UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

DEVELOPMENT OF ALGORITHMS TO DETERMINE ACCURATE PARAMETERS FOR EYE MOVEMENT DETECTION FOR VISUAL SCANNING BEHAVIOR ANALYSIS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

RICARDO PALMA FRAGA

Norman, Oklahoma

2024

DEVELOPMENT OF ALGORITHMS TO DETERMINE ACCURATE PARAMETERS FOR EYE MOVEMENT DETECTION FOR VISUAL SCANNING BEHAVIOR ANALYSIS

A DISSERTATION APPROVED FOR THE
SCHOOL OF INDUSTRIAL AND SYSTEMS ENGINEERING

BY THE COMMITTEE CONSISTING OF

Dr. Ziho Kang, Chair

Dr. Randa Shehab

Dr. Shivakumar Raman

Dr. Talayeh Razzaghi

Dr. Edward Cokely

Acknowledgements

I would first like to thank Dr. Ziho Kang, who has been my mentor and advisor over the last few years, always motivating and pushing me forward on all academic and professional pursuits through our work and all our conversations, as well as for introducing me to the world of air traffic control and eye tracking – which ended up becoming two of my favorite research areas.

I would also like to thank all members of my PhD committee. Dr. Shehab, who gave me my only B grade throughout my time at OU ISE in her Ergonomics class, which ironically became my first introduction to the world of Human Factors & Ergonomics. Dr. Cokely, whose advice and insights over the years I always appreciated and valued – this dissertation is centered around "solving the problem of solving the problem", one of many insights you shared with me. Dr. Raman, who motivated me in the first place to pursue a graduate degree from a brief chat, couldn't have been more than two minutes, one advising session at the end of a semester. Dr. Razzaghi, whose Healthcare Analytics and Timeseries Analysis courses, two of my favorite classes I always tell students to take, introduced me to so many new tools I've used throughout my PhD degree.

To my parents, Olga and Antonio, brother, Jorge, and my uncles and cousins for their support all these years. I'm also extremely grateful for the support of my fiancé, Miranda, as well as our dog, Hazel, who have been by my side all these years and are probably tired of hearing about the work you'll read in this dissertation.

In addition, I am really grateful for so many others – Melodi, Cheryl, and Jenn, who are the foundation of the school of ISE, who have always had their door open to hear me out and have helped me so many times I lost count. My colleagues, June, Jahnavi, and Saptarshi, for their help, support, and engaging conversations over the years. Dr. Dodd, for the opportunity to serve as the

TA for Ergonomics (yeah, the same class in which I got a B) and allowing me to pursue creative new ways of teaching in the lab. Dr. Crutchfield, who also introduced me to the world of aviation and air traffic control – our trips to the air traffic control towers at MCO, MSP, and DEN really shaped my passion for this field.

Lastly, the research in this dissertation was supported and funded by the FAA NextGen Organization's Human Factors Division, ANG-C1. The FAA Technical Sponsor was the Air Traffic Organization's Safety and Technical Training Service Unit's Policy and Performance Division, Safety Performance Group, Human Performance Team (AJI-342). Furthermore, it is noted that some of the contents included in the chapters of the present dissertation have been used as part of journal papers.

Table of Contents

Table of Figures

Table of Tables

Abstract

One way to investigate how humans interact with their environment is by studying how they visually search and gather relevant information in order to make decisions. Visual search can be analyzed through visual scan paths, the time ordered sequence of eye fixations and saccadic movements. To create visual scan paths, researchers often use eye fixation detection algorithms, many of which rely on threshold parameters set by the researcher, to automatically identify eye fixations from data collected by eye tracking devices.

However, the choice of threshold parameters used by eye movement detection algorithms is crucial, as many different factors, such as the participant population and the task to be completed, might affect what threshold values can accurately identify eye fixations. Inaccurate thresholds might result in visual scan paths that do not resemble the visual scan path carried out by an individual (i.e., the ideal visual scan path). For example, an inaccurate threshold might fail to identify eye fixations that took place, or combine multiple consecutive eye fixations together into a single eye fixation and place it somewhere in the environment that the individual never actually observed. As such, using inaccurate thresholds might affect our ability to understand and interpret an individual's visual search (e.g., what information was observed, as well the order it was observed in) and decision-making process.

In this dissertation, novel procedures and algorithms are introduced to facilitate the identification and selection of accurate thresholds. First, an automated procedure was developed to automatically select accurate thresholds based on the impact of threshold values on eye movement metrics (e.g., number of eye fixation), expanding upon prior research efforts by automating a process that was previously largely manual. Second, two approaches are proposed to approximate the trend of similarities between ideal visual scan paths and visual scan paths created at different thresholds,

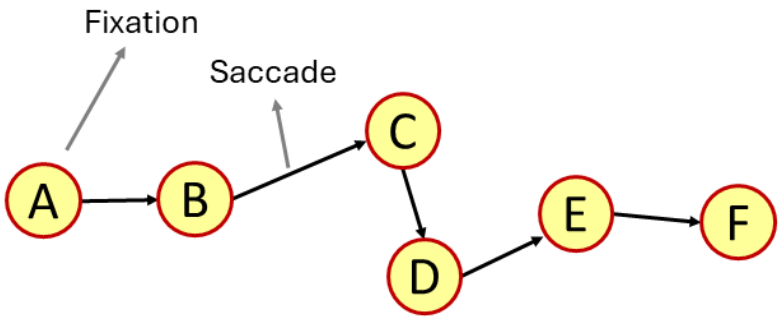
used in prior studies to determine accurate thresholds, without the need to know or use ideal visual scan paths. Using ideal visual scan paths is not always feasible, as one needs to know the expected eye movements of individuals a head of time or needs to engage in the arduous and time consuming process of manually defining the ideal visual scan paths from the data collected. Third, and lastly, a classification framework was developed to identify similar visual scan paths that might showcase variations of a common visual scanning strategy using multiple similarity metrics.

Chapter 1 – Introduction

In many tasks, our ability to visually search the environment to gather information plays a crucial role. Every year, healthcare professionals visually search and interpret millions of medical images and records, such as mammograms (Mello-Thoms et al., 2005; Gandomkar & Mello-Thoms, 2019), air traffic controllers inspect radars and runways to search for hazards in order to ensure that over 16 million flights (Federal Aviation Administration, 2023) arrive safely at their destination, and transportation security officers at airports inspect over 1.9 billion carry-on items (Transportation Security Administration, 2024). Therefore, understanding how we visually search our environment, such as what information we observe and the order we observe it, while completing a task, might help us identify opportunities to adapt the task to the user's capabilities and needs, and in turn, making the task less demanding.

One way to study how we visually search our environment is by collecting and analyzing eye movements. To determine whether a target is present in the environment, such as an aircraft in a potential conflict or a tumor on a mammogram, we must direct our attention to the target in order to identify it (Wolfe, 2010). If multiple items are present in the environment, each item might need to be inspected until the target is found (Wolfe, 2020). Throughout this search process, as our eyes move across the environment, we create what is known as a visual scan path (Josephson & Holmes, 2002) (Figure 1), the time-ordered sequence of eye fixations (when we stop to inspect an item) and saccades (when we move between items).

Figure 1. Representative examples of a visual scan path, highlighting one eye fixation and one saccadic movement.



Note. The yellow circles denote an eye fixation, and the inscribed letter indicates the order the eye fixations took place in. The black line between each eye fixation represents a saccadic movement.

In order to create visual scan paths, researchers often use eye movement detection algorithms to automatically identify eye fixations and saccadic movements. Manual identification of eye fixations and saccades from the eye movement data collected by an eye tracker is possible but considered to be both very time consuming and tedious (Blignaut & Beelders, 2009; Hooge et al., 2018; Navarro et al., 2021). Instead, researchers apply eye movement detection algorithms such as the Velocity-Threshold Identification (I-VT) and the Dispersion-Threshold Identification (I-DT) algorithms (Salvucci & Goldberg, 2000), two of the most commonly used algorithms by researchers (Strohmaier et al., 2020; Hahn & Klein, 2022; Birawo & Kasprowski, 2022). These algorithms rely on threshold parameters set by the researcher, such as the gaze velocity threshold for the I-VT algorithm or the dispersion threshold for the I-DT algorithm, in order to identify and classify eye fixations and saccades from the eye movement data collected.

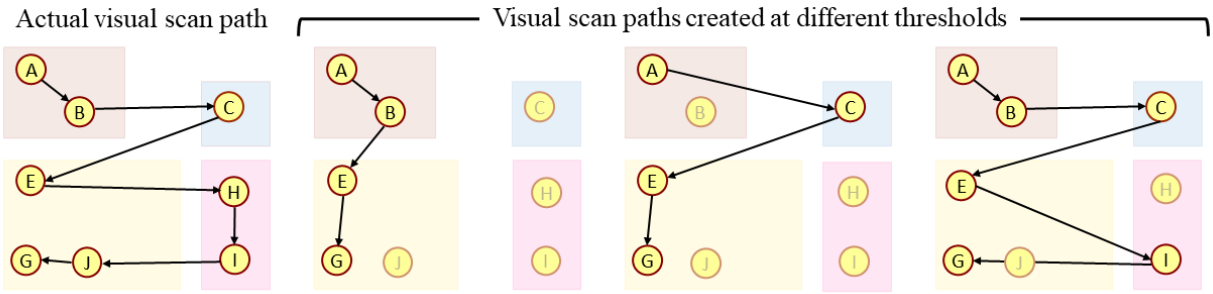
However, the choice of threshold values set by the researcher is crucial to create accurate visual scan paths. Prior studies have described how eye movement metrics calculated from the same eye movement data, such as the number of eye fixations and eye fixation durations, might be different depending upon the threshold value selected (Shic, Scassellati, & Chawarska, 2008; Komogortsev

et al., 2010; Kosel et al., 2023). Thus, due to the different number of eye fixations identified, different thresholds might also create different visual scan path from the same eye movement data (Yoo, Jeong, & Jang, 2021), where an inaccurate threshold might result in a visual scan path that is very different from the one that actually took place.

As a result, our ability to analyze the visual search carried out by a participant while completing a task might be affected by our choice of threshold value. Consider the example of the visual scan path showcased in Figure 2(a) carried out by a fictitious radiologist examining an x-ray for an abnormality. One can observe that the radiologist checked all areas of interest (AOIs) in the environment, denoted by squares colored red, blue, yellow, and pink, which represent the location of potential abnormalities in the x-ray. The visual scan paths created using inaccurate thresholds of an eye movement detection algorithm can be observed in Figures 2(b) and Figures 2(c). In this case, on the other hand, actual eye fixations carried out by the radiologist on AOIs do not appear in the visual scan path, indicating to us, erroneously, that they did not fixate on those areas while searching. Yet, the visual scan path created with a more accurate threshold, Figure 2(d), indicates that the radiologist fixated on all the AOIs on the same order (red, blue, yellow, pink, yellow) as done in the actual visual scan path, although not all eye fixations were identified. Overall, different thresholds might result in different visual scan paths, some of which may be poor representations of the actual eye movements that took place during search.

In addition, how we might be able to interpret the visual scan path is also impacted by the choice of threshold. Assuming that the pink AOI represents the location of the abnormality in the x-ray image, the visual scan paths shown in Figure 2(b) and Figure 2(c) indicate that the radiologist committed a search or sampling error (Nodine & Kundel, 1987; Waite et al., 2017), as the radiologist did not inspect the location of the lesion. However, as shown in actual visual scan path

Figure 2. Simplified example highlighting the impact of inaccurate and accurate thresholds on visual scan paths (b-d) compared to an actual visual scan path (a).



(a) (b) (c) (d)

Note. The red, blue, yellow, and pink squares represent areas of interest (AOIs) in the environment that the visual scan path inspected.

carried out by the radiologist in Figure 2(a), as well as the visual scan path created at a more accurate threshold shown in Figure 2(d), the radiologist did inspect the location of the lesion, and thus, could not have committed a search or sampling error. Thus, our ability to interpret the visual scan path of the radiologist can be very different based upon whether an accurate threshold or an inaccurate threshold is used.

Multiple different factors can influence the threshold values that should be selected, requiring researchers to identify and select accurate thresholds for their respective application. Prior studies have discussed how factors ranging from the participant population investigated, such as older participants (Blignaut & Beelders, 2009), the task to be completed (van der Lans et al., 2011), the device in which the task is completed, such as a computer or a mobile device (Trabulsi et al., 2021), as well the eye tracker used to collect the eye movement data, which can have different sampling frequencies (Olsen & Matos, 2012), might impact which threshold values are accurate. As a result,

threshold values that create accurate visual scan paths in one study might not be as accurate in a separate study.

However, current proposed approaches to determine accurate thresholds for eye movement detection algorithms might be challenging for researchers to implement in practice. For example, a literature review of Strohmaier et al (2020) of 161 eye-tracking studies focused on mathematics education found that only 61% of studies reported the eye movement detection algorithm used and among those, only 21% reported the thresholds that were applied (Strohmaier et al., 2020).

One proposed approach to determine accurate thresholds is to compare how similar the visual scan path created at a given threshold is to an ideal visual scan path, such as the expected or pre-determined sequence of a participants' eye movements (Komogortsev et al. 2010). Nonetheless, an ideal visual scan path might not always be readily available, limiting the feasibility of this procedure to specific experimental designs and applications (Startsev & Zemblys, 2023). Another proposed approach is to evaluate how the values of eye movement metrics, such as the number of eye fixations, change across threshold values (Blignaut, 2009). Yet current implementations of this approach rely on the researcher's subjective judgement to identify accurate thresholds.

In addition, once accurate thresholds have been identified, analyzing accurate visual scan paths enable us to investigate how operators search their environment in multiple different domains, ranging from aircraft piloting (Naeeri, Kang, & Palma Fraga, 2022) and air traffic control (Kang & Landry, 2014; McClung & Kang, 2016; Palma Fraga, 2021), to reading websites (Josephson & Holmes, 2002), automobile driving (Jeong, Kang, and Liu, 2019; Navarro et al., 2021), education (Tang et al., 2016; Špakov et al., 2017). Although there exist multiple ways to analyze and compare visual scan paths, two common approaches to compare visual scan paths are to evaluate them based on the similarity information observed (i.e., the AOIs in common between two visual scan paths)

or to evaluate them based on the order said information was observed in (i.e., to evaluate them based upon whether they inspected the AOIs in the same order) (Privitera & Stark, 2000).

However, analyzing and comparing visual scan paths by only considering one similarity metric might not be sufficient to identify visual scan paths that represent variations of an underlying similar scanning behaviors. In complex and dynamic environments operators might need to vary their visual scan paths while carrying out their tasks. In the case of air traffic control, prior research has described how controllers might gather the same information in a different order or from a different sources in the environment (Meyer et al., 2021). Furthermore, the same controller might gather the same information every time they issue a clearance but do might do so in a different order to save time during periods of high workload (Svensson, 2015). Therefore, using solely one similarity metric to compare visual scan paths may not fully account for the potential variations that might exist between participants applying a similar scanning behavior.

Research contributions

The present dissertation contains three major contributions to address two gaps highlighted in the introduction: (1) identifying accurate thresholds for eye movement detection algorithms; (2) identifying similar visual scan paths when variations of a common visual scanning behavior may exist.

The first contribution (Chapter 3) introduces an automated procedure to identify a range of accurate thresholds based on eye movement metric trends across threshold values, expanding upon the work of Blignaut (2009). The procedure defines a minimum and a maximum threshold across eye movement metrics given a set of eye movement data through a three-step procedure: (1) identify whether the eye movement metric trend contains any changes in direction (increasing or decreasing)

as well as concavity (convex or concave). If so, split the trend to account for said changes; (2) determine the thresholds that take place at elbow or knee points in each trend using the Kneedle algorithm (Satopaa et al., 2011), and if needed, the splits of each trend that account for changes in concavity and direction; (3) calculate a common minimum and maximum threshold among the thresholds identified between all eye movement metric trends. The procedure might enable researchers to select accurate thresholds for their respective application automatically based on the data they collect.

The second contribution (Chapter 4) consists of two methods, referred to as between-participant and within-participant comparisons, to identify accurate threshold values by quantifying the impact of threshold values on visual scan paths similarities. More specifically, the between-participant method compares how threshold values impact the similarity between the visual scan paths of participants, while the within-participant method evaluates how threshold values impact the similarity of a single participant's visual scan path created at different thresholds. Quantifying the impact of threshold values on visual scan paths can enable researchers to select more accurate thresholds without the need of ideal visual scan paths.

The third contribution (Chapter 5) introduces a classification framework to identify similar visual scan paths when potential variations might exist between individuals by applying multiple measures to quantify the similarity of visual scan paths across two different dimensions: (1) the AOIs inspected; (2) the order said AOIs were inspected in. Similar visual scan paths are empirically identified based upon their above-average values among participants across both similarity metrics, while those visual scan paths that might be potential variations are considered to be those with an above-average value in at least one similarity metric. On the other hand,

different visual scan paths are considered to be those that had less than average values for both similarity metrics.

