

Simulated visual attention modelling in digital tourism interfaces: an integrative approach to destination communication

Lucilia Cardoso ^{1*}, Carla Fraga ^{3, 4}, Patrick Moratori ^{2, 3}, Marcelo Knop ^{3, 4},
Youssef El Archi ^{1, 5}, Sérgio Silva ^{8, 9} and Francisco Muñoz-Leiva ^{6, 7}

¹ CiTUR – Centre for Tourism Research, Development and Innovation, Leiria, University of Leiria and Oeste, Portugal.
Tourism Research Group, Neurosciences and Computing (TRG:N&C).. Email: lucilia.a.cardoso@ulo.pt

² Department of Computer Science, Fluminense Federal University (UFF), Brazil

³ Tourism Research Group, Neurosciences and Computing (TRG: N&C)

⁴ Department of Tourism, Federal University of Juiz de Fora (UFJF), Brazil.

⁵ Faculty of Law, Economics and Social Sciences – Souissi, Mohammed V University in Rabat, Rabat, Morocco

⁶ Department of Marketing and Market Research, University of Granada, Granada, Spain;

⁷ Andalusian Institute of Research and Innovation in Tourism (IATUR), Granada, Spain.

⁸ Engineering Department, School of Sciences and Technology, UTAD, Vila Real, Portugal

⁹ IEETA, Aveiro, Portugal

* Corresponding author

Abstract

This paper introduces a simulation-based methodological and conceptual study addressing a critical gap in tourism research: the fragmented treatment of spatial, sequential and temporal dimensions of visual attention. It proposes the Simulated Visual Attention Integrative (SVAI) Model, combining predicted saliency heatmaps, simulated scan paths and saccadic timing within a unified analytical framework. Guided by Visual Attention Theory, Gestalt Theory, Feature Integration Theory and Predictive Coding, the model offers a transparent, theory-driven and cost-effective alternative to empirical eye-tracking for analysing digital tourism interfaces. Applied to ten island pages of an official tourism website, results indicate that simulated attention is initially allocated to functional elements such as call-to-action buttons and branding, followed by a progressive shift towards scenic imagery. Predicted scan paths support this sequence, while temporal analysis reveals a transition from rapid scanning to slower, semantically oriented inspection. The findings demonstrate the analytical potential of simulated visual attention for destination communication research, with direct implications for European DMOs, digital marketers and interface designers seeking to optimise tourism websites prior to empirical testing.

Keywords: Simulated visual attention, Visual salience, Tourism websites, Destination communication, Visual attention modelling

Citation: Cardoso, L., Fraga, C., Moratori, P., Knop, M., El Archi, Y., Silva, S., & Muñoz-Leiva, F. (2026). Simulated visual attention modelling in digital tourism interfaces: an integrative approach to destination communication. *European Journal of Tourism Research*, 44, 4407. <https://doi.org/10.54055/ejtr.v44i.5046>.

Publication history:

Received: 05/04/2026; *Revised:* 25/06/2026; *Accepted:* 06/07/2026; *Published online:* 01/10/2026;

Volume date: 01/11/2026

Coordinating editor: Stanislav Ivanov



© 2026 The Author(s)

This work is licensed under the Creative Commons Attribution 4.0 International (CC BY 4.0). To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

1. Introduction

Visual attention plays a central role in how visual information is perceived and processed within digital environments, including online travel platforms where early impressions significantly shape destination image and influence subsequent decision-making (Li, Xiao & Seekamp, 2026). As tourism increasingly relies on digital channels to attract, inform, and persuade visitors, destination marketing organizations (DMOs) face growing pressure to design interfaces that effectively capture attention and convey compelling narratives (Fernández-Cavia *et al.*, 2017; Lee & Gretzel, 2012). Online travel platforms thus function not only as functional information hubs, but also as symbolic communication spaces through which destinations express identity, values, and authenticity (Confetto *et al.*, 2023). The allocation of visual attention within these environments has critical implications for persuasive communication, user experience, and destination competitiveness.

In response to this challenge, tourism research, particularly studies on digital and immersive technologies, has examined how digital and immersive technologies influence attentional processes in mediated environments, including destination websites and online travel platforms, where visual and textual cues jointly shape attentional allocation and decision-related processes (Guerrero-Rodríguez, Stepchenkova & Kirilenko, 2020; Wu & Lai, 2025). Within this stream of research, psychophysiological methods, most notably eye-tracking, have been widely employed to investigate visual attention, information processing, and decision-making in tourism-related interfaces (Savin *et al.*, 2021).

Empirical eye-tracking studies in tourism have shown that elements such as banners, eco-labels, sustainability cues, and destination images strongly guide attentional allocation patterns and influence cognitive processing, memory, and preferences (Köchling, 2020; Muñoz-Leiva *et al.*, 2021; Espigares-Jurado *et al.*, 2020; Penz *et al.*, 2017). These findings underscore the centrality of visual attention research for understanding destination image formation and advancing consumer behaviour theories related to online persuasion. Yet, despite valuable insights, most research has examined visual attention through participant-based eye-tracking experiments (Chen *et al.*, 2025; Li *et al.*, 2016). While these methods capture rich fixation and scan path data, they also face critical limitations (Valtakari *et al.*, 2021). They are resource-intensive, requiring costly hardware, controlled laboratory environments, and participant recruitment. They also impose logistical constraints that limit their application in early design or prototyping stages, when rapid iteration is essential.

In parallel, computational models of simulated visual attention have gained traction in cognitive science, computer vision, and human-computer interaction research (Borji *et al.*, 2012; Voßkühler *et al.*, 2008; Itti & Koch, 2001). These models generate simulation-based representations of attentional allocation, such as saliency heatmaps, simulated scan paths, and temporal saccadic metrics, without the need for human participants. Their advantages include cost-efficiency, transparency, replicability, and the ability to anticipate predicted attentional flows at early stages of interface conception and evaluation. Such models have been successfully used in domains such as advertising, web design, and human-machine interaction to pre-test visual layouts and support evidence-based communication strategies.

However, a systematic Scopus search conducted in April 2025 indicates that simulation-based approaches to visual attention remain virtually unexplored in tourism research. This gap is particularly striking given that online travel platforms are inherently multimodal and symbolic, simultaneously combining functional information (e.g., navigation and booking tools) with hedonic imagery and identity-related narratives. Unlike conventional e-commerce interfaces, they address utilitarian, experiential, and affective dimensions of interaction with digital content, making them a highly relevant yet under-researched context for simulation-based attention modelling.

Equally important, existing visual attention research in tourism has tended to focus on dimensions in isolation: spatial hotspots, sequential scan paths, or temporal fixation measures (Chen *et al.*, 2025). Yet these layers are inherently interdependent: spatial saliency shapes scan path trajectories, sequential flows influence temporal dynamics, and temporal rhythms affect the interpretation of interface coherence. Treating them separately risks producing fragmented insights and overlooking the cumulative impact of attentional dynamics on perception and decision-making. This fragmentation has limited theoretical integration and constrained practical guidance for interface designers seeking to anticipate how attention unfolds across complex digital environments.

Addressing this critical gap, the present study asks: How can simulated visual attention be analysed across spatial, sequential, and temporal layers to generate an integrated interpretation of attentional allocation patterns in digital tourism interfaces? The objective is to advance both theory and practice by developing a simulation-based analytical framework that systematically integrates these three dimensions of predicted visual attention into a holistic interpretation of visual salience and attentional organisation, without inferring actual user behaviour. To achieve this, the study adopts a simulation-based methodological approach that combines predicted saliency heatmaps, simulated scan paths, and saccadic timing metrics within an integrated analytical framework. Building on this research question and the proposed integrative framework, this study contributes in three main ways. First, it introduces simulated visual attention modelling as a methodological innovation in tourism research, demonstrating its potential as a transparent, replicable, and cost-effective pre-testing and exploratory tool. Second, it advances theoretical understanding by integrating spatial, sequential, and temporal layers into a unified interpretive model, addressing the current fragmentation in visual attention studies. Third, it offers practical insights for DMOs, digital marketers, and sustainability communicators by showing how simulation-based attention modelling can support evidence-based website design, enhance the communication of sustainability cues, and strengthen destination competitiveness. Together, these contributions establish simulated visual attention not merely as a technical alternative to eye-tracking, but as a methodological bridge between cognitive modelling and applied digital tourism research.

2. Theoretical background

2.1. Visual attention and predictive mechanisms

Visual Attention Theory (VAT) establishes that attention can operate both reactively and proactively. Visual attention operates not only as a reactive process but also as an anticipatory and predictive system, providing a solid theoretical foundation for the simulation of visual attention in digital environments (Wickens, 2021). A growing body of literature demonstrates that attention functions in an anticipatory manner, continuously generating predictions about where visual information is likely to appear.

