transition-based diagrams, and sequence-oriented representations provide an interpretable interface between raw gaze data and abstract computational models. These tools are particularly valuable for exploratory analysis, hypothesis generation, and qualitative validation, allowing researchers to visually inspect temporal structures, recurrent patterns, or anomalous behaviors before formalizing them quantitatively. However, as emphasized in this review, visualization methods are predominantly qualitative and descriptive. While they offer powerful insights into gaze dynamics, they do not, on their own, provide the formal rigor required for systematic comparison, statistical inference, or model-based prediction.

Conversely, a wide range of quantitative approaches — including Markov-based models, entropy measures, pattern-mining algorithms, and consensus sequence methods — enable the formal characterization of gaze behavior. Many of these techniques originate from domains such as bioinformatics, information theory, or data mining, and have been adapted to the specific constraints of eye-tracking data. These methods make it possible to quantify visual strategies, compare observer groups, and relate gaze behavior to task demands, expertise, or cognitive load. Rather than opposing qualitative and quantitative paradigms, the results of this review highlight the necessity of their integration. Visualization supports interpretability and validation, while quantitative modeling provides formalization, generalization, and reproducibility. Effective gaze analysis workflows therefore rely on iterative interactions between these two perspectives.

Another key issue concerns the scalability and generality of existing methods. Many algorithms perform well on small or moderately sized datasets, or under assumptions of limited inter-individual variability. However, as datasets grow larger and experimental designs become more complex — particularly in naturalistic or dynamic settings — the limitations of certain approaches become apparent. Some methods struggle with computational complexity, others with sensitivity to parameter choices, and others with their ability to accommodate heterogeneous visual strategies without excessive information loss. These limitations underscore the need for adaptive, data-driven approaches capable of balancing robustness and interpretability.

The increasing methodological sophistication of gaze analysis also raises important challenges in terms of expertise and collaboration. Quantitative methods for AoI sequence analysis often require advanced knowledge in applied mathematics, statistics, or algorithm design, while neurophysiological, clinical, and ergonomic expertise remains essential for grounding computational findings in meaningful cognitive interpretations. As a result, the analysis of gaze dynamics — encompassing raw gaze data, scanpaths, and AoI-based representations — increasingly demands genuinely multidisciplinary research efforts, as well as shared conceptual frameworks and methodological vocabularies.

Finally, reproducibility and methodological transparency remain critical challenges for the field. Beyond issues of implementation and reporting, a more fundamental difficulty lies in the inherently ill-posed nature of several core steps in AoI-based gaze analysis. Different, yet equally plausible, modeling assumptions for AoI definition, sequence abstraction, or transition estimation can lead to markedly different representations of visual behavior. As a result, downstream analyses—such as transition modeling, pattern mining, or consensus sequence extraction—may vary substantially depending on upstream methodological choices.

In practice, the large volume and complexity of gaze data necessitate standardized preprocessing pipelines to ensure data quality and comparability across studies. At the same time, many algorithms used for AoI definition and sequence analysis are highly parameter-dependent, and these parameter choices are not always fully reported or systematically justified. Moreover, the absence of openly accessible implementations for some widely used methods further complicates replication, benchmarking, and meaningful methodological comparison.

Addressing these issues is essential for consolidating gaze research into a cumulative and reproducible body of knowledge. Clearer methodological reporting, sensitivity analyses with respect to parameter choices, and systematic comparative evaluations are therefore indispensable for strengthening the reliability and interpretability of AoI-based gaze analyses.

Looking ahead, several research directions emerge as particularly important for advancing the analysis of gaze dynamics. First, the definition and tracking of areas of interest in dynamic and interactive stimuli remains an open challenge: spatially similar fixations may correspond to temporally distinct or semantically different visual elements, limiting the validity of static AoI representations. Second, hybrid methodological pipelines that combine data-driven AoI inference, explicit temporal modeling, and interpretable visualization appear especially promising for balancing robustness, flexibility, and cognitive plausibility. Such approaches may help reconcile the need for abstraction with the preservation of meaningful temporal structure.

More broadly, future progress will depend on the adoption of shared evaluation protocols, benchmark datasets, and reporting standards that make methodological choices explicit and results comparable across studies, tasks, and populations. By clarifying the assumptions underlying AoI definition, sequence abstraction, and model selection, the field can move toward more principled and transparent analyses. Ultimately, strengthening these foundations will enable gaze research to better integrate qualitative insight and quantitative rigor, and to more reliably link observed visual behavior to underlying cognitive and perceptual processes.