The remaining two chapters, Chapter 2 and 6, provide readers with additional background information on the motivation and methods used to develop the contributions discussed in the present work (Chapter 2), as well as a summary of the conclusions regarding the contributions in addition to future avenues of research (Chapter 6).

Chapter 2 – Background

How threshold selection might impact the accuracy of visual scan paths

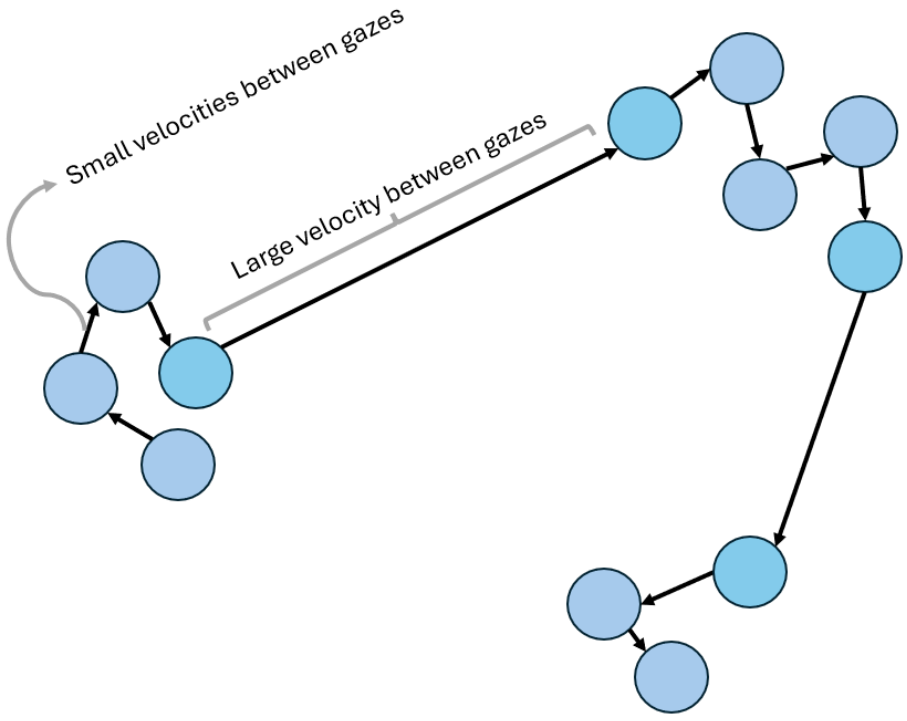
In this section, thresholds can influence and impact the accuracy of visual scan paths is explained. First, we introduce the Velocity-Threshold Identification (I-VT) algorithm, focusing on how threshold values are used by the algorithm to identify and classify eye fixations and saccadic movements from gaze data. Second, the factors that can influence the accuracy of threshold values, such as the participant population or the device used to complete the task, are described. Third, and lastly, we describe how selecting inaccurate thresholds values can impact the accuracy of visual scan paths.

The role of thresholds in the Velocity-Threshold Identification (I-VT) algorithm

The I-VT algorithm belongs to the family of eye movement detection algorithms that identifies eye fixations and saccadic movements from gaze data based on the velocity of eye movements (Salvucci & Goldberg, 2000). Eye fixations and saccadic movements possess two distinct distributions of velocities, where saccadic movements are much faster than eye fixations (Salvucci & Goldberg, 2000). More specifically, Salvucci & Goldberg (2000) describe that eye fixations tend

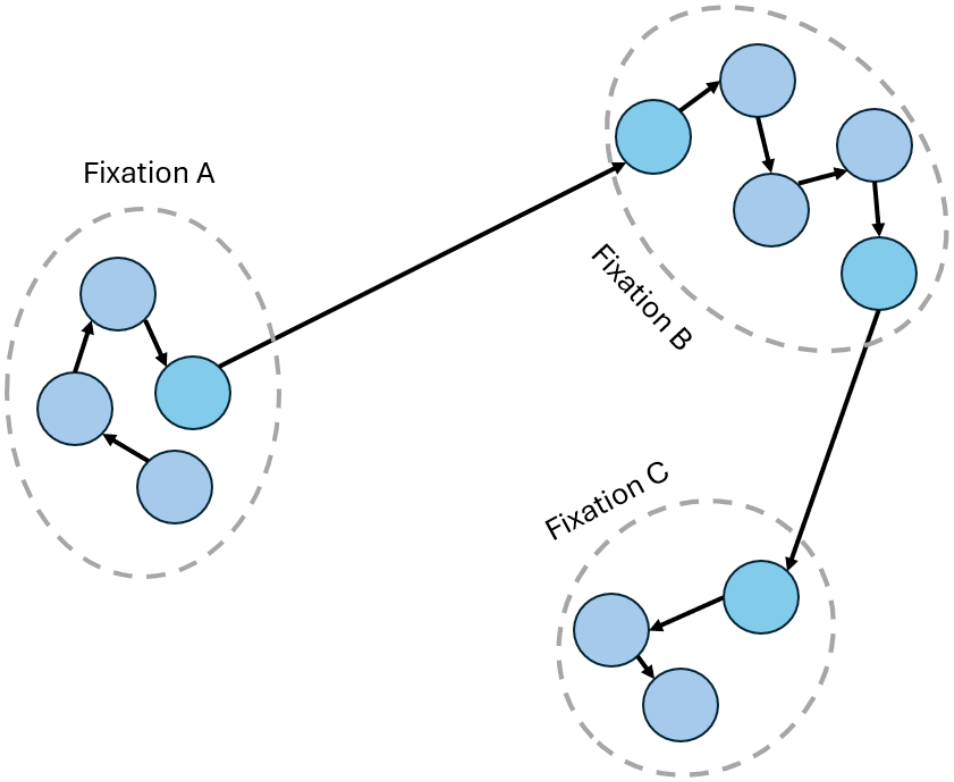
to take place at velocities smaller than 100 degrees per second, while saccadic movements tend to occur at velocities higher than 300 degrees per second. Consider the example showcasing gaze movements of a participant collected by an eye tracker shown in Figure 3 below. Here, the blue dots represent gaze movements, while the length of the arrow represents the velocity between gaze movements. Gaze movements belonging to eye fixations tend to be in close proximity of each other, possessing small velocities between consecutive gazes (i.e., shorter arrows), while those belonging to saccadic movements tend to be much farther apart, separated by larger velocities (i.e., larger arrows). At its core, the I-VT algorithm leverages that different velocity profiles exist for eye fixations and saccadic movements, represented by larger or smaller arrows in our example, in order to identify them.

Figure 3. Simplified visualization showcasing gaze movements (blue dots) of a participant that have low gaze velocities (small arrows) and high gaze velocities (large arrows).



As a result, vast amounts of collected gaze data can be rapidly classified as belonging to an eye fixation or a saccade based on a fixed gaze velocity threshold value (Salvucci & Goldberg, 2000; Andersson et al., 2017). If the velocity between two gaze points is less than a fixed gaze velocity threshold, they are classified as belonging to an eye fixation, otherwise, they are classified as belonging to a saccadic movement. Continuing the example shown in Figure 3, Figure 4 showcases how the gaze movements of a participant can be classified into three separate eye fixations. The gazes with velocities smaller than the fixed threshold are classified as belonging to an eye fixation and grouped together (i.e., grey dashed circles), while gazes with velocities larger than the threshold are used to separate eye fixations each eye fixation.

Figure 4. Simplified visualization showcasing how gaze movements of a participant are grouped together into eye fixations using a fixed velocity threshold.



Note. Blue dots indicate the gaze movements of a participant, the arrows the gaze velocity between gaze movements, where larger arrows indicate higher velocities, and the gray dashed circles indicate an eye fixation.

Furthermore, researchers have defined additional thresholds beyond the gaze velocity thresholds to increase the accuracy of eye movement detection algorithms. Thresholds such as a minimum eye fixation duration threshold (Komogortsev et al., 2010), a minimum saccade duration threshold (Salvucci & Goldberg, 2000) among others have been used. For example, a minimum eye fixation duration is used to discard eye fixations that are considered to be too short in duration for a human to gather and process any meaningful information from the environment (Olsen, 2012).

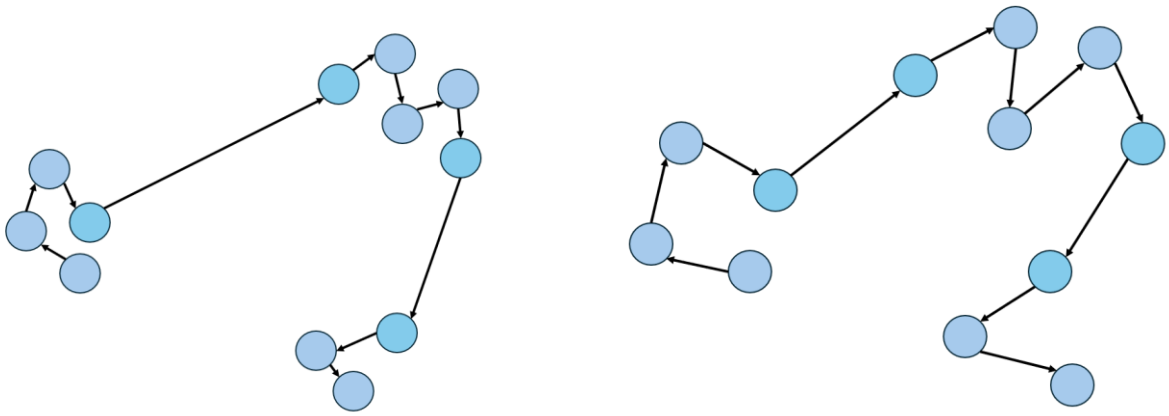
Factors that influence the accuracy of thresholds

As mentioned in the introduction, multiple different factors such as the participant population (Blignaut & Beelders, 2009), the device in which the task is completed (Trabulsi et al., 2021), the eye tracker used (Olsen & Matos, 2012), among others, might affect the threshold values that can accurately identify eye fixations and saccadic movements.

The participant population might affect the choice of threshold due to individual differences in eye movement characteristics. The work of Blignaut & Beelders (2009) explored how gaze stability can vary between individuals, due to factors such as age, the presence of ocular tremors (involuntary movements of the eye), among others. In their work, they describe how different individuals can show different distributions of point-to-point gaze velocities. As a result, a threshold value that might be accurate for one participant might not be as accurate for another participant. Consider the examples highlighted in Figure 5(a-b). The gaze movements shown in Figure 5(a) are more stable, resulting in smaller velocities (i.e., smaller arrows) between gaze movements that belong to eye fixations and much larger velocities between those movements belonging to saccades. On the other hand, Figure 5(b) contains the gaze movements of a participant

with a less stable gaze, resulting in larger movements between gazes and therefore larger gaze velocities (i.e., larger arrows). As a result, a threshold that might accurately identify eye fixations for Figure 5(a) might not have the same accuracy for Figure 5(b) due to the larger gave velocities between gazes belonging to eye fixations, which might be classified as saccades instead.

Figure 5. Simplified visualization highlighting how gaze stability between two different participants, one possessing high gaze stability (a) and one with poor gaze stability (b), may influence threshold accuracy.



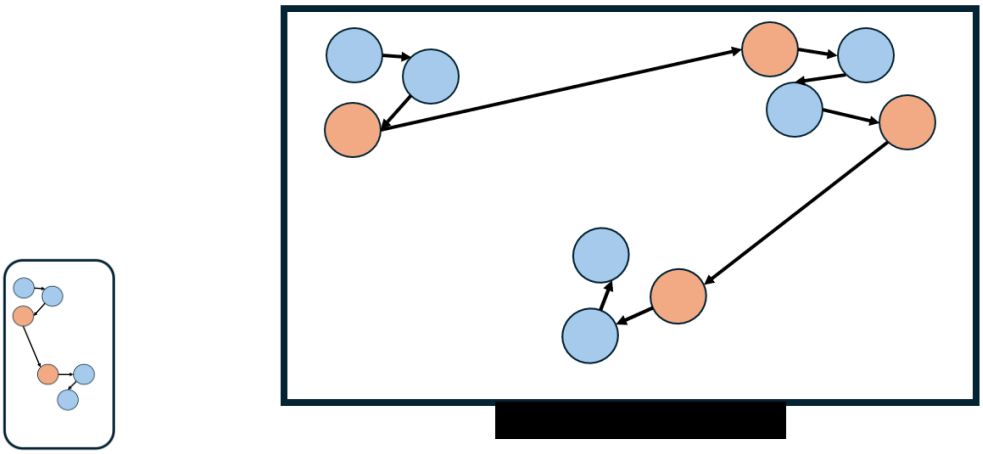
(a) (b)

Note. Poor gaze stability might increase the gaze velocities between gaze movements, resulting in threshold values that might accurately identify all eye fixations in (a) to perform poorly in (b).

In addition, the device used by participants to carry out a task might also affect what threshold values the researcher should select (Trabulsi et al, 2021). In their work, Trabulsi et al (2021) discuss how stimulus presentation on a mobile screen might influence the gaze velocities of eye movements compared to computer screens. More specifically, that a smaller screen closer to a participant might induce smaller velocities between gaze movements than a larger screen, as shown in Figure 6(a-b). Figure 6(a) showcases the gaze movements of a participant on a mobile device, while Figure 6(b) contains the gaze movements of another participant on a computer screen. Due

to the larger size difference between the two stimuli presentations, the gaze velocities between eye movements on the larger computer screen might be larger (i.e., larger arrows) than those in the mobile device (i.e., smaller arrows). As a result, the threshold values that result in the accurate classification of eye fixations and saccadic movements in one device might not be the same in another device.

Figure 6. Simplified visualization highlighting how the device used to complete a task, such as a mobile device (a) or a computer screen (b), may influence threshold accuracy.



(a) (b)

Note. Smaller screens may influence the gaze velocities of participants to be lower than those on a computer screen due to the smaller screen size.

Another factor that might influence the accuracy of thresholds is the eye tracker used to collect gaze data (Olsen & Matos, 2012). Prior studies have suggested that gaze data sampled at lower frequencies might require higher thresholds for the I-VT algorithm, particularly in cases where the data is noisy (Holmqvist, 2016). Different eye trackers can collect data at vastly different frequencies, with some eye trackers capable of collecting gaze movements as frequently as 2000 Hz (Holmqvist et al., 2011). For example, the EyeLink 1000 Plus can collect data at either 1000 Hz or 2000 Hz, while the Tobii Pro Glasses II can collect data at either 50 Hz or 100 Hz. In other

words, the EyeLink 1000 Plus is capable of collecting gaze movements every millisecond (1000 Hz) or every 0.5 millisecond (2000 Hz), while the Tobii Pro glasses II can collect gaze data every 20 milliseconds (50 Hz) or every 10 milliseconds (100 Hz).

Overall, when identifying eye fixations and saccadic movements by applying the I-VT algorithm, researchers need to identify and select thresholds that can accurately classify eye movements in their respective application.

Inaccurate gaze velocity thresholds impact visual scan path accuracy

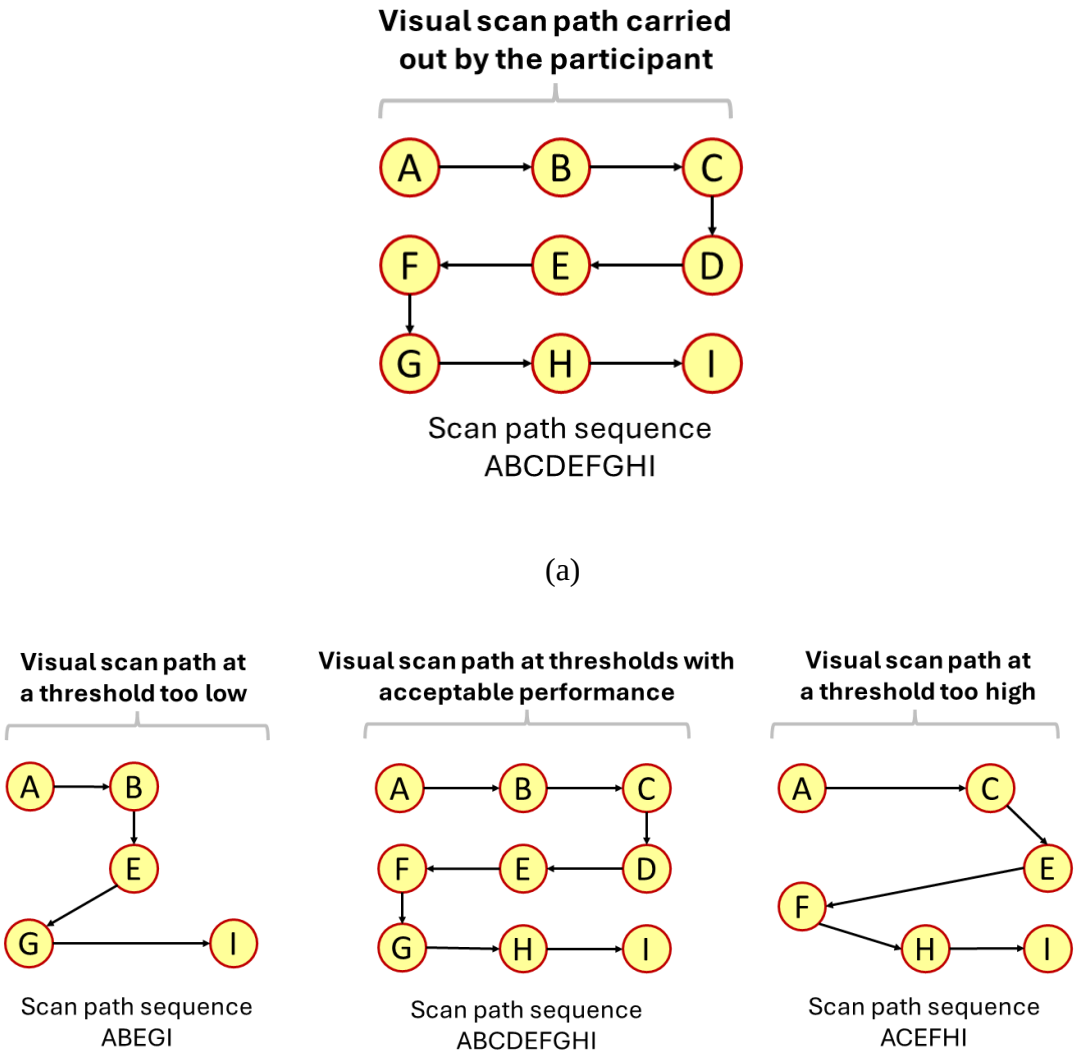
Inaccurate gaze velocity thresholds can be classified into two categories (Olsen 2012, Trabulsi et al., 2021) based upon how they impact visual scanpaths: (1) gaze velocity thresholds that are higher than the values they ought to be, which may begin to classify saccadic movements as eye fixations; (2) gaze velocity thresholds that are smaller than the value they should be, which can fail to identify eye fixations.