In computational neuroscience, the Free Energy Principle describes the brain as a system that reduces uncertainty by updating its internal models through predictive inference (Friston, 2010). Within this framework, attention emerges from the comparison between expected and actual sensory input, as described in hierarchical predictive coding models (Spratling, 2017). This predictive nature of attention supports the use of simulation-based approaches, in which visual attention can be modelled even in the absence of empirical eye-tracking data. This predictive perspective complements classical VAT, which historically distinguished between bottom-up (exogenous) and top-down (endogenous) mechanisms. Bottom-up attention is automatically captured by perceptually prominent stimuli, independent of the observer's goals, whereas top-down attention reflects deliberate, goal-driven focus influenced by prior knowledge and expectations (Corbetta & Shulman, 2002; Wolfe, 1994). These two processes often operate interactively, forming a dynamic attentional landscape shaped both by salient features and by cognitive intent. Corbetta and Shulman (2002) proposed a dual network model, highlighting the distinct neural circuits supporting endogenous control, while Wolfe (1994) developed the Guided

Search model, which formalised the interaction between stimulus-driven and goal-oriented processes. In ecologically valid environments, attentional allocation is further modulated by semantic content, task relevance and spatial layout (Tatler *et al.*, 2011). At the neural level, Kastner and Ungerleider (2000) demonstrated that higher cortical areas modulate early sensory processing, while Katsuki and Constantinidis (2014) identified overlapping functions between exogenous and endogenous pathways. Kingstone *et al.* (2003) reinforced that real-world attention reflects both physical features and cognitive expectations.

Over the past decades, these VAT-based theoretical models have evolved into empirically validated paradigms, particularly through eye-tracking research. The operationalisation of bottom-up and top-down attention has revealed how visual salience and cognitive relevance interact in complex viewing scenarios (Rayner, 1998; Corbetta & Shulman, 2002; Tatler *et al.*, 2011). These principles have also been translated into computational approaches, most notably biologically plausible saliency models (Itti & Koch, 2001), which simulate how low-level visual features compete for attention. More recently, data-driven methods, such as machine learning and adaptive saliency algorithms, have been applied to model predicted scan paths in dynamic digital interfaces (Kümmerer *et al.*, 2022; Chakraborty *et al.*, 2023). This integration of theory, empirical data and computational modelling underpins the use of simulated visual attention as a transparent and replicable tool for exploring attentional structures in complex digital environments.

In tourism communication, simulation-based attention models offer a method to anticipate predicted attentional allocation to destination imagery. Website layouts are strategically designed to capture and guide attention, shaping early impressions and persuasive narratives. Simulation-based gaze prediction provides a theory-informed alternative to empirical eye-tracking, particularly when research is constrained by technical, ethical or logistical limitations (Köchling, 2020; Lin & Lin, 2014).

By leveraging perceptual salience (Itti & Koch, 2001) and predictive coding (Friston, 2010; Spratling, 2017), saliency-based simulations allow researchers to estimate likely simulated scan paths and predicted visual hotspots before collecting empirical data. Borji, Parks and Itti (2014) demonstrated the validity of such approaches through comparative evaluations of classical and adaptive saliency algorithms, while Voßkübler *et al.* (2008) developed OGAMA (Open Gaze and Mouse Analyzer) as a user-friendly platform for simulated visual attention analysis. In tourism research, where image-based persuasion is central, these models support interface design assessment and hypothesis-driven exploration of how digital layouts guide attention and influence destination image formation (Echtner & Ritchie, 1993; Cardoso *et al.*, 2022).

2.2. Perceptual organisation and feature-based integration

Gestalt Theory explains how the visual system structures input into coherent wholes, rather than processing isolated elements. Core principles, including proximity, similarity, continuity and figure-ground distinction, govern how the visual field is segmented and prioritised (Wagemans *et al.*, 2012). Pomerantz (2006) demonstrated how grouping and colour influence perceptual salience and figure-ground organisation, supporting early automatic structuring of complex scenes.

This perceptual organisation is particularly relevant to digital environments, where layout, spacing and visual balance strongly influence early attention. Todorović (1997, 2008, 2020) examined how figure-ground assignment, perceptual discontinuities and spatial context affect attentional prioritisation. Historically, Metzger (1928) helped establish the scientific foundations of Gestalt Theory, which continues to shape contemporary visual cognition research. Computational extensions of these principles have also emerged; for example, Desolneux, Moisan and Morel (2007) proposed a

probabilistic model to estimate perceptual structure in static images, offering a formal bridge between theoretical psychology and digital layout analysis.

Beyond perceptual grouping, simulated visual attention also depends on the integration of feature-based information. Feature Integration Theory (FIT), introduced by Treisman and Gelade (1980), proposes that basic low-level visual features (colour, luminance intensity and orientation) are processed in parallel and independently, but must be bound through focused attention to form coherent visual objects. Without this attentional binding, incorrect feature combinations may occur across elements (Treisman & Schmidt, 1982). Over time, FIT has been refined to reflect the dynamic and context-dependent nature of attention. Quinlan (2003) highlighted its limitations in complex tasks, while Kristjánsson and Egeth (2020) and Humphreys (2016) showed that top-down factors, prior knowledge and learned expectations guide feature integration and confirm relevance. Empirical research has further demonstrated how attention resolves conflicts and prioritises dimensions according to task goals and contextual meaning (Liesefeld & Müller, 2019; Zhang *et al.*, 2012).

In dense digital interfaces, such as tourism promotion websites, feature-based attention explains how visual elements are organised and integrated into coherent destination representations. Rather than responding only to what is visually prominent (bottom-up, exogenous), attentional allocation is shaped by contextual relevance and expectations (top-down, endogenous), supporting the organisation of content into meaningful structures (Treisman & Schmidt, 1982; Quinlan, 2003; Tatler *et al.*, 2011; Krippendorff, 2018; Kristjánsson & Egeth, 2020). Computational saliency models combine low-level features such as colour, luminance intensity and orientation into a single integrated saliency map. This map represents predicted visual attention, reflecting the feature-binding process described by FIT, without the need to analyse each visual channel separately.

Beyond their perceptual basis, these mechanisms have direct implications for destination marketing and destination image formation. Because online travel platforms are often the first point of contact between a destination and prospective visitors, the way visual attention is captured and guided across an interface conditions which attributes become salient and, ultimately, how the destination image is constructed in the mind of the user (Echtner and Ritchie, 1993; Cardoso *et al.*, 2022). Visual attention research therefore offers DMOs a lens to understand not only whether an interface is usable, but how its visual hierarchy shapes the affective and cognitive associations that underpin destination branding and competitiveness (Lee and Gretzel, 2012). Anticipating these attentional dynamics at the design stage allows the strategic alignment of functional and identity-related cues with the intended brand narrative, reinforcing the role of visual communication in early destination image formation.

3. Methodology

3.1 *Modelling visual attention: Computational framework and justification*

This study follows an exploratory, qualitative and simulation-based approach to model visual attention in digital tourism environments. No empirical eye-tracking data were collected. Instead, the analysis relied on OGAMA (Open Gaze and Mouse Analyzer), which generates three outputs: predicted saliency heatmaps, simulated scan paths and simulated timestamps. Predicted saliency heatmaps show the spatial distribution of likely attention. Simulated scan paths represent the sequential progression of modelled fixations, and simulated timestamps allow the derivation of saccadic timing metrics. These simulation-based outputs serve as proxies for bottom-up attention and do not represent empirical gaze data (Voßkühler *et al.*, 2008; Borji, Parks & Itti, 2014).

To address the research question, the study pursued the following specific objectives:

Analyse predicted saliency heatmaps to identify attentional hotspots and describe the spatial distribution of simulated visual salience in the tourism website interface. Examine simulated scan paths

to characterise the sequential progression of attention between functional and aesthetic elements. Assess saccadic timing metrics derived from simulated timestamps to interpret the temporal rhythm of simulated visual exploration, distinguishing rapid exogenous from slower endogenous attention. Integrate the spatial, sequential and temporal layers to produce an interpretative model of simulated visual attention in digital tourism interfaces.

OGAMA is grounded in the classical saliency model proposed by Itti, Koch & Niebur (1998) and refined by Itti and Koch (2001). The model extracts three low-level visual features, which are colour, luminance intensity and orientation, and integrates them into a composite saliency map. Two mechanisms govern the simulated gaze. Winner-Take-All (WTA) iteratively selects the most salient region, and Inhibition-of-Return (IOR) suppresses previously attended areas to encourage the exploration of new regions. Through these processes, OGAMA produces simulated fixations, scan paths and timestamps that provide a structured approximation of bottom-up visual attention (Voßkühler *et al.*, 2008; Holmqvist *et al.*, 2011). This complete explanation of the modelling framework is provided here at the start of the methodology and will not be repeated in later sections.

In its standard configuration, OGAMA produces saliency heatmaps that highlight visually prominent areas based solely on low-level features. The three visual channels are computed, normalised and combined into a single composite heatmap, where brighter areas indicate higher predicted saliency. These heatmaps are purely bottom-up and do not reveal the sequence of attention or the semantic meaning of the highlighted elements (Borji, Parks & Itti, 2014). Assigning meaning is part of the qualitative interpretation conducted by the researcher, for example identifying whether a highlighted area corresponds to a beach, a logo or a booking button. Figure 1 shows how OGAMA generates saliency heatmaps from the three visual channels.