CONFLICT OF INTEREST STATEMENT

Author QL was employed by company SNCF. Author AR was employed by company Thales AVS France. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

QL: Formal Analysis, Methodology, Writing – original draft, Writing – review & editing. AR: Formal Analysis, Writing – original draft, Writing – review & editing. AA: Validation, Writing – review & editing. NV: Supervision, Methodology, Validation, Writing – review & editing. IB: Supervision, Methodology, Validation, Writing – review & editing. LO: Supervision, Methodology, Validation, Writing – review & editing.

ACKNOWLEDGMENTS

Part of this work has been founded by the Industrial Data Analytics And Machine Learning chairs of ENS Paris-Saclay.

REFERENCES

Achanta, R., Hemami, S., Estrada, F., and Susstrunk, S. (2009). Frequency-tuned salient region detection. In *2009 IEEE conference on computer vision and pattern recognition* (IEEE), 1597–1604

Agrawal, R., Imieliński, T., and Swami, A. (1993). Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD international conference on Management of data*. 207–216

Agrawal, R. and Srikant, R. (1995). Mining sequential patterns. In *Proceedings of the eleventh international conference on data engineering* (IEEE), 3–14

Agrawal, R., Srikant, R., et al. (1994). Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB* (Santiago, Chile), vol. 1215, 487–499

Ahsan, Z. and Obaidellah, U. (2023). Is clustering novice programmers possible? investigating scanpath trend analysis in programming tasks. In *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications*. 1–7

Albanesi, M. G., Gatti, R., Porta, M., and Ravarelli, A. (2011). Towards semi-automatic usability analysis through eye tracking. In *Proceedings of the 12th International Conference on Computer Systems and Technologies*. 135–141

Arbelaez, P., Maire, M., Fowlkes, C., and Malik, J. (2010). Contour detection and hierarchical image segmentation. *IEEE transactions on pattern analysis and machine intelligence* 33, 898–916

Arora, N., Fung, G., and Tunuguntla, S. (2017). T-patterns in business. *Available at SSRN 3066839*

Atweh, J. A., Hayek, J. A., and Riggs, S. L. (2023). Quantifying visual attention of teams during workload transitions using aoi-based cross-recurrence metrics. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (SAGE Publications Sage CA: Los Angeles, CA), 21695067231193683

Ayres, J., Flannick, J., Gehrke, J., and Yiu, T. (2002). Sequential pattern mining using a bitmap representation. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*. 429–435

Bakardzhiev, H., van der Burgt, M., Martins, E., van den Dool, B., Jansen, C., van Scheppingen, D., et al. (2021). A web-based eye tracking data visualization tool. In *Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10–15, 2021, Proceedings, Part III* (Springer), 405–419

Barz, M., Bhatti, O. S., Alam, H. M. T., Nguyen, D. M. H., and Sonntag, D. (2023). Interactive fixation-to-aoi mapping for mobile eye tracking data based on few-shot image classification. In *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces*. 175–178

Bednarik, R., Myller, N., Sutinen, E., and Tukiainen, M. (2005). Applying eye-movement tracking to program visualization. *Proceedings of the 2005 IEEE Symposium on Visual Languages and Human-Centric Computing* , 302–304

Bergroth, L., Hakonen, H., and Raita, T. (2000). A survey of longest common subsequence algorithms. In *Proceedings Seventh International Symposium on String Processing and Information Retrieval. SPIRE 2000* (IEEE), 39–48

Bianco, S., Gasparini, F., and Schettini, R. (2015). Color coding for data visualization. In *Encyclopedia of Information Science and Technology, Third Edition* (IGI Global). 1682–1691

Bilmes, J. A. et al. (1998). A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. *International computer science institute* 4, 126

Bishop, C. (2006). Pattern recognition and machine learning. *Springer google schola* 2, 35–42

Blascheck, T., Kurzhals, K., Raschke, M., Burch, M., Weiskopf, D., and Ertl, T. (2014). State-of-the-art of visualization for eye tracking data. In *Eurovis (stars)*. 29

Blascheck, T., Kurzhals, K., Raschke, M., Burch, M., Weiskopf, D., and Ertl, T. (2017). Visualization of eye tracking data: A taxonomy and survey. In *Computer Graphics Forum* (Wiley Online Library), vol. 36, 260–284

Blascheck, T., Kurzhals, K., Raschke, M., Strohmaier, S., Weiskopf, D., and Ertl, T. (2016). Aoi hierarchies for visual exploration of fixation sequences. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*. 111–118

Blascheck, T., Raschke, M., and Ertl, T. (2013). Circular heat map transition diagram. In *Proceedings of the 2013 Conference on Eye Tracking South Africa*. 58–61