Consider the simplified example of a visual scan path carried out by a participant in Figure 7(a), where the participant fixated on 9 locations on the screen (A-I) in a specific order.

Using a gaze velocity threshold too high may misclassify saccadic movements as belonging to an eye fixation. As a result, multiple eye fixations separated by saccadic movements could be erroneously combined into a singular eye fixation that did not actually take place in the visual scan path. Furthermore, the combined eye fixation may also be placed at a location that the participant never actually observed (Blignaut, 2009), as the centroid of an eye fixation is defined as the average of the x and y coordinates of the consecutive gazes classified as an eye fixation (Salvucci & Goldberg, 2000). Consider the example provided in Figure 7(b, right). A threshold too high combined multiple separate eye fixations together, such as eye fixations B and C into C, eye fixations E and D into E, and eye fixations F and G into F. Further, the combination of multiple

eye fixations impacted their positions, placing them at locations that the participant never actually observed.

Figure 7. Simplified example showcasing the impact of thresholds values on visual scan paths.



Using a gaze velocity threshold too high may misclassify saccadic movements as belonging to an eye fixation. As a result, multiple eye fixations separated by saccadic movements could be erroneously combined into a singular eye fixation that did not actually take place in the visual scan path. Furthermore, the combined eye fixation may also be placed at a location that the participant

never actually observed (Blignaut, 2009), as the centroid of an eye fixation is defined as the average of the x and y coordinates of the consecutive gazes classified as an eye fixation (Salvucci & Goldberg, 2000). Consider the example provided in Figure 7(b, right). A threshold too high combined multiple separate eye fixations together, such as eye fixations B and C into C, eye fixations E and D into E, and eye fixations F and G into F. Further, the combination of multiple eye fixations impacted their positions, placing them at locations that the participant never actually observed.

On the other hand, if a gaze velocity threshold too low was selected, gaze samples belonging to eye fixations may be misclassified as saccadic movements, which can lead to, for example, an eye fixation being erroneously split into multiple separate eye fixations (Salvucci & Goldberg, 2000). At the extreme, a threshold too low might classify all the gaze samples belonging to an eye fixation as a saccadic movements, completely failing to identify that eye fixation. In other words, a threshold too low might separate a singular eye fixation into multiple eye fixations or even fail to identify that the eye fixation took place at all. Consider the example presented in Figure 2(7, left). Using a threshold too low identified the eye fixations C, D, F, and H carried out by the participant to be saccadic movements, resulting in the scan path sequence ABEGI. The scan path sequence created with a threshold too low is not only missing eye fixations, but it also indicates that the participant carried out eye movement transitions that never took place, such as fixation B followed by fixation E.

In-between the thresholds that are too high and too low, as shown in Figure 7(b, center) exists a range of thresholds that lead to acceptable performance – those threshold values can most accurately identify the eye fixations and saccadic movements that were carried out.

Current approaches to determine thresholds that create accurate visual scan paths

The present section introduces two current approaches used in prior literature to identify accurate visual scan paths, as well as the limitations these approaches have, which are addressed in this dissertation. The first approach discussed revolves around comparisons to an ideal visual scan path, the expected eye movements of a participant, as they complete a task. The ideal visual scan path can be used to determine which threshold results in a visual scan path that best represents the ideal one. The second approach discussed investigates how eye movement metrics are impacted across thresholds in order to determine thresholds that result in eye movement metric values that match those expected by the researcher.

Comparing to an ideal visual scan path

One way to identify how accurately a threshold can classify the eye movements is to compare the eye movements identified by the threshold to the eye movements the participant carried out. Consider the actual visual scan path, as well as the various visual scan paths created at thresholds too low and too high, showcased in Figure 7. In this case, through visual observation, we can identify that the visual scan path shown in Figure 7(b, center) is much more accurate than the visual scan paths shown in Figure 7(left, right), as it is most closely approximates the visual scan path carried out by the participant in Figure 7(a). In other words, when the visual scan path of the participant is known (i.e., the ideal visual scan path), the most accurate threshold is the one that returns a visual scan path that most closely resembles the eye movements of the participant (Komogortsev et al., 2010; Komogortsev & Karpov, 2013).

An example of this procedure can be observed in the work of Komogortsev et al (2010). In their study, the authors instructed participants to fixate on a target bullseye on a computer monitor that

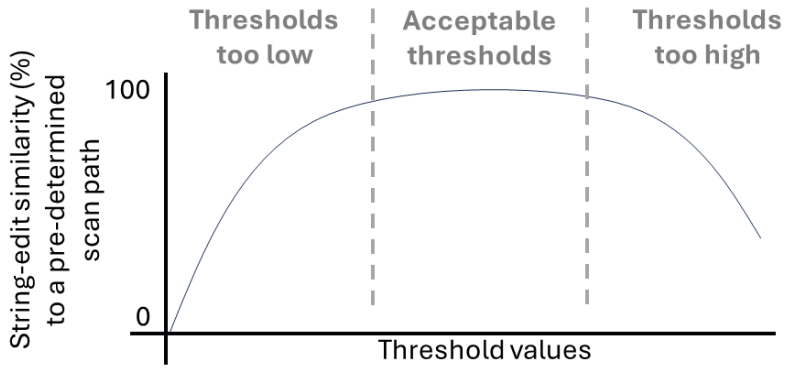
changed positions in a pre-determined manner coded by the authors. Using this information, the authors were able to define characteristics of the ideal visual scan path, such as the number of eye fixations (there were a total of 10 targets, an accurate threshold would calculate 10 eye fixations), the average eye fixation duration (each target was present for 1 second, an accurate threshold would return an average eye fixation duration of 1 second), among others. Therefore, the threshold of the eye movement detection algorithm that best matched the expected eye movements of a participant would be considered as the most accurate threshold for their particular application.

Note that other metrics can be used to compare visual scan paths created at thresholds and the ideal visual scan path. For example, another metric used in prior literature is the string-edit similarity (Blignaut & Beelders, 2009; Hareżlak & Kasprowski, 2014), a metric popular in eye tracking research (Privitera & Stark, 2000; Eraslan, Yesilada & Harper, 2016). Briefly, the string-edit similarity, explained in more detail below, calculates the number of operations needed to convert one visual scan path into another visual scan path, allowing us to determine how similar two visual scan paths are without having to define specific characteristics (e.g., average eye fixation durations expected). In this approach, the most accurate threshold would be considered to be the one that creates a visual scan path that most closely resembles the ideal visual scan path carried out by the participant.

Furthermore, prior literature suggests that there is a relationship between the string-edit similarity comparison of an ideal scan path visual scan path and the visual scan paths created at a range of threshold values appears (Blignaut & Beelders, 2009; Hareżlak & Kasprowski, 2014). More specifically, the trend appears to increase sharply at low thresholds, remain stable for a range of thresholds, after which it begins to decrease, as can be observed in Figure 8 below. In other words,

there might exist a potential range of thresholds that can lead to accurate and acceptable visual scan paths when comparing to an ideal visual scan path.

Figure 8. Example visualization of the theoretical relationship between threshold values and string-edit similarity to an ideal scan path sequence observed in prior research studies.



Nonetheless, comparing to an ideal visual scan path might not be feasible in many applications. First, it requires the researcher to define the ideal visual scan path of a participant ahead of time (Startsev & Zemblys, 2023), which might not be possible without a carefully managed study (as in the work of Komogortsev et al., 2010). Alternatively, one may choose manually to re-create the visual scan path from the raw gaze data collected, although it can be very time consuming (Blignaut & Beelders, 2009; Navarro et al., 2021). For example, identifying the visual scan path of a participant whose eye movements were collected using the Tobii Pro Glasses II eye tracker (which collects gaze data every 10 milliseconds as mentioned previously). For a short 5 second duration experiment, the researcher would have to manually process 5000 gaze points (assuming no missing data) in order to re-crate the visual scan path of the participant.

One of the major contributions of the present dissertation, discussed further in Chapter 2, introduces a novel procedure to address the challenge of identifying accurate thresholds using string-edit similarity without the need of an ideal visual scan path.

Comparing to eye movement metrics trends

In many cases, such as in complex and dynamic environments, it may be difficult for a researcher to know ahead of time what the visual scan path of a participant is, making the method defined in the previous section not feasible. Nonetheless, the need to identify accurate thresholds for eye movement detection algorithms remains.

One approach to identify accurate thresholds when ideal visual scan paths are not known is to explore how threshold values impact eye movement metrics, such as the number of eye fixations, and evaluating them against a set of criteria (Blignaut, 2009; Llanes-Jurado et al., 2020). More specifically, the method determines the range of thresholds that, simultaneously, result in the highest average number of eye fixation, the highest percentage of gazes included in eye fixations, as well as a high and stable number of eye fixations in AOIs, among other metrics selected by the researcher (Blignaut, 2009; Llanes-Jurado et al., 2020).

These metrics and their criteria (i.e., selecting the highest number of eye fixations) serve as heuristics to determine accurate thresholds. For example, Blignaut (2009) considered that the highest number of eye fixations might better match the actual participants' eye fixations than larger thresholds with fewer eye fixations. They describe that, a higher number of shorter eye fixations may take place at the locations inspected by the participants, while at larger thresholds, in which eye fixations may begin to merge, the position of eye fixations may change (Blignaut, 2009). Therefore, they state that it is preferrable to select a smaller threshold with more eye fixations than a higher threshold with fewer eye fixations at locations that may have not been observed. Another example is selecting the threshold that results in high and stable number of eye fixations in AOIs (Llanes-Jurado et al., 2020). In their work, Llanes-Jurado et al (2020) used his measure as a way to capture information regarding the spatial-temporal behavior of the participants' eye movements.

If participants are expected to primarily fixate in AOIs as part of an application or experiment, thresholds that result in an unstable number of eye fixations in AOIs, or in very few eye fixations in AOIs, may be an indication that the threshold is not an accurate model of visual attention (Llanes-Jurado et al., 2020).

Overall, understanding how different thresholds impact eye movement metrics can facilitate threshold selection, even in the absence of an ideal visual scan path, as a researcher can avoid using thresholds in eye movement detection algorithms that do not appear to have reasonable performance (Startsev & Zemblys, 2023) for the eye movement metrics chosen.

However, the application of this procedure may be challenging in practice, as it relies on the researcher's subjective judgement to identify potential thresholds. In prior applications, researchers have visually observed an eye movement metric trend in order to identify specific points in a trend (and the threshold at which it takes place), such as the point at which an increasing trend begins to slow down (Blignaut, 2009; Llanes-Jurado et al., 2020). Prior research has described how, when tasked with visually determining the points at which a trend changed, participants may vary in terms of how many points they select as well as where those points take place in the trend. Furthermore, the task of visually identifying a specific point in a trend can be difficult, especially when trends are smooth, as it may be hard to differentiate a specific point in the trend from all surrounding points (Shi et al., 2021).

One of the major contributions of the present dissertation, discussed further in Chapter 4, is the development of an automated procedure to identify thresholds based on this approach, allowing researchers to identify thresholds without relying on their subjective judgement.

Quantifying similarity in visual scan paths

The present section describes two approaches used in eye tracking literature to analyze and compare visual scan paths based on the order information was observed or based on what information was observed (Privitera & Stark, 2000). In this dissertation, within Chapter 5, these approaches are combined to develop a classification framework to identify similar visual scan paths, particularly in cases in which visual scan paths can have minor variations due to the task environment. The two similarity metrics explored are the Jaccard coefficient similarity, which measures the similarity in terms of common AOIs explored between two visual scan paths, as well as the string-edit similarity, which considers the order in which AOIs were explored.

Jaccard coefficient similarity

One approach to evaluate the similarity between visual scan paths based on the AOIs present in both scan paths is through the Jaccard similarity coefficient (Jaccard, 1901; Fletcher & Islam, 2018). The Jaccard coefficient similarity has previously been used in multiple different domains, such as machine learning (Ferdous, 2009; Huang, 2009) and computational biology (Besta et al., 2020), as well as eye tracking (Kumar et al., 2019; Burch et al., 2019). Within the context of visual scan paths, the Jaccard coefficient similarity is defined as the number of AOIs common between two visual scan paths (i.e., the intersection of AOIs) divided by the total number of AOIs present in both visual scan paths (i.e., the union of AOIs) as can be observed in equation (2.1).

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{2.1}$$

Here, A represents the set of AOIs present in one scan path sequence, and B the set of AOIs present in another scan path sequence, $A \cap B$ contains the number of AOIs in common between the two scan path sequences, and $A \cup B$ includes the number of AOIs present in both scan path sequences.

Consider the scan path sequence A = *PLH* and the scan path sequence B = *POL*. Here, the value of $A \cap B$ would be 2, as there are 2 AOIs in common between the two scan path sequences (i.e., PL), while $A \cup B$ would be 4, as there are 4 AOIs in total present between the two scanpath sequences (i.e., O and H in addition to PL). Therefore, the Jaccard coefficient similarity between these two scan path sequences would be $J(A,B) = {}^{2}/_{5} = 0.4$ – in other words, the two visual scan paths (PLH and POL) are 40% similar.

String-edit similarity

To calculate the similarity between visual scan paths by considering the order that information was inspected in (i.e., AOIs), a common metric used in the literature is the string-edit similarity (Privitera & Stark, 2000; Duchowski et al., 2010, Fahimi & Bruce, 2021). As mentioned briefly in this chapter, the string-edit distance computes the number of operations (insertions, deletions, and substitutions) required to transform one visual scan path into another (Levenshtein, 1966).

Furthermore, one can compute the normalized string-edit similarity (Privitera & Stark, 2000), where the string-edit distance is converted into similarity by subtracting one from the string-edit distance value calculated and normalize the string-edit similarity by dividing by the length of the longest scan path sequence. Such an approach allows us to directly compare multiple visual scan paths, and their similarity values, while accounting for the length (i.e., how many AOIs were observed) in the visual scan path. One or two differences in the order AOIs were inspected in between two short visual scan paths might result in very low similarity, while the same one or two differences in very large visual scan paths might not impact the similarity value. The equation to calculate the string-edit normalized similarity can be observed in as can be observed in equation (2.2) below.

$$S(A,B) = 1 - \left(\frac{i + d + s}{n}\right) \quad (2.2)$$

In this equation, A represents one scan path sequence while B another scan path sequence. The variables i, d, and s indicate the number of insertions, deletions, and substitutions, respectively, needed to convert the scan path sequence A into B. Lastly, n represents the length of the largest scan path sequence between A and B. Consider as an example the scan path sequence A = *EFD* and the scan path sequence B = *EFCD.* To convert the scan path sequence A into the sequence B, one must insert the AOI C into scan path sequence A. Thus, only one insertion operation is needed. Given the length of the largest scan path sequence is 4, the normalized string-edit similarity would be $S(A,B) = 1 - \left(\frac{1}{4}\right) = 0.75$.

Chapter 3 – Identifying accurate thresholds by analyzing eye movement metric trends automatically

In this chapter, we introduce an automated procedure to identify accurate thresholds based on the impact of threshold values on eye movement metric trends, more precisely: (1) the number of eye fixations; (2) the percentage of gaze samples in eye fixations; (3) the percentage of eye fixations in AOIs.

The chapter is structured in the following manner. First, we introduce the proposed approach as well as explain the calculations that take place at each step. Second, we apply the proposed approach to the random saccade task (RAN) in the open-source Gazebase dataset (Griffith et al., 2021). Third, and lastly, we discuss the results and derived insights from the application of the proposed approach, focusing on the accurate thresholds identified by the proposed approach.