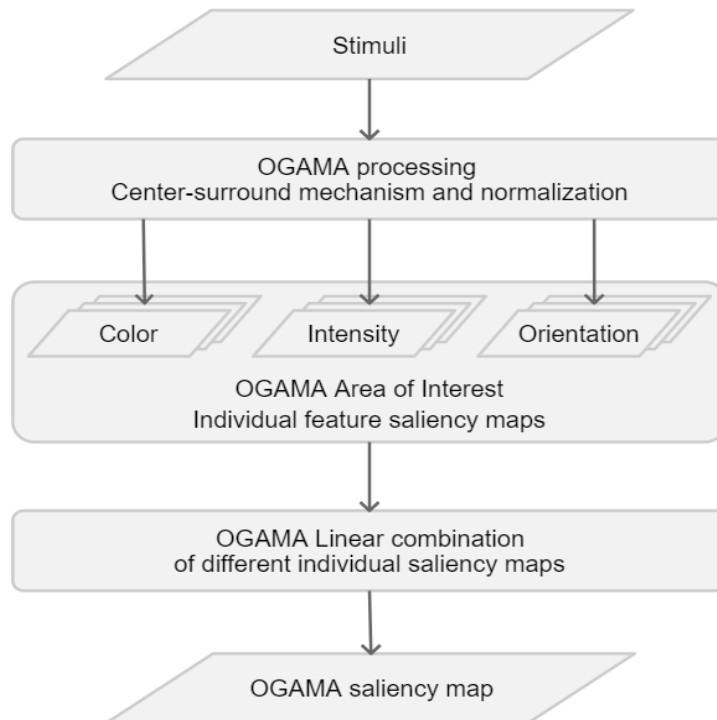


Figure 1: Generation of OGAMA saliency maps.
Source: Adapted from Voßkühler *et al.* (2008).

OGAMA, an open-source tool grounded in visual attention theory and cognitive neuroscience (Voßkühler *et al.*, 2008) was selected because it is grounded in classical bottom-up visual attention theory and produces transparent, rule-based outputs. This makes it particularly suitable for theory-driven research in tourism interfaces. In contrast, DeepGaze relies on deep neural networks to predict fixation points but does not generate scan paths and offers limited theoretical transparency. AWS (Attention based on Weighted Saliency) is another bottom-up computational model that estimates static saliency from low-level features such as intensity, colour and orientation (Garcia Diaz *et al.*, 2012). Compared with these alternatives, OGAMA provides predicted saliency heatmaps, simulated scan paths and simulated timestamps, making it appropriate for the layered methodology adopted in this study, where simulation-based outputs are integrated with higher-level qualitative interpretation (Tatler *et al.*, 2011; Holmqvist *et al.*, 2011).

With this theoretical foundation, the study advanced through six sequential stages to simulate visual attention and interpret the predicted gaze strategies across the ten islands of the Cabo Verde tourism website.

3.2 Procedural overview and implementation

The analytical design of this study followed a simulation-based approach that combined qualitative interpretation with descriptive quantitative components to capture how simulated visual attention unfolds in digital tourism environments. Rather than collecting empirical eye-tracking data, the study relied on OGAMA to generate simulation-based outputs, which included predicted saliency heatmaps, simulated scan paths and simulated timestamps (Voßkühler *et al.*, 2008; Itti & Koch, 2001). These outputs provided the spatial, sequential and temporal layers required to analyse bottom-up visual attention and to construct an integrated interpretation of simulated visual attention.

The methodology was organised into six sequential stages, summarised in Table 1, to identify which interface elements were most likely to attract attention and how simulated visual exploration evolved over time. The spatial dimension relied on predicted saliency heatmaps and the deductive content analysis (Cardoso *et al.*, 2019) of simulated scan paths to identify focal elements, exploration flows and perceptual priorities within each island webpage. The temporal dimension used simulated timestamps to calculate saccadic timing metrics, offering insights into scanning rhythm, attentional stability and the interpretation of simulated attentional patterns (Friston, 2010; Tatler *et al.*, 2011; Wolfe, 1994).

3.2.1. Stage 1 – Stimuli selection

Stage 1 involved the selection of ten promotional pages from the “Explore the Islands” section of the official Cabo Verde tourism website (<https://www.visit-caboverde.com/en>), captured on 15 September 2024. The layout remained constant across all pages, with only the central island image changing, ensuring that the destination visuals were the sole drivers of simulated attentional allocation. Screenshots were captured in full-screen desktop mode to maintain consistent resolution, composition and spatial hierarchy across all stimuli. This website was selected because its homepage provides direct access to the ten islands through a consistent top navigation bar, offering a clear and standardised layout that represents the official imagery of Cabo Verde.

The use of these images was formally authorised by the rights holder following a written request submitted on 9th September, 2024. Their inclusion complies with the Portuguese Código do Direito de Autor e dos Direitos Conexos (CDADC), the Berne Convention for the Protection of Literary and Artistic Works, and Regulation (EU) 2016/679 of the European Parliament and of the Council, ensuring full adherence to intellectual property and research ethics.

Table 1: Methodological sequence and analytical correspondence.

Stage	Procedure	Operationalisation	Type of Analysis
1 – Stimulus selection	Capture of ten promotional pages (“Explore the Islands”) from the official Cabo Verde tourism website.	Full-screen desktop mode, captured on 15 September 2024. Layout kept constant across pages, with only the central island image varying. Images used under formal authorisation and in compliance with copyright and data protection regulations.	-
2 – OGAMA simulation setup	Upload of stimuli to OGAMA in simulation mode.	Screenshots configured as individual stimuli; saliency module activated using Itti & Koch (2001) model. Generation of predicted saliency heatmaps and simulated scan paths with timestamps. Only integrated (composite) maps used, excluding channel-specific outputs.	-
3 – Predicted saliency heatmap analysis	Visual inspection of integrated heatmaps.	Analysis of 10 composite saliency heatmaps to identify focal elements and predicted attention distribution across the interface.	Qualitative
4 – Simulated scan path analysis	Qualitative and descriptive analysis of simulated scan path sequences .	Ten simulated scan paths per island were exported and coded into five predefined analytical categories (Natural Landscape, Built Elements, Branding, Interactive/Call to Action (CTA), and Lighting & Contrast) following a deductive content analysis approach. Category frequencies were calculated to identify dominance and temporal distribution patterns.	Qualitative with descriptive quantitative support
5 – Saccadic timing analysis	Temporal analysis of simulated scan paths with content-based interpretation of simulated temporal dynamics .	Simulated fixation timestamps were extracted from OGAMA to calculate saccade intervals as the temporal difference between consecutive fixations. Descriptive metrics included mean interval, standard deviation, and shortest and longest saccades, which were interpreted in relation to the visual categories defined in the deductive content analysis.	Quantitative with content-based interpretation
6 – Integrated interpretation framework	Triangulation of spatial (heatmaps), sequential (simulated scan paths) and temporal (saccadic timing) outputs.	Integration of these three layers to construct a coherent narrative of simulated visual attention, distinguishing patterns associated with exogenous (bottom-up) and endogenous (top-down) attention. This framework is guided by Visual Attention Theory, Gestalt Theory, Feature Integration Theory and complemented by Predictive Coding, explaining how predicted attention supports destination imagery formation without empirical eye-tracking data.	Triangulated (simulation-based)

3.2.2. Stage 2 – OGAMA simulation setup

Stage 2 prepared the visual stimuli in OGAMA to obtain simulation-based outputs for the visual attention analysis. Each screenshot from Stage 1 was processed in simulation mode to produce predicted saliency heatmaps, simulated scan paths and simulated timestamps for every island webpage. Only composite saliency maps were retained, excluding channel-specific outputs to ensure a consistent interpretative framework. This stage provided the complete spatial, sequential and temporal layers that formed the basis for the subsequent analyses.

3.2.3. Stage 3 – Predicted saliency heatmap analysis

Stage 3 focused on the qualitative analysis of predicted saliency heatmaps generated in OGAMA. The analysis was restricted to ten composite heatmaps that integrated colour, luminance intensity and orientation, representing the spatial distribution of simulated visual attention. Warmer colours indicated areas with a higher probability of early fixations, while cooler tones corresponded to lower predicted salience.

Predicted heatmaps served as the foundational spatial layer for simulation, as spatial saliency alone is often insufficient to explain attentional patterns in complex digital environments such as online travel

platforms (Henderson & Hayes, 2018; Tatler *et al.*, 2011). To extend beyond static predictions, the analysis focused on composite attention maps that integrate spatial saliency with simulated scan paths, allowing the approximation of how simulated attention unfolds across the interface.

The interpretation of heatmaps followed three main steps: (1) High-saliency regions were identified by reading the standard colour gradient. Red and yellow signalled high attentional saliency and predicted hotspots likely to attract early fixations. Green indicated intermediate attention, typically associated with secondary areas of interest; blue represented low attentional allocation in peripheral or background regions, and purple denoted minimal or negligible attention. This interpretation aligns with the principles of bottom-up visual attention (Itti *et al.*, 1998; Tatler *et al.*, 2011); (2) The distribution and spread of attention were assessed to determine whether visual focus was concentrated in key points or dispersed across the scene, providing insights into simulated exploration patterns (Tatler *et al.*, 2011); (3) Finally, heatmap patterns were integrated into the simulation-based analysis of visual attention. They were examined to identify initial attentional capture, likely scan path flows and neglected regions, providing a holistic view of simulated gaze allocation in line with computational attention frameworks (Holmqvist *et al.*, 2011).

This systematic approach ensured that Stage 3 produced a robust spatial baseline for the Stage 4 scan path analysis, where sequential visual trajectories would be assessed to capture the dynamic component of simulated attention.

3.2.4. Stage 4 – Simulated scan path analysis

Stage 4 combined a qualitative interpretation of simulated scan paths with descriptive quantitative support to capture the sequential progression of simulated visual attention. OGAMA generated ten simulated fixations per stimulus, automatically numbered in chronological order to form a modelled visual trajectory for each island. These fixations were exported in long table format, where each row corresponded to a single visual event defined by the island, the fixation position from 1 to 10 and the associated visual element. This structure allowed the reconstruction of predicted scan path progression across the tourism interface, offering a detailed view of how bottom-up saliency interacts with the semantic and functional structure of the page.