Boisvert, J. F. and Bruce, N. D. (2016). Predicting task from eye movements: On the importance of spatial distribution, dynamics, and image features. *Neurocomputing* 207, 653–668

Burch, M., Kull, A., and Weiskopf, D. (2013). Aoi rivers for visualizing dynamic eye gaze frequencies. In *Computer Graphics Forum* (Wiley Online Library), vol. 32, 281–290

Burch, M., Kumar, A., and Mueller, K. (2018). The hierarchical flow of eye movements. In *Proceedings of the 3rd Workshop on Eye Tracking and Visualization*. 1–5

Burch, M., Raschke, M., and Weiskopf, D. (2010). Indented pixel tree plots. In *Advances in Visual Computing: 6th International Symposium, ISVC 2010, Las Vegas, NV, USA, November 29-December 1, 2010. Proceedings, Part I 6* (Springer), 338–349

Burch, M., Wallner, G., Broeks, N., Piree, L., Boonstra, N., Vlaswinkel, P., et al. (2021). The power of linked eye movement data visualizations. In *ACM symposium on eye tracking research and applications*. 1–11

Burmester, M. and Mast, M. (2010). Repeated web page visits and the scanpath theory: A recurrent pattern detection approach. *Journal of Eye Movement Research* 3

Castner, N., Geßler, L., Geisler, D., Hüttig, F., and Kasneci, E. (2020a). Towards expert gaze modeling and recognition of a user's attention in realtime. *Procedia Computer Science* 176

Castner, N., Kuebler, T. C., Scheiter, K., Richter, J., Eder, T., Hüttig, F., et al. (2020b). Deep semantic gaze embedding and scanpath comparison for expertise classification during opt viewing. In *ACM Symposium on Eye Tracking Research and Applications*. 1–10

Chen, X. and Chen, Z. (2017). Exploring visual attention using random walks based eye tracking protocols. *Journal of Visual Communication and Image Representation* 45, 147–155

Chuk, T., Chan, A. B., and Hsiao, J. H. (2014). Understanding eye movements in face recognition using hidden markov models. *Journal of vision* 14, 8–8

Chuk, T., Chan, A. B., and Hsiao, J. H. (2017). Is having similar eye movement patterns during face learning and recognition beneficial for recognition performance? evidence from hidden markov modeling. *Vision research* 141, 204–216

Church, K. W. and Helfman, J. I. (1993). Dotplot: A program for exploring self-similarity in millions of lines of text and code. *Journal of Computational and Graphical Statistics* 2, 153–174

Claus, M., Hermens, F., and Bromuri, S. (2023). A user study of visualisations of spatio-temporal eye tracking data. *arXiv preprint arXiv:2309.15731*

Connor, M. C., Glass, B. H., Finkenstaedt-Quinn, S. A., and Shultz, G. V. (2020). Developing expertise in 1h nmr spectral interpretation. *The Journal of Organic Chemistry* 86, 1385–1395

Coutrot, A., Hsiao, J. H., and Chan, A. B. (2018). Scanpath modeling and classification with hidden markov models. *Behavior research methods* 50, 362–379

Coviello, E., Chan, A. B., and Lanckriet, G. R. (2014). Clustering hidden markov models with variational hem. *The Journal of Machine Learning Research* 15, 697–747

Coviello, E., Lanckriet, G., and Chan, A. (2012). The variational hierarchical em algorithm for clustering hidden markov models. *Advances in neural information processing systems* 25

Crowe, E. C. and Narayanan, N. H. (2000). Comparing interfaces based on what users watch and do. In *Proceedings of the 2000 symposium on Eye tracking research & applications*. 29–36

Dayan, P. (1993). Improving generalization for temporal difference learning: The successor representation. *Neural computation* 5, 613–624

Drusch, G. and Bastien, J. C. (2012). Analyzing visual scanpaths on the web using the mean shift procedure and t-pattern detection: a bottom-up approach. In *Proceedings of the 2012 conference on ergonomie et interaction homme-machine*. 181–184

Duchowski, A. T., Driver, J., Jolaoso, S., Tan, W., Ramey, B. N., and Robbins, A. (2010). Scanpath comparison revisited. In *Proceedings of the 2010 symposium on eye-tracking research & applications*. 219–226

Elbattah, M., Carette, R., Cilia, F., Guérin, J.-L., and Dequen, G. (2023). Applications of machine learning methods to assist the diagnosis of autism spectrum disorder. In *Neural Engineering Techniques for Autism Spectrum Disorder, Volume 2* (Elsevier). 99–119

Ellahi, W. (2022). *Measuring and comparing visual attention deployment: application to Quality of Experience assessment in the context of hyper-realistic content*. Ph.D. thesis, Nantes Université