Proposed approach

The proposed procedure expands upon prior research efforts in four ways: (1) computes inflection points in eye movement metric trends – points at which changes in concavity take place. Such points are used to partition the eye movement metric trend into multiple trends to account for their concavity and slope direction when using elbow-detection algorithms; (2) adapts the Kneedle elbow-detection algorithm (Satopaa et al., 2011) to automatically detect thresholds at elbow points in each eye movement metric trend; (3) identifies minimum and maximum thresholds by determining two sets of elbow (or knee) points; (4) creates aggregate visual scanpaths that simplify visual scanpaths and highlight the impact of thresholds on the overall shape of visual scanpaths and the position of eye fixations. In addition, it provides information to the researcher on what AOIs inspected are missing across thresholds.

The proposed approach consists of 5 major steps, each explained in more detail below: (1) processing the eye movement data; (2) calculating the eye movement metrics, concavity, and slope directions; (3) detecting elbow (or knee) points in eye movement metrics; (4) identifying minimum and maximum recommended thresholds; (5) plotting visual scan paths at the recommended thresholds.

Step 1. Process eye movement data

The proposed procedure processes the eye movement data collected, which consists of the following sub-steps: (1) Selecting the eye movement metrics to evaluate; (2) Selecting an eye movement detection algorithm (e.g., I-VT) and the threshold range to be explored by choosing a maximum threshold (t_{max}) and step increase (Δ); (3) Setting the size of the AOIs in the environment and mapping eye fixations to AOIs.

First, the researcher selects the eye movement metrics with which to select and evaluate thresholds. Prior studies have selected eye movement metrics such as (e.g., Blignaut 2009; Llanes-Jurado et al., 2020): (1) the number of eye fixations; (2) the percentage of gaze samples in eye fixations; (3) the percentage of eye fixations in AOIs. The intuition behind these metrics is to identify the range of thresholds in the eye movement that (Llanes-Jurado et al., 2020): (1) classify most gaze samples as eye fixations rather than saccades; (2) lead to a high number of eye fixations; (3) ensure that a sufficient amount of eye fixations identified take place on the AOIs created by the researcher. Thresholds that satisfy such criteria are considered to be thresholds with good performance.

Second, the eye fixations and saccadic movements must be identified from the collected eye movement data. Multiple eye movement detection algorithms exist, such as the I-VT and Dispersion-Threshold Identification (I-DT) algorithms (Salvucci & Goldberg, 2000), two of the most commonly implemented algorithms in eye tracking research (e.g., Strohmaier et al., 2020;

threshold (e.g., Komogortsev et al., 2010). Such a modification is done for two reasons: (1) Eye fixations lower than the minimum eye fixation duration threshold are considered to be too short to represent any meaningful user behavior (Olsen, 2012); (2) Enables comparisons with other eye movement detection algorithms that already contain a minimum eye fixation duration threshold (e.g., I-DT) (Komogortsev et al., 2010). Prior studies have used a minimum eye fixation duration threshold ranging from 60 ms (Olsen, 2012) to 100 ms (Komogortsev et al., 2010).

In addition, the researcher selects a maximum threshold (t_{max}) and step increase (Δ) to systematically explore and evaluate thresholds, a process known as creating a "gridsearch" (e.g., Syarif et al., 2016). The maximum threshold and step increase are used to define the set of thresholds $T = \{t_0 = \Delta, t_1 = t_0 + \Delta, t_2 = t_0 + 2\Delta, \dots, t_i, \dots, t_{max}\}$. For example, setting Δ = 5 °/s results and $t_{max} = 200$ °/s would create the set of thresholds $T = \{5, 10, 15, \dots, 200\}$. Prior studies using the I-VT algorithm have created different threshold ranges (Komogortsev et al., 2010; Larsson, 2010), with the highest velocity threshold used being 400 °/s, alongside step increases of either 5 °/s or 8 °/s. For the I-DT algorithm, prior studies (Komogortsev et al., 2010; Llanes-Jurado et al., 2020) have used a maximum threshold of 2.5°, using step increases of either 0.1° or 0.33°. Selecting a large increment step is a commonly used strategy when implementing a "gridsearch" for two reasons (Hsu et al., 2003). First, it is much less time-consuming to compute, which can be particularly important when using eye trackers that collect eye movements at high rates, some of which can be 1000 Hz or higher (Hosp et al., 2020). Second, if needed, the increment steps can always be decreased later for a more granular and exhaustive search (Hsu et al., 2003).

Third, each eye fixation is aligned temporally and spatially with the AOIs in the environment in order to determine if the eye fixation occurred within any AOI, a process known as "mapping"

(Kang et al., 2016). For this process, the researcher sets the maximum spatial distance (s_{max}) between an eye fixation and an AOI.

During the mapping process, the eye fixations and the AOIs are compared based on their timestamps (i.e., if the eye fixation took place when the AOI was visible) and their positions (i.e., if the eye fixation took place within the bounds of the AOI) in order to determine if the eye fixation occurred within the AOI. First, eye fixations and AOIs can be aligned by their timestamps by identifying which eye fixations took place within the timeframe AOI was present in the environment. Second, an eye fixation can be considered to be within an AOI if it falls within the bounds defined by the AOI. For example, in the case of a circular AOI, if the Euclidean distance between the centroid of the eye fixation, and the center of the AOI, was less than or equal to s_{max} (i.e., the radius of the AOI a circular AOI), then the eye fixation is considered to have taken place on the AOI. Such a process can be observed in equation (3.1), where (a_j, b_j) represents coordinates of the eye fixation in the environment, while (a_s, b_s) the coordinates of the circular AOI in the environment.

$$\sqrt{(a_j - a_s)^2 + (b_j - b_s)^2} \leq s_{max} \tag{3.1}$$

The s_{max} values must be large enough to account for the potential discrepancy between the actual gaze location of the participant and the gaze location reported by the eye tracker, also known as the visual angle error (e.g., Kang et al., 2016; Mandal & Kang, 2015). For example, the EyeLink 1000 Hz (SR Research, Ottawa, Ontario, Canada) has an average visual angle error of 0.5 (Griffith et al., 2021), and thus, the AOI size must be large enough to account for it. In addition, prior research has estimated the visual angle error by instructing participants to fixate on specific sets of validation points (e.g., Griffith et al., 2021). In this procedure, the Euclidean distance between

the gaze reported by the eye tracker and a validation point that participants were instructed to fixate on is calculated, and can be used to define s_{max}. Lastly, note that other AOI shapes such as rectangular or convex can be used, and the selection may depend upon the shape of the stimuli presented (Kang et al., 2016).

Step 2. Calculate eye movement metric values, slope direction, and concavity

At this step, the proposed procedure is implemented: (1) calculate the average values across participants for each eye movement metric (e.g., number of eye fixations; percentage of eye fixations in AOIs, percentage of gaze samples in eye fixations); (2) Afterwards, we apply a Savitzky-Golay filter (Savitzky & Golay, 1984) in order to smooth the data by fitting a polynomial function and calculates the first and second derivatives of the trend. The first and second derivatives are used to determine the concavity and slope direction.

More specifically, the Savitzky-Golay filter (Savitzky & Golay, 1984) fits the average eye movement metric value to a polynomial function, allowing us to smooth the eye movement metric trend, as well as to calculate the first and second derivatives of the (fitted) trend, which are used to determine the concavity and slope direction. The Savitzky-Golay filter (Savitzky & Golay, 1984) has been used previously in prior eye tracking research (Ouzts & Duchowski, 2012; Duchowski et al., 2016), to determine, for example, the velocity and acceleration of eye movements by calculating the first and second derivatives of gaze positions (Nyström & Holmqvist, 2010). The Savitzky-Golay filter determines the best fitting polynomial function that describes a trend via least squares minimization (Ouzts & Duchowski, 2012). In other words, the filter "finds the best polynomial function that describes the raw data" (Nyström & Holmqvist, 2010). Equation (3.2) showcases the application of the Savitzky-Golay filter to smooth average

values of an eye movement metric trend ($\bar{c}_{t_i}$), calculate the first derivative ($\bar{c}_{t_i}{}'$) and the second derivative of the trend ($\bar{c}_{t_i}{}''$).

$$\bar{c}_{t_i} = SG_0\left(\frac{1}{p}\sum_{j=1}^{p} x_{t_i j} \quad \forall\, t_i \in T\right) \tag{3.2}$$

$$\bar{c}_{t_i}{}' = SG_1\left(\frac{1}{p}\sum_{j=1}^{p} x_{t_i j} \quad \forall\, t_i \in T\right)$$

$$\bar{c}_{t_i}{}'' = SG_2\left(\frac{1}{p}\sum_{j=1}^{p} x_{t_i j} \quad \forall\, t_i \in T\right)$$

Here, SG represents the function of Savitzky-Golay filter, which includes an optional parameter to determine which (if any) derivative to calculate (0 if no derivative (SG_0), 1 if the first derivative (SG_1), and 2 if the second derivative (SG_2)), $x_{t_i j}$ the eye movement metric value at the t_ith threshold in the set T for the jth participant, and p the total number of participants present in the data set. In other words, the input values into the Savitzky-Golay filter at the average values of the eye movement metric at each threshold (i.e., the trend) between participants.

Furthermore, in existing publications, there are three representative trends that can be observed for the impact of eye movement metric trends across threshold values: concave or concave down with an increasing slope, concave or concave down with a decreasing, as well as a combination of concave down or concave up and increasing or decreasing slopes. A portion of those papers showcased how minor fluctuations in the trend can exist in the trend, which could influence what elbow points are detected, and thus, the recommended minimum and maximum thresholds identified in equations (3.7) and (3.8). To prevent those minor fluctuations to be identified as

inflection points, the Savizky-Golay filter was used to smooth the data by fitting the trend to a polynomial function.

Determining the concavity and slope direction of the trend is important, as it dictates whether an elbow or knee needs to be detected on the trend (Satopaa et al., 2011; Onumanyi et al., 2022); however, some eye movement metrics can change both their concavity and slope direction throughout the trend due to the impact of thresholds. For example, prior literature has shown that the number of eye fixations tend to increase rapidly up until a specific threshold after which they remain stable for a period of time and begin to decrease (Komogortsev et al., 2010). A similar trend can be observed for the number of eye fixations in AOIs (Llanes-Jurado et al., 2020). On the other hand, other eye movement variables, such as the number of gaze samples in eye fixations, appear to maintain the same concavity and slope direction as visualized in prior research studies (Blignaut, 2009; Llanes-Jurado et al., 2020). As a result, the proposed procedure must be able to determine when these changes in concavity and slope direction take place in order to calculate the correct elbow or knee point.

Afterwards, we can determine the concavity (v_{t_i}) and direction of the slope (l_{t_i}) at each threshold value as shown in equation (3.3). The concavity of a function is determined by the sign of the second derivative ($\bar{c}_{t_i}{}''$), calculated via the Savitzky-Golay filter, which if positive ($\bar{c}_{t_i}{}'' > 0$) can be considered as concave up, and if negative ($\bar{c}_{t_i}{}'' < 0$), concave down. Similarly, the direction of the slope can be determined by the sign of the first derivative ($\bar{c}_{t_i}{}'$), calculated via the Savitzky-Golay filter, which is increasing if positive (i.e., $\bar{c}_{t_i}{}' > 0$) , and decreasing if negative ($\bar{c}_{t_i}{}' < 0$) (e.g., Stewart et al., 2015; Jones et al., 2019).

$$v_{t_i} = \begin{cases} 0 \; if \; \bar{c}_{t_i}'' > 0 \\ 1 \; if \; \bar{c}_{t_i}'' < 0 \end{cases} \text{ and } l_{t_i} = \begin{cases} 0 \; if \; \bar{c}_{t_i}' > 0 \\ 1 \; if \; \bar{c}_{t_i}' < 0 \end{cases} \quad (3.3)$$

Lastly, at this step, the proposed procedure defines four characteristics for each eye movement metric: $\{(t_i, \bar{c}_{t_i}, v_{t_i}, l_{t_i}) | \; t_i \in T\}$, where t_i represents *i*th threshold value, $\bar{c}_{t_i}$ the average eye movement metric value from the polynomial function fit via the Savitzky-Golay filter at each threshold, and v_{t_i} and l_{t_i} contain the concavity and slope direction at each threshold value.

Step 3. Detect elbow (or knee) points between inflection points for each eye movement metric

In the present step, the proposed procedure: (1) identifies any inflection points in the eye movement metric (i.e., points at which changes in concavity take place) and partitions the eye movement metric into multiple trends between inflection points; (2) applies the Kneedle algorithm (Satopaa et al., 2011) to detect thresholds at elbow (or knee) points in the eye movement metric trend.

First, the proposed procedure identifies the thresholds at which inflection points – points at which concavity changes (e.g., Jones, 2019) – occur in the eye movement metric. An inflection point (f_{k-1}) is defined when the curvature at threshold t_{i-1} (i.e., $v_{t_{i-1}}$) has a different value than the curvature at t_i (i.e., v_{t_i}). The proposed procedure creates the set *F* that contains all thresholds at which inflection points take place (i.e., $f_1, \ldots, f_{k-1}, f_k$), if any, in the eye movement metric trend (3.4).

$$F = (f_1, \ldots, f_{k-1}, f_k) \text{ where } f_k = t_{i-1} \; only \; if \; v_{t_{i-1}} \neq v_{t_i} \quad (3.4)$$

Afterwards, the eye movement metric is partitioned into multiple trends between the inflection points to account for changes in the trend due to the impact of thresholds. For example, assume that the starting threshold ($f_0 = t_0$) and final threshold ($f_k = t_{max}$) are added to the set *F* such that *F*

$= (f_0, f_1, \ldots, f_{k-1}, f_k)$. In this case, if one inflection point was found, the proposed procedure would partition the eye movement metric trend into two trends: one beginning from f_0 (i.e., the starting threshold t_0) to f_{k-1} (i.e., the threshold where the inflection point found), while the other trend starts from f_{k-1} (i.e., the inflection point found) to f_k (i.e., the maximum threshold t_{max}). However, if no inflection points are found, only one elbow (or knee) point is found for the entire eye movement metric. This process can be observed in in equation (3.5), where $R_1, R_2, \ldots, R_k$ represent portions of the kth eye movement metric trends between inflection points identified (e.g., R_1 contains the threshold values (t_i) and eye movement metric values by the Savizky-Golay filter ($\bar{c}_{t_i}$) between the thresholds f_0 and f_1). In addition, the process can be visualized in Figure 9 below.

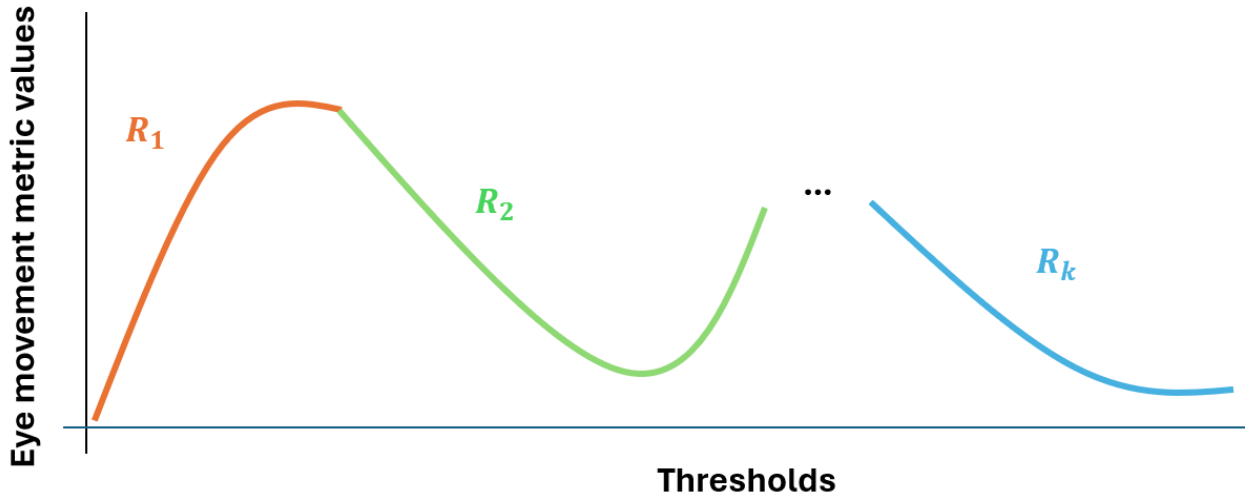
$$R_1 = \{(t_i, \bar{c}_{t_i})\} \; \forall \; t_i \in T \; where \; f_0 \leq t_i \leq f_1\} \tag{3.5}$$

$$R_2 = \{(t_i, \bar{c}_{t_i})\} \; \forall \; t_i \in T \; where \; f_1 \leq t_i \leq f_2\}$$

$$\ldots$$

$$R_k = \{(t_i, \bar{c}_{t_i})\} \; \forall \; t_i \in T \; where \; f_{k-1} \leq t_i \leq f_k\}$$

Figure 9. Application of equation (3.5) to split the eye movement metric trend values based on the thresholds identified inflection points.