Each visual element was coded into five analytical categories: Natural Landscape, Built Elements, Branding, Interactive/Call to Action (CTA), and Lighting & Contrast. The coding process followed a deductive content analysis approach (Cardoso *et al.*, 2019), where predefined categories guided the systematic classification of visual elements into a structured matrix. NVivo 14 (QSR International) was then used to generate a matrix query that calculated the frequency of each category across the ten simulated fixations per island. This step introduced a descriptive quantitative layer, enabling the identification of category dominance and the temporal dynamics of the simulated scan paths. By examining each fixation individually, the analysis described how simulated attention progresses across functional and aesthetic elements within the interface (Friston, 2010; Engel, Zhang & Wandell, 1997).

To ensure the reliability of manual coding, four independent researchers classified all visual elements into the five analytical categories. Inter-coder reliability was assessed using Krippendorff's Alpha for nominal data (Hayes & Krippendorff, 2007), calculated with the K Alpha Calculator (Marzi, Balzano & Marchiori, 2024), resulting in $\alpha = 0.84$, which indicates high reliability (Krippendorff, 2018).

This stage provided the sequential layer of the simulation, establishing the temporal order of visual attention and linking bottom-up saliency with the semantic structure of the interface. It served as the bridge to stage 5, where simulated timestamps were used to compute saccadic timing metrics.

3.2.5. Stage 5 – Saccadic timing analysis

Stage 5 focused on the processing of simulated saccade data generated by OGAMA (Open Gaze and Mouse Analyzer), an open-source tool grounded in visual attention theory and cognitive neuroscience (Voßkühler *et al.*, 2008). OGAMA implements the saliency-based model of Itti, Koch & Niebur (1998), later refined by Itti and Koch (2001), which predicts visual exploration by integrating three low-level visual attributes: colour (Engel, Zhang & Wandell, 1997), luminance intensity (Leventhal, 1991) and orientation (Greenspan *et al.*, 1994). These features are combined in a dynamic saliency map that simulates the competition of visual stimuli for attentional selection.

The model operates through two key mechanisms. WTA selects the most salient location to generate a simulated fixation, while IOR suppresses the previously attended area to encourage exploration across new regions (Itti & Koch, 2001; Itti, Koch & Niebur, 1998). Through this iterative process, the model produces sequential scan paths that approximate saccadic exploration. In parallel, the progression from short, stimulus-driven fixations to longer and more variable intervals can be interpreted as reflecting a shift from exogenous to endogenous attentional patterns (Friston, 2010).

For each of the ten island layouts of the Cabo Verde tourism website, OGAMA exported a sequence of simulated fixation timestamps in milliseconds, corresponding to the integrated saliency map. In the simulated context, these timestamps indicate the virtual onset time of each modelled fixation, functioning as a virtual clock for the predicted scan path. Accordingly, these timestamps are expressed in the model's internal time units and are not equivalent to human fixation durations. Unlike empirical eye-tracking data, which reflects real gaze behaviour and individual variability, these model-driven timestamps are stable predictions that serve as structured proxies for the rhythm and spatio-temporal distribution of bottom-up attention. Each saccade corresponded to the modelled jump between two consecutive fixations, calculated as the temporal difference between their timestamps using the standard formula:

$$\text{Saccade Interval}_i = t_i - t_{i-1}$$

This method aligns with standard procedures in eye-tracking research, where saccade intervals are defined as the temporal spacing between fixations, reflecting scanning rhythm and cognitive load (Duchowski, 2007; Holmqvist *et al.*, 2011). This distinction is essential to avoid misinterpreting cumulative fixation times as temporal dynamics. The approach is consistent with established practices in visual attention modelling (Rayner, 1998).

Short and stable intervals were interpreted as patterns associated with automatic, stimulus-driven exploration, whereas longer and more variable intervals were interpreted as patterns associated with more structured and context-dependent attentional allocation (Kristjánsson & Egeth, 2020; Tatler *et al.*, 2011). Analysing the rhythm of these intervals provides a structured proxy for temporal variation, allowing the temporal layer to complement spatial and sequential analyses. In this study, measuring the rhythm was essential to identify shifts between rapid, stimulus-driven viewing and slower, content-driven processing, thereby linking simulated temporal dynamics to the interpretative framework developed in Stage 6.

The first interval in each sequence was set as missing, since no saccade precedes the initial fixation. IBM SPSS Statistics 29 was then used to compute descriptive metrics for each island, including mean saccade interval (ms), standard deviation (SD, ms), and shortest and longest intervals (ms). These measures were selected to capture the central tendency, variability, and extreme values of the simulated saccadic rhythm, which are established indicators of temporal structure in visual attention analysis (Holmqvist *et al.*, 2011; Rayner, 1998). Analysing these parameters enables the identification of consistent patterns

and contrasts across islands, and extreme values help detect shifts between different attentional regimes (Tatler *et al.*, 2011). By quantifying these temporal characteristics, the analysis provided an objective basis for linking differences in scanning rhythm to the spatial and sequential findings, thereby strengthening the integrated interpretation developed in Stage 6.

3.2.6. Stage 6 – Integrated interpretation of simulated visual attention

This stage synthesised the outputs from the previous stages into a coherent analytical model that integrates the spatial, sequential and temporal layers of analysis. This integration provides a structured representation of how predicted visual attention develops across the ten island layouts of the Cabo Verde website, showing patterns associated with a progression from bottom-up (exogenous) capture to top-down (endogenous) inspection. The Simulated Visual Attention Integrative (SVAI) Model, presented in Figure 2, depicts this process by combining the key analytical steps and their theoretical foundations into a single interpretative framework.

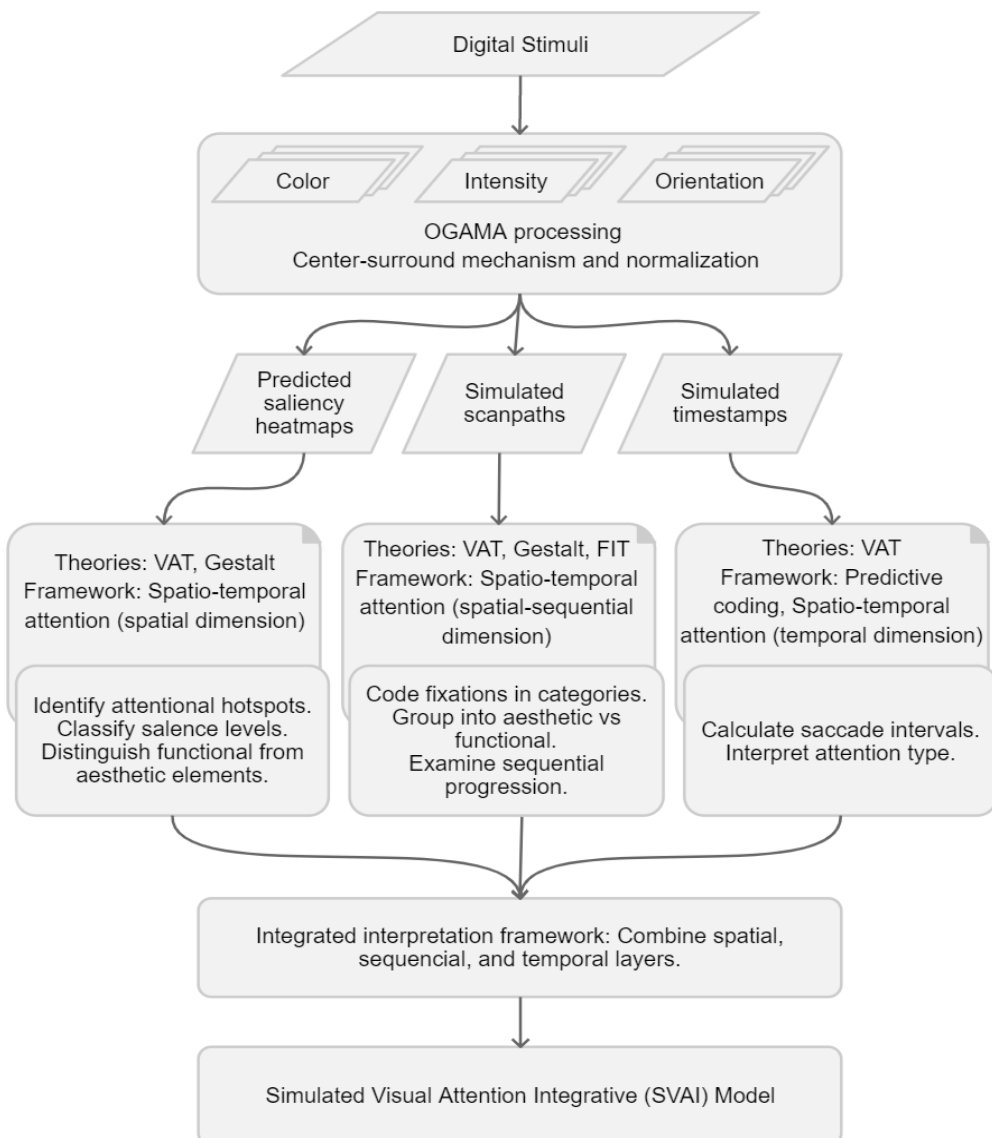


Figure 2: SVAI Model for the analysis of simulated visual attention in online travel platforms.