Eraslan, S., Yesilada, Y., and Harper, S. (2014). Identifying patterns in eyetracking scanpaths in terms of visual elements of web pages. In *Web Engineering: 14th International Conference, ICWE 2014, Toulouse, France, July 1-4, 2014. Proceedings 14* (Springer), 163–180

Eraslan, S., Yesilada, Y., and Harper, S. (2016). Scanpath trend analysis on web pages: Clustering eye tracking scanpaths. *ACM Transactions on the Web (TWEB)* 10, 1–35

Eraslan, S., Yesilada, Y., and Harper, S. (2017). Engineering web-based interactive systems: trend analysis in eye tracking scanpaths with a tolerance. In *Proceedings of the ACM SIGCHI symposium on engineering interactive computing systems*. 3–8

Eraslan, S., Yesilada, Y., Yaneva, V., and Harper, S. (2020). Autism detection based on eye movement sequences on the web: a scanpath trend analysis approach. In *Proceedings of the 17th International Web for All Conference*. 1–10

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*. vol. 96, 226–231

Fischer, P. and Peinsipp-Byma, E. (2007). Eye tracking for objective usability evaluation. In *European Conference on Eye Movements (ECEM). ECEM*

Fitts, P. M., Jones, R. E., and Milton, J. L. (2005). Eye movements of aircraft pilots during instrument-landing approaches. *Ergonomics: Major Writings*

Frey, B. J. and Dueck, D. (2007). Clustering by passing messages between data points. *science* 315, 972–976

Fuhl, W., Kübler, T., Sippel, K., Rosenstiel, W., and Kasneci, E. (2015). Arbitrarily shaped areas of interest based on gaze density gradient. In *European Conference on Eye Movements, ECEM*. vol. 8, 9

Fuhl, W., Kübler, T. C., Santini, T., and Kasneci, E. (2018a). Automatic generation of saliency-based areas of interest for the visualization and analysis of eye-tracking data. In *VMV*. 47–54

Fuhl, W., Kuebler, T. C., Brinkmann, H., Rosenberg, R., Rosenstiel, W., and Kasneci, E. (2018b). Region of interest generation algorithms for eye tracking data. In *Proceedings of the 3rd Workshop on Eye Tracking and Visualization (ETVIS'18)* (ACM)

Fukunaga, K. and Hostetler, L. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on information theory* 21, 32–40

Geisler, D., Castner, N., Kasneci, G., and Kasneci, E. (2020). A minhash approach for fast scanpath classification. In *ACM Symposium on Eye Tracking Research and Applications*. 1–9

Gibbs, A. J. and McIntyre, G. A. (1970). The diagram, a method for comparing sequences: Its use with amino acid and nucleotide sequences. *European journal of biochemistry* 16, 1–11

Goldberg, J. and Helfman, J. (2009). Pattern analysis in eye tracking data: Dotplots revisited. *Proceedings of the CHI 2009*

Goldberg, J. H. and Helfman, J. I. (2010). Scanpath clustering and aggregation. In *Proceedings of the 2010 symposium on eye-tracking research & applications*. 227–234

Goldberg, J. H. and Kotval, X. P. (1999). Computer interface evaluation using eye movements: methods and constructs. *International journal of industrial ergonomics* 24, 631–645

Guo, C., Ma, Q., and Zhang, L. (2008). Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform. In *2008 IEEE conference on computer vision and pattern recognition* (IEEE), 1–8

Haji-Abolhassani, A. and Clark, J. J. (2013). A computational model for task inference in visual search. *Journal of vision* 13, 29–29

Haji-Abolhassani, A. and Clark, J. J. (2014). An inverse yarbus process: Predicting observers' task from eye movement patterns. *Vision research* 103, 127–142

Hao, Q., Ma, L., Sbert, M., Feixas, M., and Zhang, J. (2020). Gaze information channel in van gogh's paintings. *Entropy* 22, 540

Hao, Q., Sbert, M., and Ma, L. (2019). Gaze information channel in cognitive comprehension of poster reading. *Entropy* 21, 444

Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, vol. 2 (Springer)

Hayes, T. R., Petrov, A. A., and Sederberg, P. B. (2011). A novel method for analyzing sequential eye movements reveals strategic influence on raven's advanced progressive matrices. *Journal of Vision* 11, 10–10

Hejmady, P. and Narayanan, N. H. (2012). Visual attention patterns during program debugging with an ide. In *proceedings of the symposium on eye tracking research and applications*. 197–200

Hellinger, E. (1909). Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen. *Journal für die reine und angewandte Mathematik* 1909, 210–271