Second, the proposed approach applies the Kneedle algorithm (Satopaa et al., 2011) to each trend between inflection as implemented in (Arvai, 2020) to detect elbow (or knee) points based on the concavity and slope direction detected. A trend is considered to have an increasing slope if the initial value of the slope direction (i.e., $l_{f_{i-1}}$) in the trend is 0 and decreasing if the initial value is 1. Similarly, the trend is considered to be concave down if the value of concavity is 1 (i.e., $v_{f_{i-1}}$), while it is concave up if the value is 0.

In more detail, the intuition behind the Kneedle algorithm is to detect the knee in a trend by identifying the point with the largest vertical distance from a straight line formed by the start and end points of the trend (Satopaa et al., 2011). The following summarized steps of the Kneedle algorithm (Satopaa et al., 2011) describe the implementation in (Arvai, 2020) as used in the proposed procedure: (1) The eye movement metric trend is smoothed using linear interpolation; (2) Both the eye movement metric values and the threshold values are normalized to be between $[0,1]$ based on their corresponding minimum and maximum values; (3) The Kneedle algorithm assumes that the trend is concave down and increasing to identify a knee point (Satopaa et al., 2011; Arvai, 2020). If the shape of the eye movement metric trend was determined to be something other than concave down and increasing, it must be transformed (Satopaa et al., 2011; Arvai, 2020) prior to calculating the difference trend; (4) Calculate the difference trend by subtracting the normalized eye movement metric values minus the normalized threshold values; (5) The local maxima points are identified in the difference trend. A local maxima is the point in the difference trend where the points before and after have a smaller value than the local maxima point; (6) The elbow (or knee) point of an eye movement metric trend is considered to be the first local maxima point identified in the difference trend. The sensitivity parameter described in (Satopaa et al., 2011), used when

identifying knees in an 'online' setting (i.e., when additional points are added to the original trend over time) was not utilized as no additional points are added to the trend (Satopaa et al., 2011).

The detected threshold at the elbow (or knee) point (e_k) for the eye movement metric at a given inflection point is computed (3.6). Where K represents the application of the Kneedle algorithm for the trend split (R_k) while accounting for the concavity ($v_{f_{k-1}}$) and slope direction ($l_{f_{k-1}}$). As the Kneedle algorithm detects the local maxima point in the difference trend, the proposed procedure identifies the threshold at which the local maxima point takes place.

$$e_k = \arg\max K(R_k |\, v_{f_{k-1}},\, l_{f_{k-1}}) \,\forall\, k \tag{3.6}$$

Note that multiple inflection points and elbow (or knee) points can be detected, the proposed procedure only considers at most two elbow points to calculate a recommended minimum and maximum threshold. The set of all first elbow points identified is used to identify the minimum threshold, while the set of second elbow points identified is used to detect the maximum threshold.

Step 4. Identifying minimum and maximum recommended thresholds

The proposed procedure identifies the minimum and maximum thresholds based on the elbow points detected by applying the following heuristics: (1) set the minimum recommended threshold to be the maximum threshold among the first set of elbow points detected; (2) set the maximum recommended threshold as the minimum threshold among the second set of elbow points detected.

When selecting a minimum threshold, prior research has described that it is "less critical to err with a threshold that is too large than to have it be too small" (Blignaut, 2009), as a threshold too low will erroneously classify gaze samples belonging to eye fixations as saccadic movements. To facilitate the selection of a minimum threshold, the proposed procedure identifies a minimum threshold between all *m* eye movement metrics investigated through a formulation of the heuristic

described by (Blignaut, 2009), which can be observe in equation (3.7). The minimum recommended threshold (z_{min}) is defined as the maximum threshold between the first set (i.e., from the eye movement metric partitioned to start from t_0 to f_1) of elbow (or knee) points identified among all m eye movement metrics (i.e., $e_0, \dots, e_m$). For example, assume that three eye movement metrics are investigated, number of eye fixations, the percentage of eye fixations in AOIs, and the percentage of gaze samples in eye fixations – all with at least one elbow point. In this case, e_0 would be the threshold at first elbow point identified for metric the number of eye fixations, e_1 would be the first elbow point identified for the percentage of eye fixations in AOIs, and e_3 would be the first elbow point identified for the percentage of gaze samples in eye fixations.

$$z_{min} = \max\{e_0, \dots, e_m\} \tag{3.7}$$

To select a maximum recommended threshold, the proposed procedure extends the heuristic used to select a minimum recommended threshold (Blignaut, 2009). Prior studies have described how selecting a threshold too large may lead to gaze samples belonging to saccadic eye movements to be classified erroneously as eye fixations (e.g., Salvucci & Goldberg, 2000), which can affect the visual scanpath created (Blignaut, 2009). However, what constitutes a threshold "too high" may vary between applications. To address, the proposed procedure identifies a maximum through the heuristic described in (3.8), extending the minimum threshold heuristic discussed by (Blignaut, 2009). The maximum recommended threshold (z_{max}) is defined as the minimum threshold between the second set (i.e., from the eye movement metric partitioned to start from the first inflection point to the second inflection point or t, if there is no second inflection point) of elbow (or knee) points identified among all k eye movement metrics investigated (i.e., $e_{1_2}, \dots, e_{k_2}$). If the eye movement metric has no second elbow point, the eye movement metric was not considered for the calculation of the maximum recommended threshold.

$$z_{max} = \min \{e_0, \dots, e_w\} \tag{3.8}$$

Step 5. Plot aggregated visual scan paths at the recommended thresholds

Visual scanpaths can be complex in nature, particularly in applications where eye movements are investigated for extended periods of time (e.g., Kang & Landry, 2014), which can make their visual representation challenging. To address this, the proposed procedure creates aggregated visual scanpaths by grouping consecutive eye fixations landing on the same AOI (Salvucci & Goldberg, 2000; Goldberg & Helfman, 2010a; Kang & Landry, 2014). Consider the expected visual scanpath shown in Figure 1(A). The eye fixations *A* and *B* can be grouped together into a singular eye fixation *D* as they both take place consecutively within the same AOI. Similarly, the eye fixations *E*, *H*, *I*, *J*, and *G* also take place consecutively within the same AOI, and can be grouped into an eye fixation *F*. As such, the visual scanpath shown in Figure 1(A), represented as the sequence *ABCEHIJG*, can be aggregated into the sequence *DCF*. Note that the visual scanpaths can be created not only over a period of time, but also for a specific number of eye fixations.

Although prior research has visualized aggregated visual scanpaths using the known location of the AOIs they landed on (e.g., Goldberg & Helfman, 2010b), the proposed procedure uses the aggregated position of the eye fixations. Considering the position of the eye fixations when visualizing may facilitate threshold selection in two ways. First, the aggregated eye fixation position highlights where in the AOI participants were fixating on, which can be particularly important when participants are instructed or expected to fixate on a specific portion of the AOI. Second, the shift in eye fixation positions as the thresholds increases can be observed. Higher thresholds classify more gaze samples as eye fixations, which can change the location of the eye fixations (Blignaut, 2009). Thus, using the average position of consecutive eye fixations can provide additional

information regarding the visual scanpath of the participant and how it is impacted by the thresholds.

Lastly, the proposed procedure investigates and compares aggregated visual scanpaths at the minimum and maximum thresholds for a sample of participants. Every threshold within the minimum and maximum threshold is considered to be an adequate threshold option. As such, only the impact at the minimum and maximum threshold needs to be visually and compared. Aggregated visual scanpaths are compared by identifying differences in the shape of the visual scanpath through visual observation, as well as by determining if any AOIs are missing between the visual scanpath.

Experiment

The proposed procedure was applied to the Random Saccade (RAN) task in the GazeBase dataset (Griffith et al., 2021). In this task, participants fixate and follow a bullseye target as it changes positions on the environment. Because the position of the bullseye target is recorded, and participants were instructed to follow the bullseye target, each participant has an expected visual scanpath that can be used to evaluate the accuracy of the minimum and maximum thresholds identified by the proposed procedure. The expected visual scanpaths, and the aggregated visual scanpaths created at the minimum, maximum, and two baseline thresholds were visually observed and compared. In addition, the normalized string-edit similarity between the expected visual scanpath and the aggregated visual scanpaths created at the minimum, maximum, and baseline thresholds were calculated.

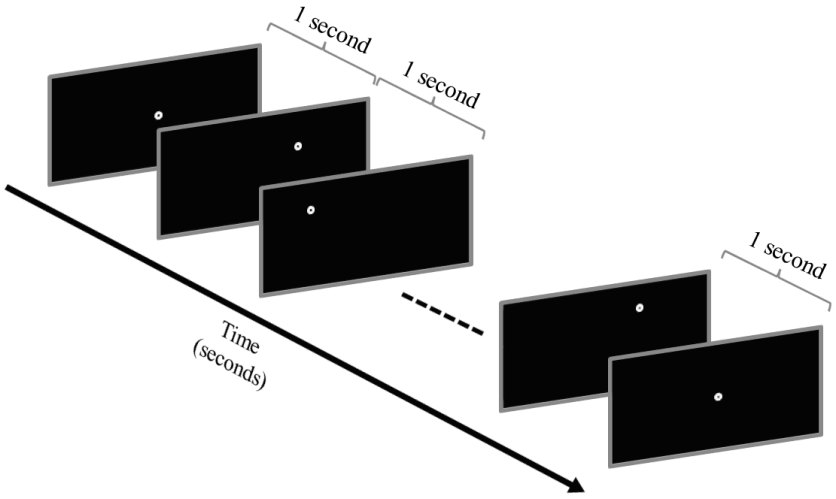
The section is structured in the following manner. First, key elements of the Gazebase dataset, summarized from the work of (Griffith et al., 2021) are explained. Second, the eye movement data processing steps are described. Third, the procedure and metrics used to evaluate the performance

of the proposed procedure are introduced. Fourth, and lastly, the data analysis procedures are described.

Random saccade task

A total of 322 college students (171 self-identify as male, 151 as female; average 21.99 years of age, *SD*: 4.22) participated in the RAN task. A representative example of the task can be observed in Figure 10. Participants were instructed to fixate and follow a bullseye target. Each target appeared on the screen for 1 second, after which it would change position. A total of 100 targets were presented to the participants. The bullseye targets were placed at a random locations in the environment. As such, the expected visual scanpath of each participant is different. Two exceptions were the start and final bullseyes, which were both placed at the center of the screen. During the task, monocular eye movements of the left eye were collected using the EyeLink 1000 (SR Research, Ottawa, Ontario, Canada) at a sample rate of 1000 Hz. Participants were seated 550 mm away from a 1680x1050 pixels (474 x 297 mm) computer monitor. The participants' heads were stabilized using a chin and forehead rest. Prior to the task, the eye tracker was calibrated to the participants eye movements following a 9-point calibration procedure. The collected gaze samples and target bullseye positions were converted to degrees of visual angle (dva).

Figure 10. Representative example of the RAN task conducted in the Gazebase dataset (Griffith et al., 2021).



Note. Participants were instructed to fixate and follow a bullseye target (white target) that appeared on the screen for 1 second prior to changing position. A total of 100 targets were presented to the participants. The position of each target, with the exception of the first and last targets at the center of the screen, was random. As such, each participant has a different expected visual scanpath.

Processing eye movement data

A total of 312 participants out of 322 participants were included in the data analysis. Participants with more than 10% of gaze samples missing were not included in the analysis. In the GazeBase dataset, the authors describe that gaze samples may be missing as a result of the participant blinking or partial occlusions of the eye (Griffith et al., 2021). Only 10 participants exceeded the 10% missing samples criteria.

To classify eye movements, the I-VT algorithm with a minimum eye fixation duration of 60 ms was set, used in prior implementations of the I-VT algorithm (Olsen, 2012). The threshold range was set at a maximum threshold (i.e., t) of 400 °/s in increments of 10 °/s (i.e., t_{step}).

In order to create visual scanpaths and calculate the eye movement metrics (i.e., percentage of eye fixations in AOIs) each bullseye target present in the task was considered as an AOI. A total of

100 unique AOIs were created as each of the targets appears in the environment once at a random location. The shape of the AOIs were defined as circle in order to match the shape of the bullseye targets. To account for the average visual angle error (0.5) reported for the eye tracker by the manufacturer (Griffith et al., 2021), as well as the highest median visual angle error error (1.6 dva) reported in the GazeBase dataset (Griffith et al., 2021), the size of the AOI (i.e., s_{max}) was set at a 2 dva radius. Increasing the size of the AOIs higher than the stimuli is a common approach to account for the visual angle error (e.g., Mandal & Kang, 2015; Kang et al, 2016).

Data analysis

The proposed procedure was implemented to identify the minimum and maximum thresholds recommended for the RAN task in the GazeBase dataset. The three-eye movement metric trends used to identify the minimum and maximum thresholds (the average number of eye fixations, the average percentage of eye fixations in AOIs, the average percentage of gaze samples in eye fixations) were visualized across the entire threshold range (i.e., 10 °/s to 400 °/s) and compared to their evaluation criteria. In addition, the Savitzky-Golay filter was used to fit the average eye movement metric trends to a 2nd degree polynomial using a window length of 10 samples, values used in prior eye tracking research (Nyström & Holmqvist, 2010; Ouzts & Duchowski, 2012).

The outputs of the proposed procedure, the aggregated visual scanpaths of a sample of two participants were plotted and compared when considering the first 15 AOIs and all AOIs. In addition, the string-edit similarity between the aggregated and expected visual scanpaths at each threshold was calculated, as well as the average number of missing AOIs in the aggregated visual scanpaths. In addition, participants were assigned into one of the following categories based on their string-edit similarities to the expected visual scanpath: (1) [0%, 40%]; (2) (40%, 60%]; (3) (60%, 90%]; (4) (80%,100%]. The threshold with the highest average string-edit similarity and

lowest average number of AOIs to the expected visual scanpath when considering all AOIs was identified.

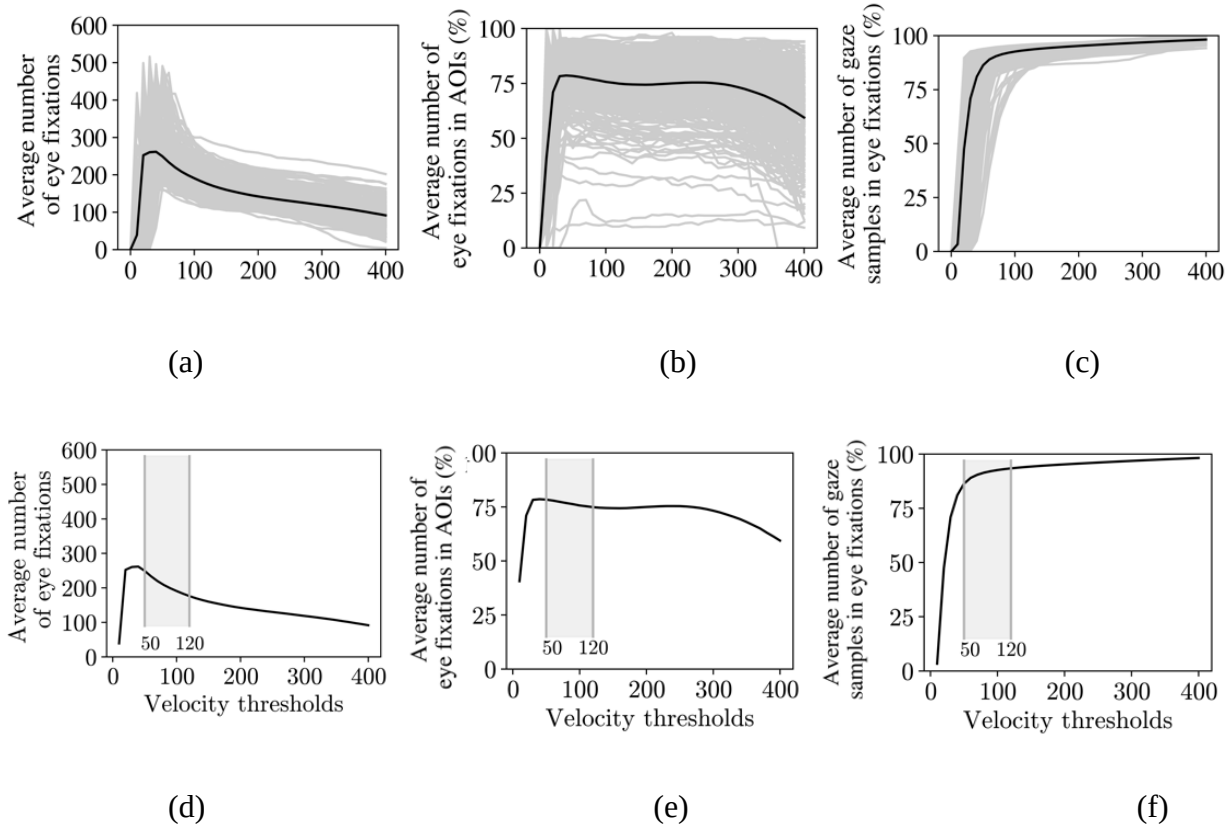
For each of these analysis, the recommended thresholds were compared to two baseline thresholds: 30 °/s and 350 °/s. The 30 °/s threshold has been set as the default threshold in prior implementations of the I-VT algorithm (e.g., Olsen, 2012), while 350 °/s is a threshold closer to the distribution of gaze velocities of saccadic movements (>300 °/s) (Salvucci & Goldberg, 2000), which would result in more saccadic movements classified as eye fixations.