The first step of the proposed SVAI Model is the Predicted Saliency Heatmap Analysis. The standard colour gradient is interpreted to identify areas of high, medium, low and minimal salience, which are then classified as functional or aesthetic elements. High salience (red and yellow) indicates predicted hotspots with a high probability of attracting early fixations, medium salience (green) reflects secondary areas of interest, low salience (blue) corresponds to peripheral or background regions, and minimal salience (purple) marks negligible attention. This step captures the starting point of bottom-up (exogenous) attention, in which predicted attentional allocation is directed towards visually prominent elements. VAT explains that high-salience regions are more likely to attract initial fixations (Corbetta & Shulman, 2002; Itti, Koch & Niebur, 1998). Gestalt Theory clarifies how grouping principles such as proximity, continuity and figure-ground organisation influence early attentional structuring (Wagemans *et al.*, 2012). The Spatio-Temporal Attention (STA) framework is applied here to the spatial dimension (Engel, Zhang & Wandell, 1997; Tatler *et al.*, 2011).

The second step is the Simulated Scan path Analysis. Each predicted fixation is coded into one of five predefined categories: Natural Landscape, Built Elements, Branding, Interactive or Call to Action, and Lighting and Contrast. These categories are further grouped into functional and aesthetic elements. The sequence shows that early fixations often target functional elements, indicating patterns consistent with bottom-up (exogenous) capture, while later fixations progressively shift towards aesthetic or semantically rich elements, suggesting the involvement of top-down (endogenous) attentional processes. FIT explains how visual features such as colour, intensity and orientation are combined into coherent fixation sequences (Treisman & Gelade, 1980). VAT frames the shift from exogenous to endogenous control (Corbetta & Shulman, 2002), while Gestalt Theory accounts for smooth scan path transitions based on continuity and proximity (Wagemans *et al.*, 2012). The STA framework is applied here to the spatial-sequential dimension, linking the position of visual elements in the interface with the predicted order in which they are attended. This connection supports the interpretation of how simulated attention transitions across elements, combining spatial distribution with the sequential flow of attention (Engel, Zhang & Wandell, 1997; Tatler *et al.*, 2011).

The third step is the Saccadic Timing Analysis. Simulated timestamps from OGAMA are used to calculate the mean, standard deviation (SD), minimum and maximum intervals between consecutive fixations. These values measure the rhythm of predicted gaze and support the classification of attention type. The mean saccade interval reflects the average scanning rhythm, with lower means indicating rapid bottom-up (exogenous) attention capture and higher means suggesting slower, semantically oriented top-down (endogenous) exploration.

The standard deviation (SD) shows the stability or variability of attentional shifts: low SD values point to consistent exogenous scanning, while high SD values indicate transitions towards endogenous inspection and more structured attentional allocation. The shortest and longest intervals define the range of attentional patterns, where very short intervals correspond to rapid stimulus-driven shifts and longer intervals indicate extended inspection of visually salient elements. The interpretative logic follows that short and stable intervals, combining a low mean with a low SD, characterise patterns associated with exogenous, stimulus-driven exploration, whereas long and variable intervals, combining a high mean with a high SD, indicate patterns associated with more structured and context-dependent attentional allocation. From a theoretical perspective, VAT supports the classification of attentional patterns (Corbetta & Shulman, 2002), while Predictive Coding explains the progression from rapid exogenous capture to slower endogenous confirmatory processes (Friston, 2010; Kristjánsson & Egeth, 2020; Rayner, 1998; Tatler *et al.*, 2011).

The final step is the Integrated Interpretation Framework. In this stage the spatial, sequential and temporal layers are combined to produce a single, coherent narrative of simulated visual attention. The

integration begins by cross-checking the spatial hotspots from the heatmaps with the first fixations in the simulated scan paths to identify patterns consistent with initial bottom-up (exogenous) capture. Sequential trajectories are then overlaid with the saccadic timing metrics to identify where simulated attention remains rapid and stimulus-driven and where it becomes slower and more structured.

This integrated reading reconstructs the predicted visual pathway: attention enters the page through functional elements that trigger exogenous capture, moves towards aesthetic or semantically rich areas under endogenous guidance, and finally slows for prolonged inspection of visually meaningful content. VAT supports the classification of attentional patterns (Corbetta & Shulman, 2002), FIT explains coherent scan path trajectories (Treisman & Gelade, 1980), and Gestalt Theory justifies spatial grouping and scan path continuity (Wagemans *et al.*, 2012). Predictive Coding accounts for the transition between exogenous and endogenous attentional processes (Friston, 2010; Kristjánsson & Egeth, 2020). The STA framework is applied here to the integration process, connecting where attention begins, how it progresses and how temporal rhythm varies, thus reconstructing the complete pathway of simulated visual attention (Engel, Zhang & Wandell, 1997; Tatler *et al.*, 2011).

4. Results and discussion

4.1. Predicted heatmap analysis

The predicted saliency heatmaps (Figure 3) of the ten Cabo Verde Island layouts reveal a consistent attentional structure aligned with the spatial dimension of the SVAI Model.

High salience (red and yellow) consistently fell on the Discover Island Button CTA and island name, indicating patterns consistent with early exogenous capture (Corbetta & Shulman, 2002; Itti & Koch, 2001). These elements were the most salient in about 90% of simulated sequences, reinforcing their role in the functional visual hierarchy. Their later transition towards semantically rich features suggests the involvement of endogenous attentional processes (Friston, 2010; Wolfe, 1994), in line with Tatler *et al.* (2011), who describe how predicted fixations shift from reactive to goal-directed patterns as simulated attention progresses towards semantically rich content.

Intermediate salience (green) appeared on natural and built features such as volcanic peaks, dunes and coastal rocks. In Boa Vista and Santa Luzia, the island name in white lettering also attracted secondary attention. Gestalt principles explain how the CTA and island name operate as a coherent attentional cluster (Pomerantz, 2006; Wagemans *et al.*, 2012). Yet salience alone does not ensure functional integration (Treisman & Gelade, 1980), and initial capture does not imply deep processing unless supported by contextual relevance (Rayner, 1998).

Low salience (blue) extended across peripheral areas of sky, sea and sand, with the national logo also falling consistently in this category. Minimal salience (purple) marked uniform background regions with negligible impact. These findings support evidence that isolated or irrelevant features fail to guarantee capture (Henderson & Hayes, 2018; Liesefeld & Müller, 2019) and that luminance intensity, when not linked to task relevance, can generate noise rather than guide perception (Holmqvist *et al.*, 2012; Wagemans *et al.*, 2012). Across layouts, common patterns included the dominance of CTA and island name, the presence of green salience on secondary features, and wide peripheral blue zones that stabilised the visual field. Differences lay in the type of secondary features (volcanic peaks in Fogo, São Vicente and Santa Luzia versus dunes and coastal formations in Sal, Maio and Brava), in distribution (concentrated in Fogo and Santa Luzia versus diffuse in Boa Vista and Santiago), and in background dominance, which was greater in open islands such as Sal, Boa Vista and Maio than in rugged ones like Fogo and São Vicente.



Figure 3: OGAMA predicted saliency heatmaps of Cabo Verde's 10 islands.

The interpretation of these results confirms the complementary role of different theoretical perspectives. VAT explains that exogenous attention is first drawn to perceptually salient elements such as the CTA and island name (Corbetta & Shulman, 2002; Itti & Koch, 2001). The early binding of shape, colour and spatial position into coherent functional objects aligns with Feature Integration Theory (Treisman & Gelade, 1980). Gestalt Theory clarifies why grouped and centrally located elements act as unified attentional entry points through principles of proximity, figure-ground separation and continuity (Pomerantz, 2006; Wagemans *et al.*, 2012). Layouts with stronger Gestalt coherence also support more efficient attentional organisation and predictability, reinforcing their effectiveness in digital tourism communication (Holmqvist *et al.*, 2012). The extensive blue and purple zones confirm the stabilising role of peripheral areas lacking contrast or grouping. This pattern illustrates the predictive nature of bottom-up simulated attention, where only structured and salient elements guide the initial scan path. From a predictive coding perspective, it reflects the interplay between visual salience and functional relevance, shaped by expectations and interface structure (Friston, 2010).

This spatial layer defines the baseline of the SVAI Model, identifying the hotspots that are associated with exogenous capture patterns and the peripheral zones that stabilise the hierarchy, providing the foundation for the sequential and temporal analyses that follow.

4.2. Simulated scan path analysis

Table 2 presents the distribution of visual elements by category along the ten simulated scan path points, revealing a clear sequential progression of simulated attention across the ten island layouts of the Cabo Verde website.

The first stage of the simulated scan path, corresponding to Saliency Points (SP) 1 to 3, is entirely monopolised by Branding (100%), as the initial simulated fixations consistently fall on the island name and the Discover Island Button CTA. This pattern indicates a concentration of predicted attentional allocation on textual and identity cues in the initial stage of the scan path, reflecting the exogenous salience embedded in the interface design. This is consistent with the predominance of exogenous attention associated with low-level visual feature salience (Itti & Koch, 2001) and aligns with the progression towards endogenous attention described by Wolfe (1994), Friston (2010) and Tatler *et al.* (2011).

Table 2. *Distribution of elements by category along the predictive scan path.*

	Natural Landscape (%)	Built Elements (%)	Branding (%)	Interactive / CTA (%)	Lighting & Contrast (%)
1	0.0	0.0	100.0	0.0	0.0
2	0.0	0.0	100.0	0.0	0.0
3	0.0	0.0	100.0	0.0	0.0
4	30.0	0.0	70.0	0.0	0.0
5	30.0	10.0	60.0	0.0	0.0
6	60.0	0.0	40.0	0.0	0.0
7	90.0	0.0	0.0	0.0	10.0
8	100.0	0.0	0.0	0.0	0.0
9	80.0	10.0	0.0	0.0	10.0
10	80.0	0.0	10.0	10.0	0.0

Note: SP = Saliency Point (predicted fixation point along the simulated scan path).