Holmqvist, K., Holsanova, J., Barthelson, M., and Lundqvist, D. (2003). Reading or scanning? a study of newspaper and net paper reading. In *The Mind's Eye* (Elsevier). 657–670

Holmqvist, K., Nystrom, M., Andersson, R., Dewhurst, R., Jarodzka, H., and Van de Weijer, J. (2011). Eye tracking: A comprehensive guide to methods and measures. *OUP Oxford*

Hooge, I. and Camps, G. (2013). Scan path entropy and arrow plots: Capturing scanning behavior of multiple observers. *Frontiers in psychology* 4, 996

Hou, X. and Zhang, L. (2007). Saliency detection: A spectral residual approach. In *2007 IEEE Conference on computer vision and pattern recognition* (Ieee), 1–8

Hsiao, J. H., Lan, H., Zheng, Y., and Chan, A. B. (2021). Eye movement analysis with hidden markov models (emhmm) with co-clustering. *Behavior Research Methods* 53, 2473–2486

Huang, H., Doebler, P., and Mertins, B. (2024). Short-time aois-based representative scanpath identification and scanpath aggregation. *Behavior Research Methods* , 1–16

Huang, T.-H., Cheng, K.-Y., and Chuang, Y.-Y. (2009). A collaborative benchmark for region of interest detection algorithms. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE), 296–303

Huang, Y. and Zhang, L. (2004). Rapid and sensitive dot-matrix methods for genome analysis. *Bioinformatics* 20, 460–466

Itoh, K., Hansen, J.-P., and Nielsen, F. (1998). Cognitive modelling of a ship navigator based on protocol and eye-movement analysis. *Le Travail Humain* , 99–127

Itoh, K., Tanaka, H., and Seki, M. (2000). Eye-movement analysis of track monitoring patterns of night train operators: Effects of geographic knowledge and fatigue. In *Proceedings of the human factors and ergonomics society annual meeting* (SAGE Publications Sage CA: Los Angeles, CA), vol. 44, 360–363

Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence* 20, 1254–1259

Jayawardena, G. and Jayarathna, S. (2021). Automated filtering of eye movements using dynamic aoi in multiple granularity levels. *International Journal of Multimedia Data Engineering and Management (IJMDEM)* 12, 49–64

Jraidi, I., Chaouachi, M., Ben Khedher, A., Lajoie, S. P., and Frasson, C. (2022). Understanding clinical reasoning through visual scanpath and brain activity analysis. *Computation* 10, 130

Just, M. A. and Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological review* 87, 329

Kanan, C., Bseiso, D. N., Ray, N. A., Hsiao, J. H., and Cottrell, G. W. (2015). Humans have idiosyncratic and task-specific scanpaths for judging faces. *Vision research* 108, 67–76

Kanan, C., Ray, N. A., Bseiso, D. N., Hsiao, J. H., and Cottrell, G. W. (2014). Predicting an observer's task using multi-fixation pattern analysis. In *Proceedings of the symposium on eye tracking research and applications*. 287–290

Kaufman, L. and Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis* (John Wiley & Sons)

Kopácsi, L., Barz, M., Bhatti, O. S., and Sonntag, D. (2023). Imeta: An interactive mobile eye tracking annotation method for semi-automatic fixation-to-aoi mapping. In *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces*. 33–36

Kosel, C., Holzberger, D., and Seidel, T. (2021). Identifying expert and novice visual scanpath patterns and their relationship to assessing learning-relevant student characteristicshao2019gaze. In *Frontiers in Education* (Frontiers Media SA), vol. 5, 612175

Krejtz, I., Szarkowska, A., and Krejtz, K. (2013). The effects of shot changes on eye movements in subtitling

Krejtz, K., Duchowski, A., Szmidt, T., Krejtz, I., González Perilli, F., Pires, A., et al. (2015). Gaze transition entropy. *ACM Transactions on Applied Perception (TAP)* 13, 1–20

Krejtz, K., Szmidt, T., Duchowski, A. T., and Krejtz, I. (2014). Entropy-based statistical analysis of eye movement transitions. In *Proceedings of the Symposium on Eye Tracking Research and Applications*. 159–166

Kübler, T. C., Rothe, C., Schiefer, U., Rosenstiel, W., and Kasneci, E. (2017). Subsmatch 2.0: Scanpath comparison and classification based on subsequence frequencies. *Behavior research methods* 49, 1048–1064

Kurzhals, K., Burch, M., Blascheck, T., Andrienko, G., Andrienko, N., and Weiskopf, D. (2017). A task-based view on the visual analysis of eye-tracking data. In *Eye Tracking and Visualization: Foundations, Techniques, and Applications. ETVIS 2015 1* (Springer), 3–22