Note that comparisons to the expected visual scanpaths are possible in this analysis procedure due to the controlled experimental task instructing participants to fixate on a specific stimulus.

Results

Figure 11 visualizes the average and participant eye movement metric trends across the range of thresholds evaluated 10 °/s to 400 °/s, and showcases the recommended minimum (50 °/s) and maximum (120 °/s) thresholds identified through the proposed procedure. The recommended thresholds contained (1) A high average number of eye fixations in AOIs (78.39% at the minimum threshold and 74.9% at the maximum threshold. The highest value achieved was 78.59%); (2) Most gaze samples classified as eye fixations (86.31% at the minimum threshold, 93.41% at the maximum threshold); (3) Led to a high average number of eye fixations (249.44 at the minimum threshold and 176.21 at the maximum threshold. The highest value achieved was 261.70). In addition, the eye movement metrics of participants are closely distributed around the average number of eye fixations and the average percentage of gaze samples in eye fixations. On the other hand, the distribution of participants for the percentage of eye fixations in AOIs eye movement metric had the most observable variation.

Figure 11. Average and participant eye movement metric trends: (1) number of eye fixation (a and d); the percentage of eye fixations in AOIs (b and e); (3) the percentage of gaze samples in eye fixations (c and f) across the range of threshold values evaluated.

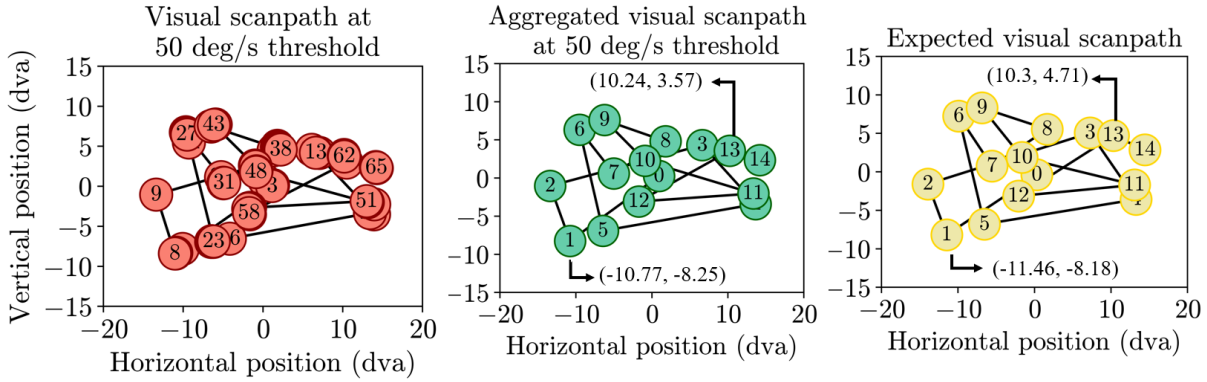


Note. The range of recommended thresholds identified by the proposed procedure, the minimum (50 °/s) and maximum (120 °/s) thresholds, is highlighted across figures d-f.

Figure 12 contains a representative example of the aggregation procedure of the first 15 AOIs in the visual scanpath of a participant at the recommended minimum threshold (50 °/s). In addition, the expected visual scanpath is included as reference. The visual scanpath contains all 58 eye fixations carried out by the participant, which the aggregation procedure simplifies into a total of 14 eye fixations. The coordinates eye fixations shared between the aggregated visual scan path and the expected visual scan path are very similar. Consider the eye fixations 1 and 13 highlighted in Figure 12. Eye fixation 1 was positioned at (-10.77, -8.25) in the aggregated visual scanpath, while its expected position was (-11.46,-8.18), while eye fixation 13 was positioned at (10.24, 3.57) and

its expected location was (10.3, 4.71). Overall, the aggregated visual scanpath at the recommended minimum threshold value of 50 °/s closely approximates the shape and positions of the expected visual scanpath by placing the eye fixations in proximity.

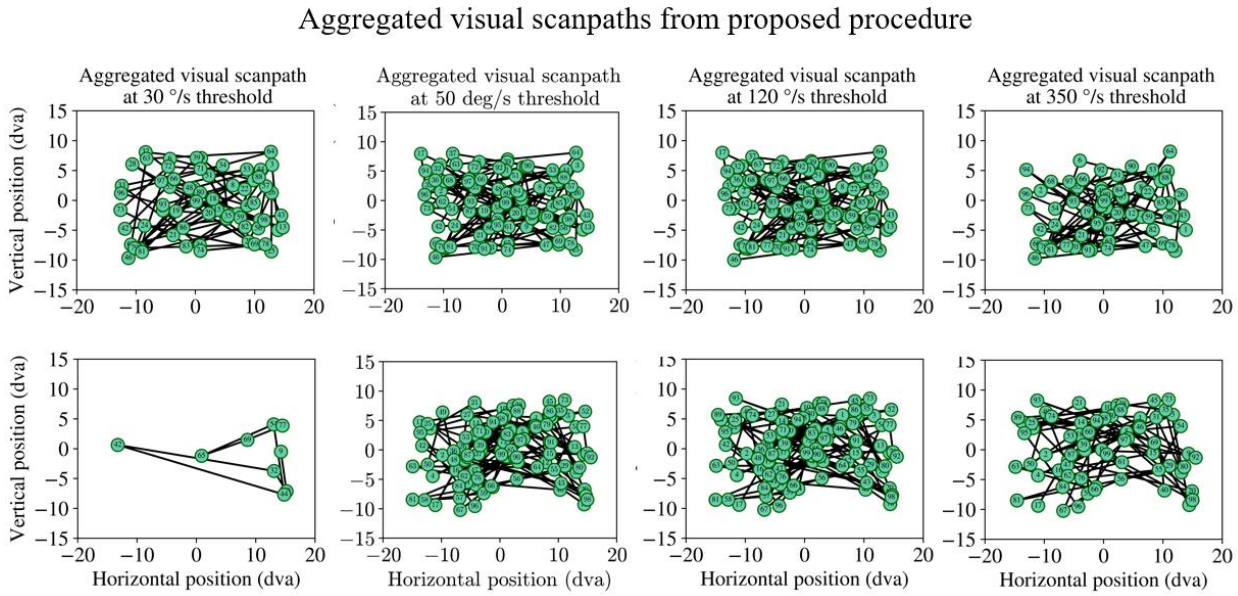
Figure 12. Representative example of a visual scanpath created at the minimum eye fixation threshold (red circles) transformed into an aggregated visual scanpath (green circles) alongside the expected visual scanpath carried out by the participant (yellow circles).



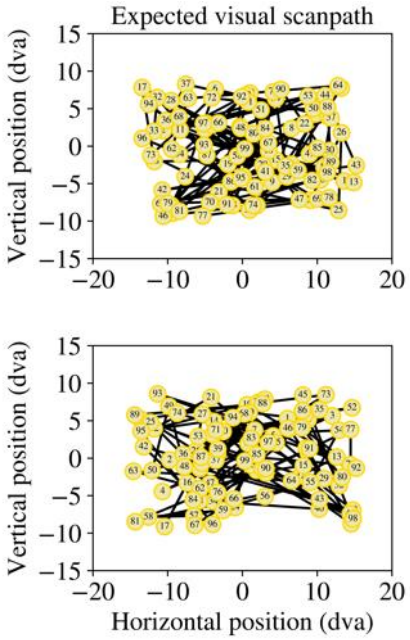
Note. Only the first 15 AOIs were considered in this example. The number inscribed within each circle represents the order in which the eye fixations took place. The coordinates of AOI 1 and AOI 13 are highlighted for both the aggregated visual scanpath and the expected visual scanpath.

To evaluate the recommended thresholds, Figure 13 showcases the aggregated visual scanpaths for two participants for the entire task at the minimum (50 °/s), maximum (120 °/s), and baseline thresholds (30 °/s and 350 °/s). The recommended thresholds had fewer AOIs missing than the baseline thresholds (Table 1) and maintained similar visual scanpath shapes.

Figure 13. Two representative examples (top and bottom rows) of the impact of threshold selection for all AOIs (green circles) at the minimum (50 °/s), maximum (120 °/s), and baseline thresholds (30 °/s and 350 °/s) for two participants (a).



(a)



Note. In addition, the expected visual scanpath for both examples are highlighted (b). The number inscribed within each circle represents the order in which the eye fixations took place.

Table 1. The number of AOIs missing and the string-edit similarity at the minimum (50 °/s), maximum (120 °/s), and baseline thresholds (30 °/s and 350 °/s) to the expected visual scanpath for the aggregated visual scanpath examples shown in Figure 13.

Visual scanpath example	Thresholds	Number of AOIs missing	String-edit similarity
Participant A	30 °/s	33 AOIs	67%
	50 °/s	2 AOIs	98%
	120 °/s	1 AOIs	99%
	350 °/s	33 AOIs	67%
Participant B	30 °/s	91 AOIs	9%
	50 °/s	3 AOIs	97%
	120 °/s	0 AOIs	100%
	350 °/s	32 AOIs	68%

Note. The visual scan paths of participant A are shown in the top row of Figure 13, while the visual scan paths of participant B are included in the bottom row of Figure 13.

Across both representative examples, the recommended thresholds had fewer AOIs missing and more accurately represented the visual scan paths the participants were instructed to follow. More specifically, for participant B (Figure 13(bottom row), the 30 °/s threshold created a visual scanpath that contained only 9 AOIs, while the 350 °/s had 32 AOIs missing. On the other hand, the recommended thresholds had either 3 AOIs missing, or did not have any AOIs missing at all. , similarly complex had more complete visual scan paths, as they missed fewer AOIs. For participant A, both recommended thresholds had AOIs missing (2 at 50 °/s and 3 at 120 °/s) but maintained similar visual scanpath shapes. The two baseline thresholds had a higher number of AOIs missing (33 AOIs), resulting in different visual scanpath shapes.

To evaluate the performance of the thresholds to the entire participant population, the average number of AOIs missing and the string-edit similarity to the expected visual scanpath were

calculated for the aggregated visual scanpaths created at the minimum, maximum, and baseline thresholds for all 312 participants (Table 2).

Table 2. Average number of AOIs missing and average string-edit similarity to the expected scanpath at the aggregated visual scanpaths for the minimum, maximum, and baseline thresholds across participants and all AOIs in the task.

			Participants grouped by string-edit similarity to the expected visual scanpath			
Thresholds	**Avg. number of AOIs missing**	**Avg. string-edit similarity (%)**	**[0%, 40%]**	**(40%, 60%]**	**(60%, 80%]**	**(80%, 100%]**
30 °/s	12.61	87.39%	24	10	26	252
50 °/s	5.32	94.68%	2	6	17	287
120 °/s	5.22	94.78%	3	3	19	287
350 °/s	31.96	68.04%	39	63	112	98

More specifically, the recommended minimum and maximum thresholds were able to achieve higher average string-edit similarities (94.68% at 50 °/s and 94.78% at 120 °/s), and thus missing fewer AOIs (5.32 AOIs at 50 °/s and 5.22 AOIs at 120 °/s) than the baseline thresholds (87.39% and 12.61 for 30 °/s; 68.04% and 31.96 AOIs for 350 °/s). In addition, at the 30 °/s threshold, 60 participants (19.23%) had a string-edit similarity of 80% or less, while the recommended thresholds only had 18 (5.76%) and 25 (8.01%) participants with less than 80% similarity. At the 30 °/s threshold, 24 participants (7.69%) had a string-edit similarity less than or equal to 40%, while at the recommended 50 °/s and 120 °/s thresholds, the number of participants was 2 (0.64%) and 3 (0.96%), respectively. At the 350 °/s, 214 participants (68.58%) had a string-edit similarity of 80% or less, while 39 participants (12.5%) had a similarity lower than or equal to 40%.

Discussion

The proposed procedure was capable of automatically identifying and recommending a minimum and maximum thresholds (50 °/s and 120 °/s) that satisfied the evaluation criteria for the three eye movement metrics selected. More specifically, it classified most gaze samples as eye fixations and led to a high number of eye fixations that took place in AOIs. In addition, the visualization aggregation procedure created aggregated visual scanpath that maintained both the shape and position of the visual search carried out by participants during the RAN task of the GazeBase dataset. The accuracy of the aggregated visual scanpaths was evaluated through visual observation, as well as by calculating the sting-edit similarity to the expected visual scanpaths. The recommended thresholds (50 °/s and 120 °/s) had higher string-edit similarities on average, and thus fewer AOIs missing, than the baseline thresholds used for comparison (30 °/s and 350 °/s). Overall, these contributions may help both researchers and eye tracking practitioners to select better performing thresholds through an automated procedure that recommends thresholds and visualizes how they are impacted across thresholds in cases when an expected visual scanpath is not available.

A reason behind the performance difference observed is that the baseline thresholds were ‘too high’ (i.e., 350 °/s) or ‘too low’ (i.e., 30 °/s), which may have led to a higher number of gaze samples being misclassified. More specifically, a threshold ‘too high’ may begin to misclassify gaze samples belonging to saccadic movements as eye fixations. The misclassification may lead to multiple eye fixations being combined into a singular eye fixation, which can potentially remove the resulting eye fixation from the bounds of the AOI. Such an effect can be observed in Figure 13, where the number of eye fixations in AOIs in the visual scanpath goes from 100 at the 120 °/s threshold to 68 at the 350 °/s threshold. On the other hand, at a threshold ‘too low’, gaze samples

belonging to an eye fixation may be misclassified as an eye fixation, which would lead to fewer eye fixations in the visual scanpath, as observed in Figure 13. The number of eye fixations in AOIs in the visual scanpath increases from 9 at 30 °/s to a total of 100 at 50 °/s threshold.

The baseline thresholds were selected from available literature employing the I-VT algorithm; however, the thresholds used in eye movement detection algorithms that can accurately classify eye movements can vary between studies based on a number of factors, such as due to individual differences between participants (e.g., Blignaut & Beelders, 2009). As shown in Table 2, the groups of participants based on their sting-edit similarity to the expected visual scanpath differed across the baseline and recommended thresholds. More specifically, at the 30 °/s threshold, 60 participants (19.23%) had a string-edit similarity of 80% or less to their expected visual scanpath. Such low similarity values may affect the researchers' interpretation of the visual search strategy carried out by the participant in the context of a task. However, through the application of the proposed procedure, the lowest recommended threshold identified was 50 °/s. The increase in threshold value from the 30 baseline °/s led to only 17 participants (5.44%) having a similarity lower of 80% or lower. As such, selecting a threshold based on the eye movement data collected from participants allowed us to identify a threshold that was capable of increasing the string-edit similarity of 35 participants (11.21%) past 80%.

In addition, the proposed visual scanpath aggregation procedure highlighted how the visual scanpaths of participants can vary across thresholds in terms of shape and the eye fixation that are present or absent. Understanding the impact of thresholds selection can be crucial when investigating key metrics such as the time to first fixate on an AOI or the number of eye fixations in AOIs, important measures in healthcare research (e.g., Van der Gijp et al., 2017; Brunyé et al., 2020). At the 30 °/s threshold, most of the eye fixations of Participant B in Figure 13 are not

included in the visual scan path, and one may erroneously consider that the participant never fixated on a specific area of the screen during the task. However, at the 50 °/s threshold, one can observe that the participant had a more thorough and complex visual scan path that covered a much larger area. Thus, arriving at a very different interpretation of the visual search strategy of the participant due to the impact of the threshold selected. Visually observing the visual search strategies of participants can help researchers make better decisions regarding what thresholds to select.

Overall, our results agree with Orquin & Holmqvist (2018) idea of 'hidden defaults', in which using parameters from an experimental design available in the literature, such as the thresholds applied, without any validation in place may lead to suboptimal results (Orquin & Holmqvist, 2018). In the present study, this can be seen as the selection of thresholds for the I-VT algorithm from comparative literature led (i.e., 30 °/s) to suboptimal results. However, when applying the proposed procedure to automatically identify thresholds that satisfy the evaluation criteria, we were able to identify thresholds that led to much more accurate visual scanpaths. The proposed computational procedure may help researchers select thresholds that are a better match to their experimental design (e.g., participant population, task design, etc) when expected visual scanpaths are not available.

Chapter 4 – Algorithms to determine accurate eye fixations when the ideal visual scan path is known vs. unknown

In this chapter, we introduce two methods, namely between-participant and within-participant similarity, that were designed to approximate the relationship between thresholds and string-edit similarity shown in Figure 8.

The chapter is structured in the following manner. First, we introduce both methods, including their underlying rationale alongside worked out examples to showcase how they are calculated. Second, we apply both methods to two eye movement data sets, one containing a simple task in which participants followed a dot changing locations on a computer screen from open-source Gazebase dataset (Griffith et al., 2021), and the other containing a much more complex task involving air traffic controllers issuing clear to take off clearances in a high-fidelity simulator. Third, and lastly, we discuss the insights derived from the application of both methods and their feasibility to determine accurate thresholds for eye movement detection algorithms.