A gradual transition towards the natural environment emerges between SP4 and SP6. At SP4, Natural Landscape rises to 30%, competing with Branding at 70%. By SP5, the first Built Elements appear at 10% (dinghy and wooden jetty bordering the sea), while Natural Landscape remains at 30% and Branding declines to 60%. SP6 consolidates this transition, with Natural Landscape reaching 60% and Branding decreasing to 40%, indicating a shift in simulated attentional allocation from functional to scenic elements. This stage also illustrates that chromatic salience alone does not guarantee functional integration, as initial capture by visually striking natural elements requires contextual relevance to sustain attention (Rayner, 1998; Treisman & Gelade, 1980).

The scenic exploration phase dominates SP7 and SP8, where Natural Landscape accounts for 90% and 100%, respectively. Lighting & Contrast begins to participate in this stage, with 10% at SP7, exemplified by elements such as sunset light zone and light transition on tree branches. This phase corresponds to a stage in which simulated attention is predominantly allocated to landscape elements, reflecting a transition from functional cues to visually rich regions. Consistent with prior evidence, luminance intensity may increase visual noise and attentional dispersion when not anchored to task relevance (Holmqvist, Nyström & Mulvey, 2012; Wagemans *et al.*, 2012).

The final stage of the simulated scan path (SP9 and SP10) exhibits a mixed profile in which Natural Landscape remains predominant (80%), while Built Elements and Interactive/CTA cues reappear in peripheral positions (10% each). Lighting & Contrast also recurs at SP9, suggesting that luminance-based features help guide the last moments of visual exploration. This late phase indicates a redistribution of simulated attentional allocation across both scenic and functional elements, maintaining attention within visually salient regions (Andersson, Larsson & Holmqvist, 2010; Holmqvist *et al.*, 2012; Tatler *et al.*, 2011).

4.3. Saccadic timing analysis

Table 3 presents the saccadic timing metrics for the ten island layouts of the Cabo Verde website, providing a temporal perspective on simulated visual exploration. The analysis is based on the mean interval, standard deviation (SD), and the shortest and longest simulated saccades, which together form a proxy for the rhythm of attentional shifts across the interface.

The mean saccade intervals range from 60.86 ms (São Vicente) to 66.94 ms (Boa Vista). Lower mean intervals, as observed in São Vicente and Sal (62.30 ms), are associated with rapid stimulus-driven scanning patterns, while higher means, such as those of Boa Vista (66.94 ms) and Brava (65.91 ms), are associated with slower and more structured attentional allocation patterns.

Table 3. *Saccadic timing metrics for the 10 island layouts.*

Island	Mean (ms)	SD (ms)	Shortest Saccade (ms)	Longest Saccade (ms)
Santo Antão	64.52	12.60	43.8	78.0
São Vicente	60.86	6.50	47.2	67.9
São Nicolau	63.43	11.06	45.2	73.3
Sal	62.30	7.63	42.1	67.7
Boa Vista	66.94	15.37	43.8	83.4
Maio	63.68	10.55	44.2	77.7
Santiago	63.86	9.95	45.1	78.8
Fogo	64.45	11.03	44.3	78.0
Brava	65.91	13.43	46.0	83.0
Santa Luzia	65.94	11.92	47.6	78.4

The standard deviation of saccade intervals provides insight into the stability of simulated attention over time. São Vicente (6.50 ms) and Sal (7.63 ms) show the lowest variability, which indicates a highly consistent scanning rhythm typical of exogenous capture where attention is tightly driven by stimulus features. In contrast, Boa Vista (15.37 ms) and Brava (13.43 ms) present the highest variability, indicating more variable simulated attentional patterns that are consistent with transitions towards endogenous attentional processes.

The shortest saccades cluster around 42–47 ms, illustrating moments of rapid stimulus-driven shifts typical of exogenous capture. As Rayner (1998) notes, such short, reactive saccades are often guided by local contrast rather than semantic content, whereas longer and more stable fixations tend to reflect goal-directed processes (Andersson, Larsson & Holmqvist, 2010). The longest saccades, ranging from 67.7 ms (Sal) to 83.4 ms (Boa Vista), indicate extended intervals that can be interpreted as reflecting sustained allocation of simulated attention to visually salient regions.

Taken together, this temporal profile aligns with VAT and Predictive Coding. The progression from short and stable intervals to longer and more variable ones is consistent with a transition from exogenous to endogenous attentional patterns (Kristjánsson & Egeth, 2020; Rayner, 1998; Tatler *et al.*, 2011). This is consistent with the predictive coding view that attention evolves from stimulus-driven scanning to confirmatory exploration of scene elements (Friston, 2010; Tatler *et al.*, 2011). Stability and precision in saccade patterns, as observed in islands with consistent scanning rhythms, are also recognised as valid proxies for assessing interface effectiveness, even in simulation-based studies (Holmqvist, Nyström & Mulvey, 2012). It should be emphasised that this shift towards endogenous attention is a theory-informed interpretation of the simulated outputs and not a process generated by the saliency model itself.

4.4. Findings: An integrated interpretation framework

This study integrated the spatial, sequential and temporal dimensions of simulated visual attention, achieving the objective of developing and applying a simulation-based analytical framework that combines these three predictive layers into a coherent interpretation of attentional patterns in digital tourism interfaces.

The results indicate that the SVAI Model consistently integrates these dimensions, identifying where attention begins, how it progresses and how its rhythm varies across different stages of the simulated scan path. The methodology represents an original contribution by combining predicted saliency maps, simulated scan paths and temporal metrics within a single interpretative model. This approach goes beyond counting fixation points, reconstructing the attentional path dynamically and with theoretical grounding, offering potential for application across multiple digital contexts.

The value of the model as a pre-testing tool is clear, as it allows the diagnosis of attention patterns before investing in empirical eye-tracking, providing a fast and cost-effective solution for designers and researchers. Its methodological transparency, based on the triangulation of VAT, Gestalt Theory and FIT with complementary frameworks such as Predictive Coding, ensures that results are both descriptive and interpretable within cognitive science.

The inclusion of temporal patterns through saccadic timing adds a rarely explored dimension to the literature, which typically relies on heatmaps or scan paths alone. The transition from rapid and stable scanning to slower and more variable inspection, as reflected in the simulated temporal patterns, constitutes a significant methodological contribution.

The SVAI Model also shows potential for generalisation, being applicable to other types of digital layouts, and its validity can be strengthened in future studies combining this approach with real eye-tracking data. Its differentiation from opaque approaches, such as some deep learning methods, lies in its transparency, replicability and theoretical grounding, which make it possible to understand not only the results but also the process that generates them.

Overall, the study demonstrates that visual attention analysis can combine scientific rigour with practical applicability. The SVAI Model supports the anticipation and interpretation of attentional patterns, aligning predictive metrics with meaningful interpretations for the design of digital experiences. Furthermore, it demonstrates the internal consistency of simulated attentional patterns in digital interfaces, reinforcing the relevance of classic attention theories in computational contexts. It also demonstrates the importance of transparency and interpretability at a time when many data-driven approaches lack theoretical grounding. Practically, the SVAI Model offers a fast and cost-effective pre-testing tool that can diagnose attentional patterns before investing in empirical eye-tracking. It supports the optimisation of digital tourism interfaces by identifying which elements attract immediate attention, how attention progresses across content and when functional or aesthetic features re-emerge within the simulated scan path, thereby prioritising candidate elements for subsequent empirical testing rather than prescribing final design decisions. The model is transferable to other digital contexts, such as booking platforms, cultural heritage websites or promotional campaigns, providing designers and destination managers with a transparent and replicable method to align interface design with predicted attentional patterns.

5. Theoretical and practical contributions

This study addressed the critical gap in tourism research concerning the absence of predictive models of visual attention and the fragmented treatment of spatial, sequential and temporal dimensions. By developing and applying the SVAI Model, it demonstrated that simulated saliency maps, scan paths and temporal metrics can be integrated into a coherent analytical framework that dynamically reconstructs simulated attentional patterns in digital tourism interfaces. In doing so, the study fulfils its objective of advancing a transparent, replicable and theory-driven approach to understanding predicted visual attention patterns in online travel platforms.

Theoretically, the study contributes to visual attention research by triangulating Visual Attention Theory, Gestalt Theory and Feature Integration Theory with two complementary frameworks, Predictive Coding and Holistic Attention Models. This integration extends cognitive science perspectives into the tourism domain and moves beyond reliance on isolated metrics, offering a holistic account of how simulated attention unfolds across spatial, sequential and temporal dimensions, and how these patterns can be interpreted in relation to destination image formation, persuasion and consumer decision-making (Ortanderl, 2025). This theoretical integration of multidisciplinary models in Tourism is grounded in the seminal studies of Friston (2010) on the Free Energy Principle and Predictive Coding, combined with the work of Treisman and Gelade (1980) on Feature Integration Theory (FIT), Wagemans *et al.* (2012) on Gestalt principles, and Corbetta and Shulman (2002) on Visual Attention Theory (VAT). In turn, the interdependence of spatial, sequential, and temporal layers, from a holistic perspective on visual attention, is aligned with Kristjánsson and Egeth (2020), who demonstrate how expectations and prior knowledge (top-down processes) guide the integration of stimuli. Likewise, the shift from a reactive gaze to a goal-directed focus, grounded in a dynamic understanding of image formation as revealed through temporal analysis, is supported by the research of Rayner (1998) and Tatler *et al.* (2011).