Kurzhals, K., Hlawatsch, M., Heimerl, F., Burch, M., Ertl, T., and Weiskopf, D. (2015). Gaze stripes: Image-based visualization of eye tracking data. *IEEE transactions on visualization and computer graphics* 22, 1005–1014

Kurzhals, K., Hlawatsch, M., Seeger, C., and Weiskopf, D. (2016). Visual analytics for mobile eye tracking. *IEEE transactions on visualization and computer graphics* 23, 301–310

Kurzhals, K. and Weiskopf, D. (2013). Space-time visual analytics of eye-tracking data for dynamic stimuli. *IEEE Transactions on Visualization and Computer Graphics* 19, 2129–2138

Kurzhals, K. and Weiskopf, D. (2015). Aoi transition trees. In *Graphics Interface*. 41–48

Laborde, Q., Roques, A., Bargiotas, I., Armougum, A., Vayatis, N., and Oudre, L. (2024a). Vision part 1: Neurophysiology and experimental paradigms [manuscript submitted for publication]. *Université Paris Saclay, Université Paris Cité, ENS Paris Saclay, CNRS, SSA, INSERM, Centre Borelli, F-91190, France*

Laborde, Q., Roques, A., Bargiotas, I., Armougum, A., Vayatis, N., and Oudre, L. (2024b). Vision part 2: Oculomotor processing [manuscript submitted for publication]. *Université Paris Saclay, Université Paris Cité, ENS Paris Saclay, CNRS, SSA, INSERM, Centre Borelli, F-91190, France*

Laborde, Q., Roques, A., Bargiotas, I., Armougum, A., Vayatis, N., and Oudre, L. (2024c). Vision part 3: Scanpaths and derived representations [manuscript submitted for publication]. *Université Paris Saclay, Université Paris Cité, ENS Paris Saclay, CNRS, SSA, INSERM, Centre Borelli, F-91190, France*

Lagmay, E. A. D., Rodrigo, M. M. T., et al. (2022). Enhanced automatic areas of interest (aoi) bounding boxes estimation algorithm for dynamic eye-tracking stimuli. *APSIPA Transactions on Signal and Information Processing* 11

Latimer, C. R. (1988). Eye-movement data: Cumulative fixation time and cluster analysis. *Behavior Research Methods, Instruments, & Computers* 20, 437–470

Lempel, A. and Ziv, J. (1976). On the complexity of finite sequences. *IEEE Transactions on information theory* 22, 75–81

Li, A. and Chen, Z. (2018). Representative scanpath identification for group viewing pattern analysis. *Journal of Eye Movement Research* 11

Li, A., Zhang, Y., and Chen, Z. (2017). Scanpath mining of eye movement trajectories for visual attention analysis. In *2017 IEEE International Conference on Multimedia and Expo (ICME)* (IEEE), 535–540

Lounis, C., Peysakhovich, V., and Causse, M. (2020). Complexity of dwell sequences: visual scanning pattern differences between novice and expert aircraft pilots. *In 1st International Workshop on Eye-Tracking in Aviation*

Lounis, C., Peysakhovich, V., and Causse, M. (2021). Visual scanning strategies in the cockpit are modulated by pilots' expertise: A flight simulator study. *PLoS one* 16, e0247061

Ma, R. and Angryk, R. (2017). Distance and density clustering for time series data. In *2017 IEEE international conference on data mining workshops (ICDMW)* (IEEE), 25–32

Ma, X., Liu, Y., Clariana, R., Gu, C., and Li, P. (2023). From eye movements to scanpath networks: A method for studying individual differences in expository text reading. *Behavior research methods* 55, 730–750

Mackworth, N. H. and Morandi, A. J. (1967). The gaze selects informative details within pictures. *Perception & psychophysics* 2, 547–552

Magnusson, M. S. (2000). Discovering hidden time patterns in behavior: T-patterns and their detection. *Behavior research methods, instruments, & computers* 32, 93–110

Maizel Jr, J. V. and Lenk, R. P. (1981). Enhanced graphic matrix analysis of nucleic acid and protein sequences. *Proceedings of the National Academy of Sciences* 78, 7665–7669

Marchal, F., Castagnos, S., and Boyer, A. (2016). Tell me what you see, i will tell you what you remember. In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*. 293–294

Mast, M. and Burmester, M. (2011). Exposing repetitive scanning in eye movement sequences with t-pattern detection. In *Proceedings IADIS International conference interfaces and human computer interaction (IHCI)*. 137–145

McGrory, C. A. and Titterington, D. (2009). Variational bayesian analysis for hidden markov models. *Australian & New Zealand Journal of Statistics* 51, 227–244