Proposed approach

Between-participant similarity

A group of participants with similar characteristics, such as expertise (Underwood, 2007), may carry out a visual search of the environment that may be very similar to each other under controlled experimental conditions, such as when they are given the same task and set of instructions (DeAngelus & Pelz, 2008; Borji & Itti, 2014). One of the first studies to highlight this behavior among participants was the work of Yarbus (1967), who found that participants viewing the same painting but given a different set of instructions, such as “give the ages of the people” and “remember the clothes worn by the people”, showed different visual scan paths over the same

stimulus. In other words, the visual scan paths of participants were dependent upon the instructions given, suggesting that eye movements may be influenced by top-down task demands(DeAngelus & Pelz, 2008; Borji & Itti, 2014).

Under such circumstances, it might be possible that potential differences in the visual scan path sequences between participants to may be attributed to inaccuracy the thresholds selected. Thresholds with accurate performance might lead to higher similarities between scan path sequences of participants, as they accurately represent the eye movements of participants, while thresholds too low or too high may have lower similarities, primarily due to missing eye fixations and/or the presence of unlikely transitions in scan path sequences.

Based upon this assumption that we may be able to quantify the impact of threshold selection by comparing visual scan paths between participants, we adapt the between-participant scan path similarity metric to evaluate the string-edit similarity between participants. Comparing the similarity between participants is a common practice in eye tracking research (Duchowski et al., 2010, Anderson et al., 2015). In addition, other approaches to evaluate threshold performance, such as when comparing visual scan path sequence similarity to an ideal visual scan path of each participant, and then averaged similarity values between participants (Blignaut & Beelders, 2009). However, in our implementation, the between-participant similarity is calculated by comparing the visual scan paths created between participants across all thresholds evaluated in order to approximate the relationship between string-edit similarity and threshold values.

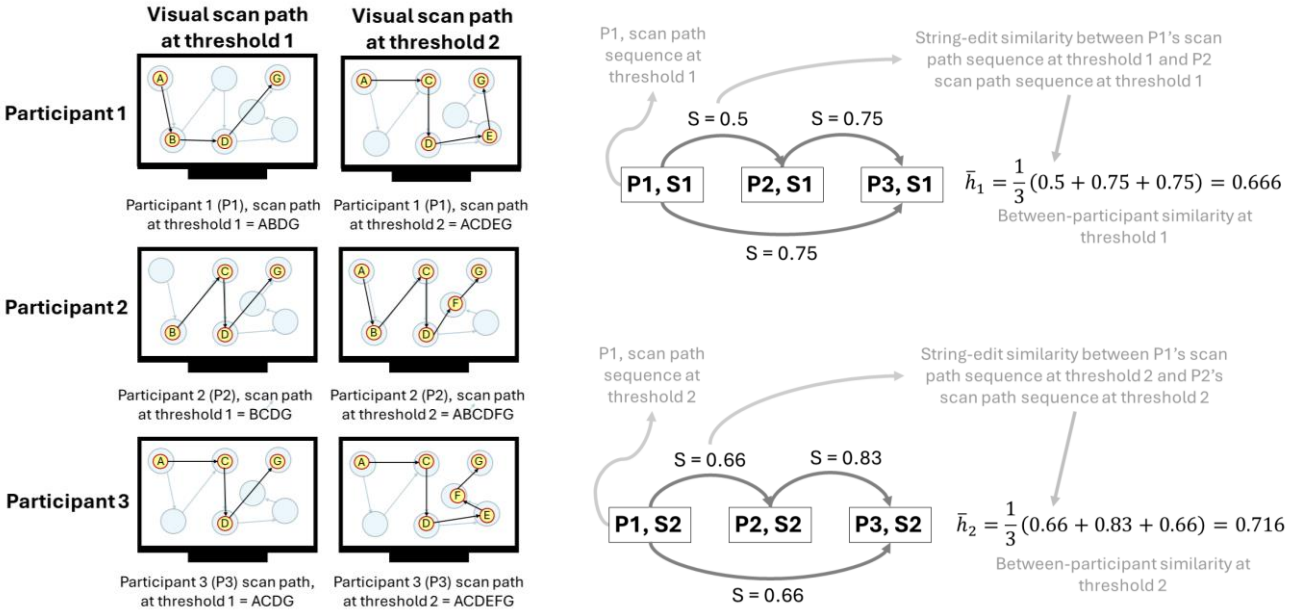
The calculation of between-participant similarity at each threshold is as follows. Let $X_i = \{x_1, x_2, \dots, x_j \mid 1 \leq j \leq p, j \in \mathbb{N}\}$, where X_i contains the set of scan path sequences at the ith threshold for all p participants and x_j represents the scan path sequence of the jth participant. From the set X_i, one can define all m two-scan path sequence combinations between participants

at the i th threshold as $W_i = \binom{X_i}{2} = \{(x_j, x_k) \mid 1 \leq j < k \leq p, j \in \mathbb{N}, k \in \mathbb{N}\}$. For all combinations, the average string-edit similarity $\bar{h}_i$ at the threshold ith between all participants can then be calculated using equation (4.1).

$$\bar{h}_i = \frac{1}{m} \sum_{(x_j, x_k) \in W_i} S(x_j, x_k) \quad \forall\, i \in \{1,2 \ldots, t\} \tag{4.1}$$

Figure 14(a) showcases an example of the between-participant similarity calculated for the scanpath sequences created at two thresholds (1 and 2) carried out by three participants (P1, P2, and P3). Equation (2.2) is applied in Figure 14(b) to compare all possible combinations of 2 scan path sequences between participants. The results indicate that the threshold with the highest average string-edit similarity between participants would be threshold B (0.716 at threshold B vs 0.666 at threshold A). As such, based on these results, one ought to select threshold B as it is the most accurate threshold.

Figure 14. Example of between-participant method calculations for three participants who were instructed to follow the movement of a blue dot on a display.



(a) (b)

Within-participant similarity

An alternative to quantify the impact of thresholds on the visual scan paths of participants, particularly when these may be very different from each other in more complex environments or tasks, might be to compare the scan path sequences of one participant created at a threshold to every other threshold. As mentioned previously, scan path sequences created at threshold too low or too high might have missing eye fixations, as well as contain transitions between eye fixations that might not have taken place. On the other hand, scan path sequences created at an acceptable threshold are not as impacted by either missing eye fixations or unlikely eye movement transitions. Therefore, one might expect that the similarity between the scan path sequences created at an acceptable threshold, and one created at threshold too low, or too high, to be lower than the similarity between scan path sequences created at acceptable thresholds.

Based on this assumption, we propose a novel within-participant similarity metric that calculates the average string-edit similarity of the scan path sequence of one threshold to the scan path sequences created at every other threshold. The calculation of the proposed within-participant similarity for a single participant is as follows. For a single participant, let $F = \{f_1, f_2, \dots, f_i \mid 1 \leq i \leq t, i \in \mathbb{N}\}$ represent the set of scan paths sequences created at each threshold, where f_i represents the scan path sequence created at the ith threshold, and t the total number of thresholds. The average string-edit similarity v_i of the scanpath sequence at threshold i to the scan path sequence at every other threshold can be calculated using equation (4.2).

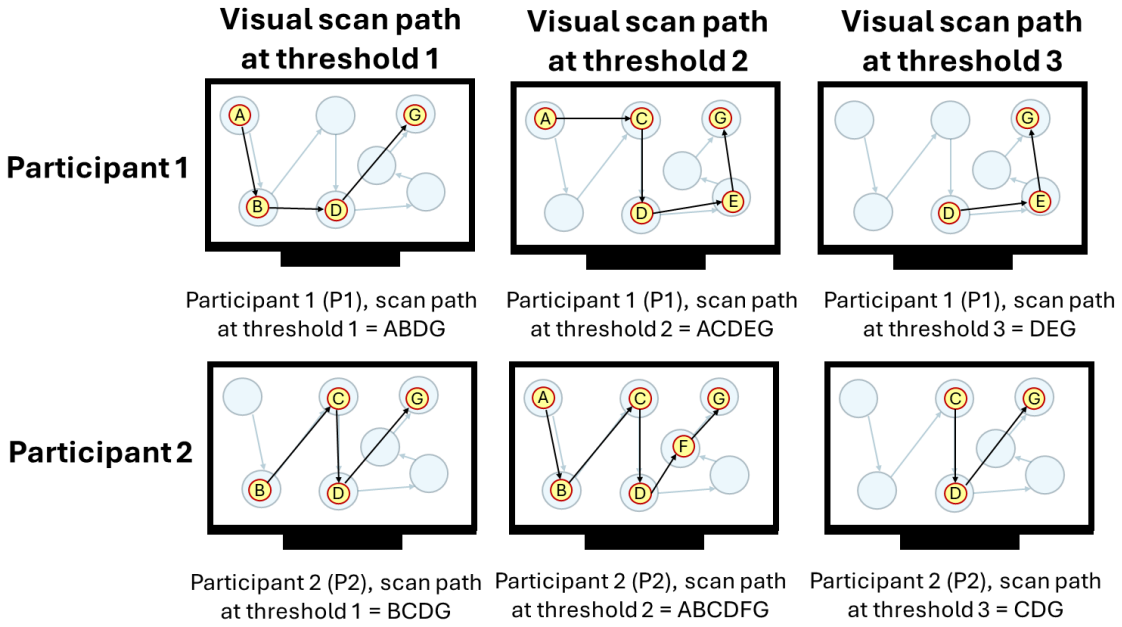
$$v_i = \frac{1}{t-1} \sum_{j=1; i \neq j}^{t} S(f_i, f_j) \qquad \forall\, i \in \{1,2 \dots, t\} \tag{4.2}$$

Here, S represents the string-edit similarity function (explained in detailed below in the methods section), f_i and f_j the scan path sequences at thresholds i and j in the set F. Note that the average is calculated by subtracting $t-1$, instead of simply t, as there is a total of $t-1$ similarity calculations (the similarity between one scan path sequence to itself is never calculated (i.e., $i \neq j$)). The outputs of equation (4.2) can be averaged across participants to identify the threshold that creates the most similar scan path sequence for all participants. Let $V = \{v_{ij} \mid 1 \leq i \leq t, 1 \leq j \leq p, i \in \mathbb{N}, j \in \mathbb{N}\}$ contain the set of average string-edit similarities at each threshold for each participant, where v_{ij} represents the average string-edit similarity at the ith threshold for the jth participant, and p the total number of participants. The average string-edit similarity $\bar{v}_i$ of the scan path sequence at threshold i to the scan path sequences created every other threshold across participants can be calculated using equation (4.3) below.

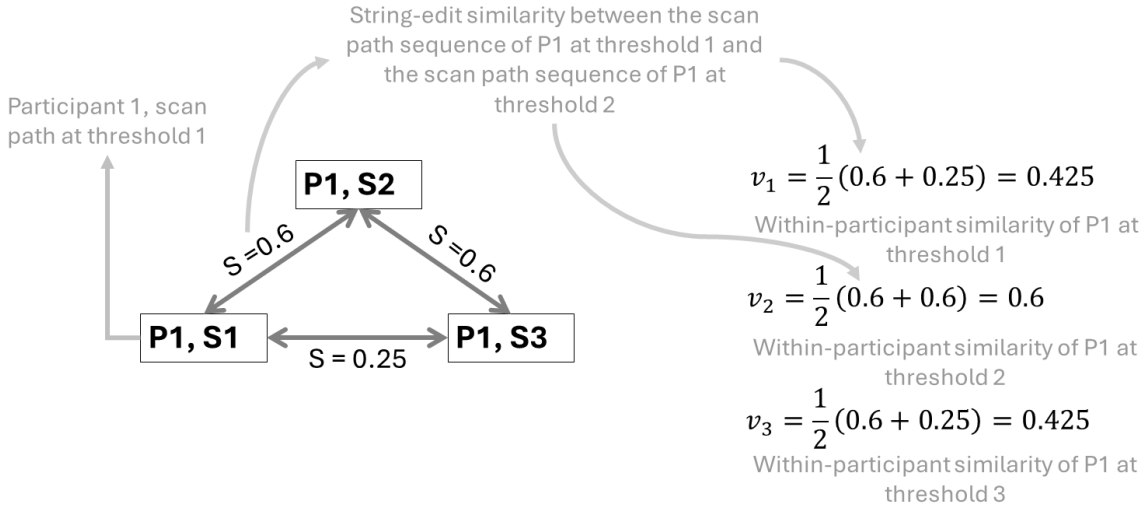
$$\bar{v}_i = \frac{1}{p}\sum_{j=1}^{p} v_{ij} \qquad \forall\, i \in \{1,2 \ldots, t\} \tag{4.3}$$

Figure 15(a) shows an example of the within-participant similarity being calculated for the scan path sequences at three thresholds (A, B, and C) carried out by two participants (P1 and P2). Equation (2.2) is applied in Figure 15(b and c) to calculate the threshold that creates the scan path sequence most similar to the scan path sequences at every other threshold for each participant. Lastly, equation (2) is applied in Figure 15(d), showcasing that the threshold with the highest average similarity to every other scan path sequence across the two participants would be threshold B (0.59 at threshold B vs 0.565 at threshold A and 0.525 at threshold C). As such, based on this example, the researcher should select threshold B.

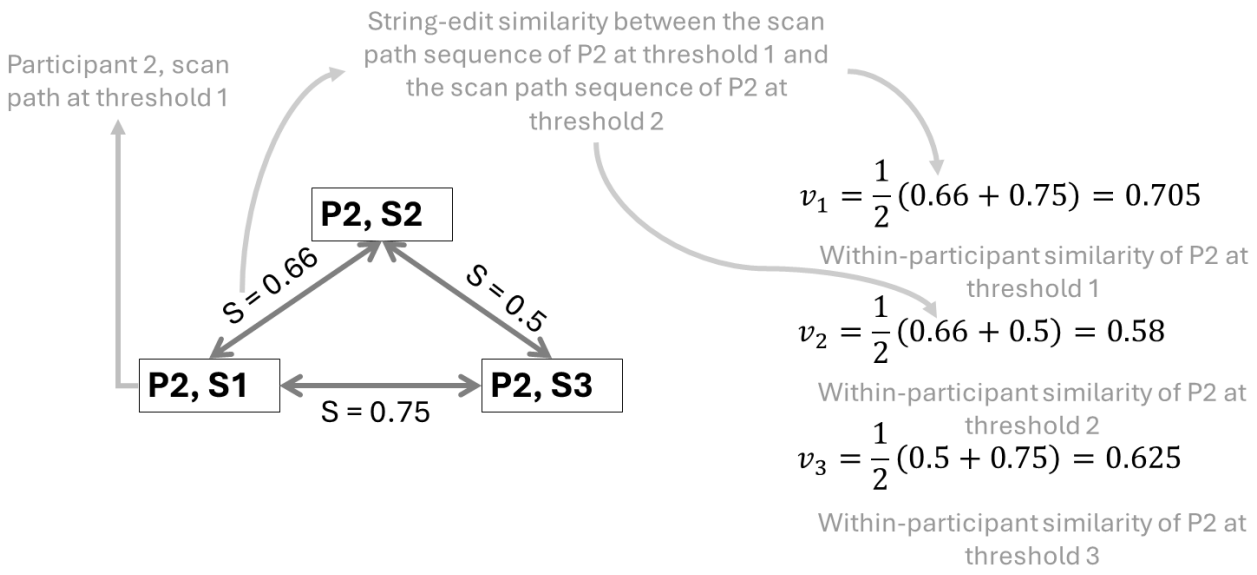
Figure 15. Representative example of the calculation of the average string-edit similarity at each threshold across two participants who were instructed to follow the movement of a blue dot on a display.



(a)



(b)



(c)

Within-participant similarity of P1 at threshold 1

Within-participant similarity of P2 at threshold 1

$$\bar{v}_1 = \frac{1}{2}(0.425 + 0.705) = 0.565$$

Within-participant similarity at threshold 1

Within-participant similarity of P1 at threshold 2

Within-participant similarity of P2 at threshold 2

$$\bar{v}_2 = \frac{1}{2}(0.6 + 0.58) = 0.59$$

Within-participant similarity at threshold 2

Within-participant similarity of P1 at threshold 3

Within-participant similarity of P2 at threshold 3

$$\bar{v}_3 = \frac{1}{2}(0.425 + 0.625) = 0.525$$

Within-participant similarity at threshold 3

(d)

Experiments

In this section, we describe the two eye movement datasets analyzed to showcase the two methods introduced in the previous section: the Random Saccade Task (RAN) from the Gazebase dataset and an air traffic control task collected by the authors. The first dataset was selected as it contains a task with an ideal visual scan path, as participants were instructed to fixate and follow a dot changing positions on a computer display in a pre-determined manner. The second dataset was selected as it contains the eye movements of air traffic controllers in a more complex and dynamic environment, where the eye movements between participants might be different.