Methodologically, the study establishes simulated visual attention as a novel tool for tourism research, highlighting its capacity to anticipate attentional flows without the logistical and financial constraints

of empirical eye-tracking. The SVAI Model advances predictive modelling by introducing saccadic timing as a rarely explored temporal dimension in tourism research, showing how scanning pace shifts from functional elements to aesthetic and identity-related features. Its transparency and theoretical grounding also differentiate it from opaque machine learning approaches, ensuring results are interpretable as well as replicable.

Practically, the findings provide DMOs, digital marketers and sustainability communicators with a cost-effective pre-testing tool to diagnose attention patterns early in the design process. By aligning predictive outputs with cognitive science theories, the SVAI Model delivers actionable insights into structuring interface elements to capture attention, communicate sustainability cues and strengthen destination competitiveness. The feasibility and advantages of adopting computational models as a low-cost alternative to empirical experiments, aimed at the efficient use of resources, are supported by Voßkühler *et al.* (2008) and Borji, Parks, and Itti (2014). Within this perspective, the use of exploratory tools in the early stages of design addresses the logistical limitations of traditional eye-tracking tests, enabling the agile prototyping of promotional campaigns, as highlighted by Valtakari *et al.* (2021). Furthermore, the model supports agile prototyping by enabling designers to pre-test layouts at an early stage, ensuring that functional and aesthetic elements are strategically aligned with predicted attentional patterns before committing resources to full-scale development. Its transferability goes beyond destination websites, offering applications in booking platforms, cultural heritage sites and promotional campaigns. For European DMOs in particular, the SVAI Model provides a scalable, resource-efficient tool to benchmark and optimise the visual communication of destinations across digital channels, supporting evidence-based design decisions without the logistical constraints of empirical eye-tracking. The importance of pre-testing visual persuasion across the entire chain of the destination's digital channels is corroborated by the studies of Lee and Gretzel (2012) on mental imagery processing in tourism websites, demonstrating the model's potential for scalability and transferability.

To make these benefits concrete, consider a European DMO preparing to launch a new island or regional landing page. Before commissioning costly eye-tracking sessions or A/B tests, the DMO can run the candidate layouts through the SVAI Model and obtain, within hours, a layered diagnosis of where simulated attention is likely to land first, how it is predicted to travel across the page and where its rhythm slows. If the analysis shows that simulated attention is monopolised by a call-to-action button and the destination wordmark, while the scenic imagery that carries the brand promise is reached only late in the scan path, the DMO has an early and low-cost signal that the visual hierarchy may be working against its communication goals. The team can then reposition or restyle the hero image, adjust contrast, or re-sequence identity cues so that the elements conveying the destination's distinctive values compete earlier for attention. In this sense, the SVAI Model operates less as a usability checker and more as a brand-communication instrument: it helps DMOs verify that the visual elements most central to destination identity are also among the most likely to be attended, thereby supporting more coherent and persuasive destination branding and brand communication strategies. Looking ahead, embedding this kind of predictive, attention-aware pre-testing into the design workflow points to a broader shift in destination management, moving from evaluating digital interfaces only after publication towards anticipating and shaping how destination identities are visually communicated and perceived from the earliest stages of design.

6. Limitations and future research

Modelling approaches rooted in computational and cognitive research remain relatively underexplored within tourism studies, which have been traditionally oriented towards direct empirical methods and observational designs. A key limitation of this study is the absence of empirical validation. However, this limitation should be understood in light of the exploratory and simulation-based nature of the proposed framework, as well as the persistence of established analytical lenses in the field that are not

always fully suited to accommodate emerging approaches situated at the interface with computation and the neurosciences.

From this perspective, the absence of empirical validation represents a limitation that is consistent with the methodological positioning of this research; while also signalling a necessary inflection as new paradigms are incorporated into tourism studies. Accordingly, at this stage, the model is not intended to directly replicate observed behaviour, but rather to enable controlled experimentation and the systematic exploration of visual attention dynamics in digital tourism environments. In this context, the lack of direct empirical data also presents a methodological opportunity, enabling the development of controlled, replicable, and scalable modelling approaches that can guide subsequent empirical investigation.

Future research should prioritise the empirical validation of the SVAI Model by comparing its predicted outputs with eye-tracking data collected from participants interacting with tourism websites, thereby establishing convergent validity. Beyond validation, the framework may be extended to account for dynamic and interactive content, including video-based environments and social media platforms, where attentional processes unfold over time. Comparative studies across different destination types, cultural contexts, and device formats (e.g., mobile versus desktop) would further enhance the generalisability of the model. In addition, integrating the model with affective and behavioural measures would strengthen its explanatory power, linking predicted attentional patterns with downstream outcomes such as engagement, recall, and booking intention.

From an epistemological standpoint, this research suggests a shift in how tourism interfaces can be studied, moving from reactive and post-hoc evaluation methods towards predictive and model-based approaches. This transition opens new possibilities for understanding how attention is structured, guided, and potentially designed within digital environments. Importantly, visual attention should not be understood as an end in itself, but as part of a broader cognitive-affective and experiential system. In this sense, the findings open a pathway for future research to investigate how predicted attentional patterns relate to affective responses and destination-related outcomes, such as engagement, decision-making, and loyalty intentions. This integrative perspective aligns with emerging interdisciplinary approaches that seek to connect perception, emotion, and action within tourism experiences.

At the same time, the study highlights the importance of methodological pluralism. Simulated attention should not replace empirical observation but rather complement it, forming part of a broader analytical ecosystem that integrates computational, behavioural, and contextual data. Ultimately, this research contributes to an emerging epistemological repositioning within tourism studies, one that moves towards predictive, integrative, and cognitively informed frameworks capable of addressing complex experiential phenomena in digitally mediated environments.

References

- Andersson, R., Larsson, L., & Holmqvist, K. (2010). Vision and occlusion: How and why occluders guide the eyes. *Journal of Eye Movement Research*, 3(3), 1-15. <https://doi.org/10.16910/jemr.3.3.3>
- Borji, A., Parks, D., & Itti, L. (2014). Complementary effects of gaze direction and early saliency in guiding fixations during free viewing. *Vision Research*, 98, 1-14. <https://doi.org/10.1016/j.visres.2014.02.002>
- Borji, A., Sihite, D. N., & Itti, L. (2012). Probabilistic learning of task-specific visual attention. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 470-477. <https://doi.org/10.1109/CVPR.2012.6247710>

- Cardoso, L., Dias, F., Lima Santos, L., & Moutinho, L. (2022). Exploring the influence of digital platforms on destination image formation. *Journal of Tourism and Cultural Change*, 20(4), 522-538. <https://doi.org/10.1080/14766825.2021.1879826>
- Cardoso, L., Estevão, C., Fernandes, C., & Alves, H. (2019). A deductive content analysis of tourism destination image formation. *Journal of Destination Marketing & Management*, 14, 100380. <https://doi.org/10.1016/j.jdmm.2019.100380>
- Chakraborty, S., Mandal, B., & Bhattacharyya, S. (2023). Adaptive visual saliency prediction using deep learning. *Pattern Recognition Letters*, 165, 47-55. <https://doi.org/10.1016/j.patrec.2022.11.015>
- Chen, X., Zhang, Q., Liu, W., Rahimi, R., & Qi, J. (2025). Exploring visual attention and perception in hospitality and tourism: a comprehensive review of eye-tracking research. *Journal of Hospitality and Tourism Insights*, 8(7), 2726-2743. <https://doi.org/10.1108/jhti-06-2024-0621>
- Confetto, M. G., Conte, F., Palazzo, M., & Siano, A. (2023). Digital destination branding: A framework to define and assess European DMOs practices. *Journal of Destination Marketing & Management*, 30, 100804. <https://doi.org/10.1016/j.jdmm.2023.100804>
- Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, 3(3), 201-215. <https://doi.org/10.1038/nrn755>
- Desolneux, A., Moisan, L., & Morel, J. M. (2007). *From gestalt theory to image analysis: a probabilistic approach*. Springer Science & Business Media. <https://doi.org/10.1007/978-0-387-74378-3>
- Duchowski, A. T. (2007). *Eye Tracking Methodology: Theory and Practice* (2nd ed.). Springer. <https://doi.org/10.1007/978-1-84628-609-4>
- Echtner, C. M., & Ritchie, J. R. B. (1993). The measurement of destination image: An empirical assessment. *Journal of Travel Research*, 31(4), 3-13. <https://doi.org/10.1177/004728759303100402>
- Engel, S. A., Zhang, X., & Wandell, B. A. (1997). Colour tuning in human visual cortex measured with functional magnetic resonance imaging. *Nature*, 388(6637), 68-71. <https://doi.org/10.1038/40398>
- Espigares-Jurado, F., Muñoz-Leiva, F., Correia, M. B., Sousa, C. M. R., Ramos, C. M. Q., & Faisca, L. (2020). Visual attention to the main image of a hotel website based on its position, type of navigation and belonging to Millennial generation: An eye tracking study. *Journal of Retailing and Consumer Services*, 52, 101906. <https://doi.org/10.1016/j.jretconser.2019.101906>
- Fernández-Cavia, J., Marchiori, E., Haven-Tang, C., & Cantoni, L. (2017). Online communication in Spanish destination marketing organizations: The view of practitioners. *Journal of Vacation Marketing*, 23 (3), 264-273. <https://doi.org/10.1177/1356766716640840>
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127-138. <https://doi.org/10.1038/nrn2787>
- García-Díaz, A., Leboran, V., Fdez-Vidal, X. R., & Pardo, X. M. (2012). On the relationship between optical variability, visual saliency, and eye fixations: A computational approach. *Journal of Vision*, 12(6), 17. <https://doi.org/10.1167/12.6.17>
- Greenspan, H., Belongie, S., Goodman, R., Perona, P., Rakshit, S., & Anderson, C. H. (1994). Overcomplete steerable pyramid filters and rotation invariance. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 222-228. <https://doi.org/10.1109/CVPR.1994.323833>
- Guerrero-Rodríguez, R., Stepchenkova, S. & Kirilenko, A. (2020). Experimental investigation of the impact of a destination promotional video on tourists' perceptions, attitudes, and intent to visit that destination. *Tourism Management Perspectives*, 33, 100625. <https://doi.org/10.1016/j.tmp.2019.100625>
- Hayes, A. F., & Krippendorff, K. (2007). Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods and Measures*, 1(1), 77-89. <https://doi.org/10.1080/19312450709336664>
- Henderson, J. M., & Hayes, T. R. (2018). Meaning guides attention in real-world scene images: Evidence from eye movements and meaning maps. *Journal of Vision*, 18(6), 10. <https://doi.org/10.1167/18.6.10>