McGuire, M. P. and Chakraborty, J. (2016). Directional scan path characterization of eye tracking sequences: A multi-scale approach. In *2016 Future Technologies Conference (FTC)* (IEEE), 51–61

McIntyre, N. A., Mainhard, M. T., and Klassen, R. M. (2017). Are you looking to teach? cultural, temporal and dynamic insights into expert teacher gaze. *Learning and Instruction* 49, 41–53

Melsted, P. and Pritchard, J. K. (2011). Efficient counting of k-mers in dna sequences using a bloom filter. *BMC bioinformatics* 12, 1–7

Meyer, C. (2000). Perron–frobenius theory of nonnegative matrices. *Matrix analysis and applied linear algebra, SIAM*

Naqshbandi, K., Gedeon, T., and Abdulla, U. A. (2016). Automatic clustering of eye gaze data for machine learning. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (IEEE), 001239–001244

Nguyen, A., Chandran, V., and Sridharan, S. (2006). Gaze tracking for region of interest coding in jpeg 2000. *Signal Processing: Image Communication* 21, 359–377

Obaidellah, U., Blascheck, T., Guarnera, D. T., and Maletic, J. (2020). A fine-grained assessment on novice programmers' gaze patterns on pseudocode problems. In *ACM Symposium on Eye Tracking Research and Applications*. 1–5

Opach, T., Golebiowska, I., and Fabrikant, S. I. (2014). How do people view multi-component animated maps? *The Cartographic Journal* 51, 330–342

Petitjean, F., Ketterlin, A., and Gançarski, P. (2011). A global averaging method for dynamic time warping, with applications to clustering. *Pattern recognition* 44, 678–693

Pillalamarri, R. S., Barnette, B. D., Birkmire, D., and Karsh, R. (1993). Cluster: A program for the identification of eye-fixation-cluster characteristics. *Behavior Research Methods, Instruments, & Computers* 25, 9–15

Plummer, P., DeWolf, M., Bassok, M., Gordon, P. C., and Holyoak, K. J. (2017). Reasoning strategies with rational numbers revealed by eye tracking. *Attention, Perception, & Psychophysics* 79, 1426–1437

Ponsoda, V., Scott, D., and Findlay, J. (1995). A probability vector and transition matrix analysis of eye movements during visual search. *Acta Psychologica* 88, 167–185

Privitera, C. M. and Stark, L. W. (1998). Evaluating image processing algorithms that predict regions of interest. *Pattern Recognition Letters* 19, 1037–1043

Privitera, C. M. and Stark, L. W. (2000). Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *IEEE Transactions on pattern analysis and machine intelligence* 22, 970–982

Qazi, M., Tunuguntla, S., Lee, P., Kanchinadam, T., Fung, G., and Arora, N. (2019). Discovering temporal patterns from insurance interaction data. In *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 33, 9573–9580

Räihä, K.-J., Aula, A., Majaranta, P., Rantala, H., and Koivunen, K. (2005). Static visualization of temporal eye-tracking data. In *Human-Computer Interaction-INTERACT 2005: IFIP TC13 International Conference, Rome, Italy, September 12-16, 2005. Proceedings 10* (Springer), 946–949

Raschke, M., Chen, X., and Ertl, T. (2012). Parallel scan-path visualization. In *Proceedings of the Symposium on Eye Tracking Research and Applications*. 165–168

Reani, M., Peek, N., and Jay, C. (2018). An investigation of the effects of n-gram length in scanpath analysis for eye-tracking research. *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications* , 1–8

Rees, D., Laramee, R. S., Brookes, P., and D'Cruze, T. (2020). Interaction techniques for chord diagrams. In *2020 24th International Conference Information Visualisation (IV)* (IEEE), 28–37

Richardson, D. C. and Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive science* 29, 1045–1060

Riehmann, P., Hanfler, M., and Froehlich, B. (2005). Interactive sankey diagrams. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.* (IEEE), 233–240

Ristovski, G., Hunter, M., Olk, B., and Linsen, L. (2013). Eyec: Coordinated views for interactive visual exploration of eye-tracking data. In *2013 17th International Conference on Information Visualisation* (IEEE), 239–248

Rodriguez, A. and Laio, A. (2014). Clustering by fast search and find of density peaks. *science* 344, 1492–1496

Salah, A. A., Pauwels, E., Tavenard, R., and Gevers, T. (2010). T-patterns revisited: mining for temporal patterns in sensor data. *Sensors* 10, 7496–7513

Santella, A. and DeCarlo, D. (2004). Robust clustering of eye movement recordings for quantification of visual interest. In *Proceedings of the 2004 symposium on Eye tracking research & applications*. 27–34