- Holmqvist, K., Nyström, M., & Mulvey, F. (2012). Eye tracker data quality: what it is and how to measure it. In *ETRA '12 Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 45-52). Association for Computing Machinery (ACM). <https://doi.org/10.1145/2168556.2168563>
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press. <https://global.oup.com/academic/product/eye-tracking-9780198738596>
- Humphreys, G. W. (2016). Feature confirmation in object perception: Feature Integration Theory 26 years on from the Treisman Bartlett Lecture. *Quarterly Journal of Experimental Psychology*, 69 (10), 1910-1940. <https://doi.org/10.1080/17470218.2014.988736>
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2(3), 194-203. <https://doi.org/10.1038/35058500>
- Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11), 1254-1259. <https://doi.org/10.1109/34.730558>
- Kastner, S., & Ungerleider, L.G. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience*, 23, 315-41. <https://doi.org/10.1146/annurev.neuro.23.1.315>
- Katsuki, F., & Constantinidis, C. (2014). Bottom-up and top-down attention: different processes and overlapping neural systems. *The Neuroscientist*, 20(5), 509-521. <https://doi.org/10.1177/1073858413514136>
- Kingstone, A., Smilek, D., Ristic, J., Kelland Friesen, C., & Eastwood, J. D. (2003). Attention, Researchers! It Is Time to Take a Look at the Real World. *Current Directions in Psychological Science*, 12(5), 176-180. <https://doi.org/10.1111/1467-8721.01255>
- Köchling, A. (2020). "Dream now, travel later": Pre-travel online destination experiences on destination websites. *Journal of Qualitative Research in Tourism*, 1(1), 51-72. <https://doi.org/10.4337/jqrt.2020.01.03>
- Krippendorff, K. (2018). *Content Analysis: An Introduction to Its Methodology* (4th ed.). Sage. <https://doi.org/10.4135/9781071878781>
- Kristjánsson, Á., & Egeth, H. (2020). How feature integration theory integrated cognitive psychology, neurophysiology, and psychophysics. *Attention, Perception, & Psychophysics*, 82 (1), 7-23. <https://doi.org/10.3758/s13414-019-01803-7>
- Kümmerer, M., Bethge, M., & Wallis, T. S. A. (2022). DeepGaze III: Modeling free-viewing human scanpaths with deep learning. *Journal of Vision*, 22(5), 7. <https://doi.org/10.1167/jov.22.5.7>
- Lee, W., & Gretzel, U. (2012). Designing persuasive destination websites: A mental imagery processing perspective. *Tourism Management*, 33(5), 1270-1280. <https://doi.org/10.1016/j.tourman.2011.10.012>
- Leventhal, A. G. (1991). *The neural basis of visual function*. London: Macmillan Press.
- Li, P., Xiao, X., & Seekamp, E. (2026). Unveiling the Complexity of Tourists' Visual Attention and Emotions Toward Destination Pictures. *Journal of Travel Research*, 65(4), 991-1014. <https://doi.org/10.1177/00472875251331517>
- Li, Q., Huang, Z. J., & Christianson, K. (2016). Visual attention toward tourism photographs with text: An eye-tracking study. *Tourism Management*, 54, 243-258. <https://doi.org/10.1016/j.tourman.2015.11.017>
- Liesefeld, H. R., & Müller, H. J. (2019). Distractor handling via dimension weighting. *Current Opinion in Psychology*, 29, 160-167. <https://doi.org/10.1016/j.copsyc.2019.03.003>
- Lin, R., & Lin, W. (2014). A computational visual saliency model based on statistics and machine learning. *Journal of Vision*, 14(9). <https://doi.org/10.1167/14.9.1>
- Marzi, T., Balzano, M., & Marchiori, M. (2024). K Alpha Calculator: An open-source tool for computing Krippendorff's Alpha inter-rater reliability coefficient. *MethodsX*, 12, 102545. <https://doi.org/10.1016/j.mex.2023.102545>
- Metzger, W. (1928). *Gesetze des Sehens*. Frankfurt am Main: Kramer.

- Muñoz-Leiva, F., Faísca, L. M., Ramos, C. M., Correia, M. B., Sousa, C. M., & Bouhachi, M. (2021). The influence of banner position and user experience on recall: The mediating role of visual attention. *Spanish Journal of Marketing - ESIC*, 25(1), 85-114. <https://doi.org/10.1108/SJME-04-2020-0050>
- Ortanderl, F. (2025). Complex destination image formation: The changes through visitation among domestic tourists. *European Journal of Tourism Research*, 41, 4115. <https://doi.org/10.54055/ejtr.v41i.4100>
- Penz, E., Hofmann, E., & Hartl, B. (2017). Fostering sustainable travel behaviour: Role of sustainability labels and goal-directed behaviour regarding touristic services. *Sustainability*, 9(6), 1056. <https://doi.org/10.3390/su9061056>
- Pomerantz, J. R. (2006). Visual Grouping and Attention. In R. R. Boothe (Ed.), *Perception of Visual Stimuli* (Vol. 1, pp. 211-231). Oxford University Press.
- Quinlan, P. T. (2003). Visual feature integration theory: Past, present, and future. *Psychological Bulletin*, 129(5), 643-673. <https://doi.org/10.1037/0033-2909.129.5.643>
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372-422. <https://doi.org/10.1037/0033-2909.124.3.372>
- Savin, G.-D., Fleşeriu, C., & Batrancea, L. (2021). Eye tracking and tourism research: A systematic literature review. *Journal of Vacation Marketing*, 28(3), 285-302. <https://doi.org/10.1177/13567667211053387>
- Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain and Cognition*, 112, 92-97. <https://doi.org/10.1016/j.bandc.2015.11.003>
- Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, 11(5), 5. <https://doi.org/10.1167/11.5.5>
- Todorović, D. (1997). Gestalt principles and the perception of visual stimuli. *Psychology of Perception*, 14(2), 223-235.
- Todorović, D. (2008). Gestalt principles. *Scholarpedia*, 3(12), 5345. <https://doi.org/10.4249/scholarpedia.5345>
- Todorović, D. (2020). Gestalt perception and saliency: Integration of features in visual scenes. *Advances in Psychology*, 24(4), 447-469.
- Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97-136. [https://doi.org/10.1016/0010-0285\(80\)90005-5](https://doi.org/10.1016/0010-0285(80)90005-5)
- Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, 14(1), 107-141. [https://doi.org/10.1016/0010-0285\(82\)90006-8](https://doi.org/10.1016/0010-0285(82)90006-8)
- Valtakari, N. V., Hooge, I. T. C., Viktorsson, C., Nystrom, P., Falck-Ytter, T., & Hessels, R. S. (2021). Eye tracking in human interaction: Possibilities and limitations. *Behaviour Research Methods*, 53, 1592-1608. <https://doi.org/10.3758/s13428-020-01517-x>
- Voßkühler, A., Nordmeier, V., Kuchinke, L., & Jacobs, A. M. (2008). OGAMA (Open Gaze and Mouse Analyzer): Open-source software designed to analyze eye and mouse movements in slideshow study designs. *Behaviour Research Methods*, 40(4), 1150-1160. <https://doi.org/10.3758/brm.40.4.1150>
- Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin*, 138(6), 1172-1217. <https://doi.org/10.1037/a0029333>
- Wickens, C. (2021). Attention: Theory, Principles, Models and Applications. *International Journal of Human-Computer Interaction*, 37(5), 403-417. <https://doi.org/10.1080/10447318.2021.1874741>
- Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, 1(2), 202-238. <https://doi.org/10.3758/bfo3200774>
- Wu, X., Lai, I.K.W. and Zhou, H. (2025). How visual and textual information influences repeat tourists: Insights from self-report data and an eye-tracking experiment. *Tourism Management Perspectives*, 58, 101396. <https://doi.org/10.1016/j.tmp.2025.101396>
- Zhang, P., Jamison, K., Engel, S., He, B., & He, S. (2012). Binocular rivalry requires visual attention. *Neuron*, 71(2), 362-369. <https://doi.org/10.1016/j.neuron.2011.05.035>