Setubal, J. C., Meidanis, J., and Setubal-Meidanis (1997). *Introduction to computational molecular biology*. 04; QH506, S4. (PWS Pub. Boston)

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal* 27, 379–423

Shi, J. and Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence* 22, 888–905

Smith, T. F., Waterman, M. S., et al. (1981). Identification of common molecular subsequences. *Journal of molecular biology* 147, 195–197

Špakov, O. and Miniotas, D. (2007). Application of clustering algorithms in eye gaze visualizations. *Information Technology and Control* 36

Tablatin, C. L. and Rodrigo, M. M. (2018). Identifying common code reading patterns using scanpath trend analysis with a tolerance. In *Proceedings of thee 26th International Conference for Computers in Education (ICCE 2018), Metro Manila, Philippines*

Tang, H., Day, E., Kendhammer, L., Moore, J. N., Brown, S. A., and Pienta, N. J. (2016). Eye movement patterns in solving science ordering problems. *Journal of eye movement research* 9

Tavenard, R., Salah, A. A., and Pauwels, E. J. (2008). Searching for temporal patterns in ami sensor data. In *Constructing Ambient Intelligence: AmI 2007 Workshops Darmstadt, Germany, November 7-10, 2007 Revised Papers* (Springer), 53–62

Tory, M., Atkins, M. S., Kirkpatrick, A. E., Nicolaou, M., and Yang, G.-Z. (2005). Eyegaze analysis of displays with combined 2d and 3d views. In *VIS 05. IEEE Visualization, 2005.* (IEEE), 519–526

Trajkovska, K., Kljun, M., and Pucihar, K. Č. (2024). Gaze2aoi: Open source deep-learning based system for automatic area of interest annotation with eye tracking data. *arXiv preprint arXiv:2411.13346*

Tsang, H. Y., Tory, M., and Swindells, C. (2010). eseetrack—visualizing sequential fixation patterns. *IEEE Transactions on Visualization and Computer Graphics* 16, 953–962

Van Erven, T. and Harremos, P. (2014). Rényi divergence and kullback-leibler divergence. *IEEE Transactions on Information Theory* 60, 3797–3820

Vandeberg, L., Bouwmeester, S., Bocanegra, B. R., and Zwaan, R. A. (2013). Detecting cognitive interactions through eye movement transitions. *Journal of Memory and Language* 69, 445–460

Vassallo, S. and Douglas, J. (2021). Visual scanpath training to emotional faces following severe traumatic brain injury: A single case design. *Journal of Eye Movement Research* 14

von der Malsburg, T., Kliegl, R., and Vasishth, S. (2015). Determinants of scanpath regularity in reading. *Cognitive science* 39, 1675–1703

Von Landesberger, T., Kuijper, A., Schreck, T., Kohlhammer, J., van Wijk, J. J., Fekete, J.-D., et al. (2011). Visual analysis of large graphs: state-of-the-art and future research challenges. In *Computer graphics forum* (Wiley Online Library), vol. 30, 1719–1749

Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J. R., Van der Helm, P. A., et al. (2012). A century of gestalt psychology in visual perception: Ii. conceptual and theoretical foundations. *Psychological bulletin* 138, 1218

Wang, F. S., Gianduzzo, C., Meboldt, M., and Lohmeyer, Q. (2022). An algorithmic approach to determine expertise development using object-related gaze pattern sequences. *Behavior Research Methods* 54, 493–507

Wang, Y., Chen, X., and Chen, Z. (2016). Towards region-of-attention analysis in eye tracking protocols. *Electronic Imaging* 2016, 1–6

Wang, Z. (2021). Eye movement data analysis and visualization. In *Eye-Tracking with Python and Pylink* (Springer). 197–224

Wattenberg, M. and Viégas, F. B. (2008). The word tree, an interactive visual concordance. *IEEE transactions on visualization and computer graphics* 14, 1221–1228

Weibel, N., Fouse, A., Emmenegger, C., Kimmich, S., and Hutchins, E. (2012). Let's look at the cockpit: exploring mobile eye-tracking for observational research on the flight deck. In *Proceedings of the symposium on eye tracking research and applications*. 107–114

West, J. M., Haake, A. R., Rozanski, E. P., and Karn, K. S. (2006). eyepatterns: software for identifying patterns and similarities across fixation sequences. In *Proceedings of the 2006 symposium on Eye tracking research & applications*. 149–154

Yang, C.-K. and Wacharamanotham, C. (2018). Alpscarf: Augmenting scarf plots for exploring temporal gaze patterns. In *Extended abstracts of the 2018 CHI conference on human factors in computing systems*. 1–6