Experiment 1 – Random Saccade Task

The participant and eye movement processing steps are the same as those described in Chapter 3, as the random saccade task (RAN) available in the open source Gazebase dataset (Griffith et al., 2021) is used again. Thus, we start this section by discussing the data analysis.

Data analysis

The between-participant and within-participant scan path sequence methods were calculated for all participants at each gaze velocity threshold ranging from 10 °/s and 400 °/s in increments of 10 °/s. The string-edit similarity between the scan path sequence at each gaze velocity threshold to the ideal scan path sequence participants were instructed to follow was calculated for each participant.

Spearman's rank correlations (r_s) were calculated to evaluate the strength of the association between the string-edit similarity values at each proposed method, and the string-edit similarity values to the ideal scan path sequence. A significance level of $\alpha = 0.05$ was used for the statistical test to evaluate whether the correlations were statistically significant.

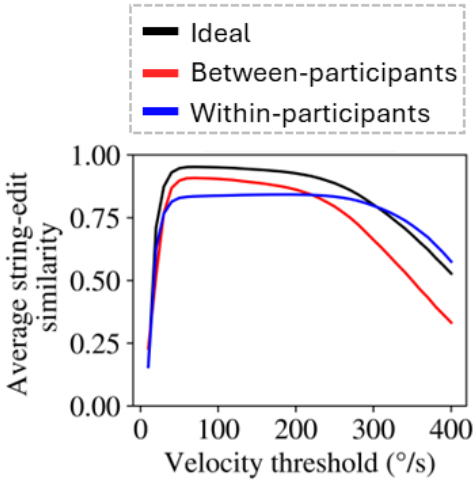
Accurate thresholds were identified by visually observing visualizations of the string-edit similarity values created for the two methods, as well as the comparisons to the ideal visual scan path sequence, over the range of thresholds investigated. More specifically, the thresholds were identified by visually observing a high and stable region of similarity values, which both researchers agreed upon. The upper and lower bounds of the range of acceptable thresholds identified by each method were compared to those found for the ideal scan path sequence.

Results

The average string-edit similarity values for both the between-participants (r_s = 0.99, p-value < 0.01) and within-participants (r_s = 0.77, p-value < 0.01) similarity trends were highly correlated with the ideal similarity values (Figure 16(b)). In addition, both methods were highly correlated with each other (r_s = 0.77, p-value < 0.01).

Figure 16(a) showcases the trends of average string-edit similarity for each method and the ideal similarity values followed similar trends across threshold values. Across these trends, one can observe a sharp increase in the similarity scores at low thresholds values, followed by a period of high and stable similarity values, after which the similarity values continuously decrease. More specifically, the ideal trend increased sharply until approximately threshold 50 °/s (0.946) and began to rapidly decrease at 200 °/s (0.926), the between-participant trend quickly increased until threshold 50 °/s (0.898) and began to decrease at around 180 °/s (0.875), and lastly, the within-participant trend rapidly increased until approximately threshold 50 °/s (0.828) and remained stable until the threshold 230 °/s (0.839).

Figure 16. Plot of the string-edit similarity over the range of thresholds evaluated (10 °/s to 400 °/s) for each method and the ideal scan path sequence string-edit similarities, as well as the Spearman correlations between the trends.



(a)

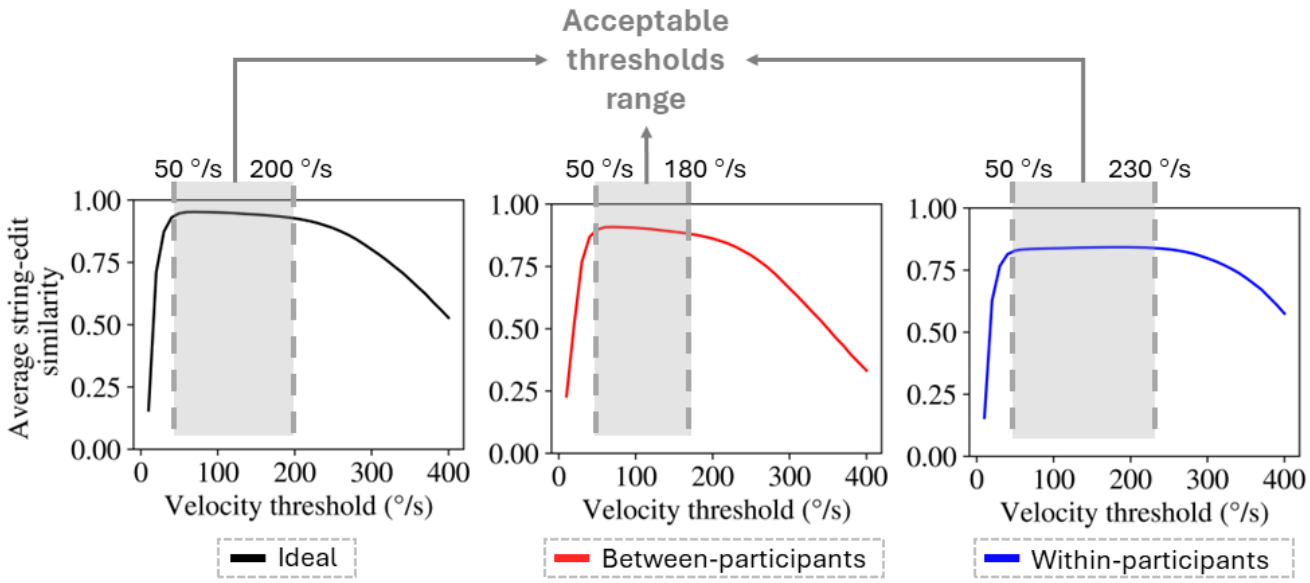
Method	1	2	3
1. Ideal	-		
2. Between-participants	0.99 **	-	
3. Within- participants	0.77 **	0.77 **	-

* denotes *p-value* < .05, ** denotes *p-value* < .01

(b)

Figure 17 showcases the accurate threshold ranges identified from the between-participant trend, the within-participant trend, as well as the trend comparing to the ideal visual scan path. The accurate threshold range selected for the between-participant method was from 50 °/s to 180 °/s, the within-participant threshold range was from 50 °/s to 230 °/s, and the threshold range for the ideal similarity values was between 50 °/s to 200 °/s. Note that the threshold range identified via the within-participant method was 30 °/s larger than the range identified when comparing to the ideal scan path sequence trend. The threshold ranges identified using both methods largely overlap with the thresholds determined for the ideal visual scan path.

Figure 17. Plot of the string-edit similarity over the range of thresholds values evaluated (10 °/s to 400 °/s) for each method. The accurate threshold ranges identified for each method are highlighted in gray.



In addition, the threshold ranges identified with the between-participant and within-participant methods had an average similarity score of 90% or higher on the ideal visual scan path trend (Table 3). In more detail, the thresholds range identified by the within-participant had 0.946 (50 °/s) and 0.907 (230 °/s) average similarity scores to the ideal scan path sequence. For the between-participants method, the average similarity values at the bounds of the threshold range identified were 0.946 (50 °/s) and 0.934 (180 °/s).

Table 3. String-edit similarity values when comparing to the ideal scan path sequences at the lower and upper bounds of the threshold range identified through the within-participants and between-participants method.

Method	Lower bound threshold (ideal scan path sequence similarity)	Upper bound threshold (ideal scan path sequence similarity)
Between-participants	50 °/s (0.946)	180 °/s (0.934)
Within-participants	50 °/s (0.946)	230 °/s (0.907)

Experiment 2 – Local air traffic controllers clear to take off clearances

The two proposed approaches were applied to identify a threshold capable of creating accurate scan path sequences to an air traffic control task in a high-fidelity tower cabin simulator. Due to the complexity of the task, ideal scan path sequences for each participant are not readily available, and thus were created manually. Furthermore, the visual scan path of each participant can be different due to the effects of the task and environment, unlike in the previous experiment, where participants were instructed to follow the stimulus presented that appeared in a pre-determined order.

Participant and apparatus

A total of 14 retired local controllers, employed (at the time of the experiment) as instructors by the Federal Aviation Administration's (FAA) Academy in Oklahoma City, Oklahoma, with an average of 26 years of experience (range between 10 and 42 years) participated in the experiment.

The controllers managed simulated landing and departing air traffic on a high-fidelity Adacel tower simulator used at the FAA's Civil Aeronautical Medical Institute (CAMI) in Oklahoma City, Oklahoma. Twelve 55" HD (1080p) monitors, wrapped greater than 180° *around the participant,* were used to simulate the out the window view of tower cabin. The simulators included flight strips of the aircraft in the scenarios, as well as working Bright Radar Indicator Terminal Equipment (BRITE) and Airport Surface Detection (ASDE) radar displays. The participants issued verbal clearances to the simulated aircraft using a headset.

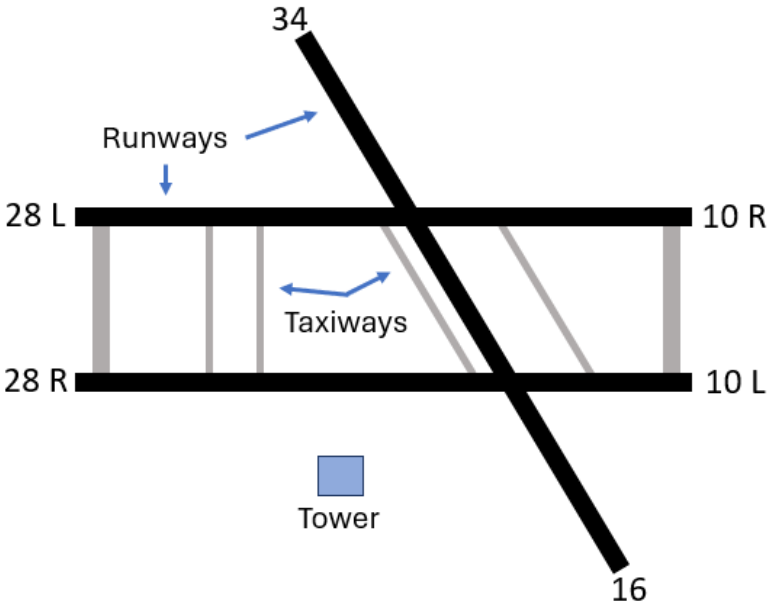
The Tobii Pro Glasses II (100 Hz) were used to capture participant eye movements (equipped with prescription lenses as necessary). In addition, the Tobii Pro Lab software was used to visually observe the recorded video from the participant's point of view that contained the raw gaze data

overlaid in order to identify ideal scan path sequences. In addition, the Tobii Pro Lab software was used to apply the I-VT algorithm in order to identify eye fixations and saccadic movements.

Task and scenario

Participants were tasked with managing landing and departing air traffic as local controllers during a scenario that lasted approximately 22 minutes. The scenario had high visibility conditions during daylight hours. The airport layout consisted of two parallel active runways (28L/10R and 28R/10L), as well as one crossing active runway (34/16) (Figure 18). The scenario contained 19 arriving aircraft and 14 departing aircraft, totaling 33 aircraft. Due to the dynamic nature of the environment and the use of a high-fidelity simulator, the order and time at which participants issued clearances to aircraft might be different. The scenario stopped once the 22-minute mark was reached, which might have occurred prior to all aircraft being issued a clearance.

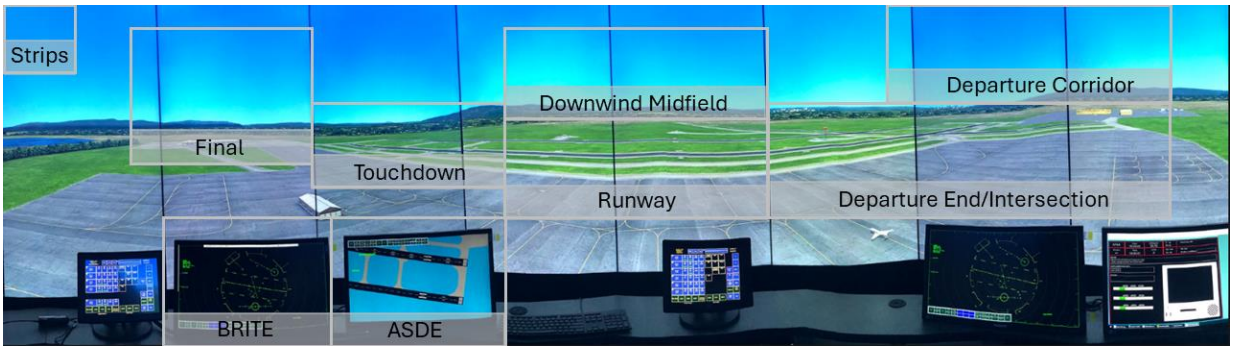
Figure 18. Diagram of airport layout managed by the participants.



Note. Three active runways were present in the experiment: two parallel runways (28L/10R and 28R/10L) and one crossing runway (34/16). The blue square represents the approximate location of the tower in the simulated environment. Two runways and two taxiways are highlighted.

AOIs in the airport environment (Figure 19) were identified and created in collaboration with subject matter experts based on their operational significance. The AOIs consisted of known hotspots in tower air traffic control (Crutchfield et al., 2021), such as the ends of an active runway (Touchdown AOI and Departure/End Intersection AOI). In addition to the active runways themselves (Runway AOI), the ASDE (ASDE AOI) and BRITE (BRITE AOI) radars. Furthermore, other AOIs considered were the arrival corridor that aircraft must fly through during their approach to the airport (Final AOI), the downwind area that arrival aircraft must cross prior to their final approach (Downwind Midfield), as well as the departure corridor (Departure Corridor AOI), where aircraft must fly through when they take off from airport. Lastly, gazes on the flight strips, which contain information regarding the aircraft present in the scenario (e.g., callsign arriving or departing from airport, etc), as well as gazes on notes written by participants on a notepad (e.g., writing the sequence of aircraft departing) were manually mapped using the Tobii Pro Lab software to the upper left corner (Strips AOI) of the image.

Figure 19. The high-fidelity tower control simulator used in the experiment with the areas-of-interest (AOIs) highlighted (gray squares).



Note. Gazes that took place on information available to the participants not included in the snapshot, such as the flight strips or a notepad the participant used to write down aircraft information were manually mapped to the upper-left corner of the image (i.e., Strips AOI). Because such information (i.e., flight strips) were not present in the snapshot, the Tobii Pro Lab Software could not automatically the gazes to the AOI.

Processing eye movement data

The eye movements of the controllers were analyzed during the time period they issued a clear to take off clearance, as prior research has suggested that that is the time where controllers are more likely to show focused attention (Manske & Schier, 2015). The period of time to issue the clear to take off clearance was defined as the time from when the aircraft states that it is ready for departure to when the controller finishes issuing the verbal clear to take off clearance to the aircraft. On average, the time to issue the clear to take off clearance was approximately 24.8 seconds across participants.

The ideal scan path sequence of each participant while issuing the clear to take off clearance was manually identified, created, and agreed upon by the researchers by visually observing the recorded video of the participant's point-of-view with their gaze overlaid. In other words, the ideal scan path sequences were obtained by observing each raw gaze data point collected by the eye tracker (Navarro et al., 2021). For example, if the participant observed the Touchdown AOI, followed by Runway AOI, the Touchdown AOI again, and lastly the Final AOI, all while they are issuing the clear to take off clearance, then the ideal scan path sequence would be Touchdown -> Runway -> Touchdown -> Final.

The I-VT algorithm implemented in Tobii Pro Lab was used to identify eye fixations and saccadic movements in order to create scan path sequences. All settings of the algorithm were left at their default values (e.g., minimum eye fixation duration of 60 ms) (Olsen, 2012) with the exception of the gaze velocity threshold, which was varied every between 10 °/s and 400 °/s in increments of 10 °/s as done for Experiment 1.

The eye fixations identified at each threshold value were used to create the scan path sequences of participants. AOIs were included in the participant's scan path sequence if at least one eye fixation

took place within the AOI. An eye fixation was determined to take place within an AOI through the automated mapping feature of Tobii Pro Lab. In cases of an incorrect mapping (i.e., where the gaze automated mapping indicates that the gaze took place on AOI A when it actually took place in AOI B) or missing mapping (i.e., where the gaze is not mapped to any AOI due to a failure in the automated mapping process), the researchers manually manual mapped the gaze to the corresponding AOI observed by the participant.

Data analysis

The analysis carried out for Experiment 2 was identical to the one carried out for Experiment 1.

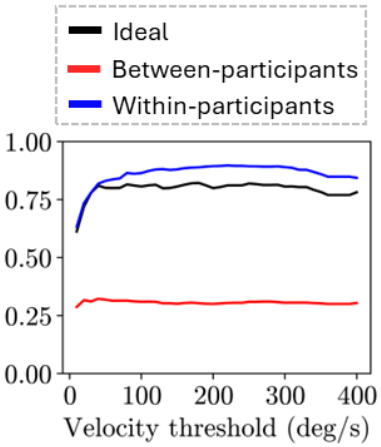
Results

The average string-edit similarity values for the between-participants (r_s = 0.36, p-value < 0.05) had a low positive correlation with the ideal similarity trend, while the average similarity values of the within-participants (r_s = 0.64, p-value < 0.01) method had a positive moderate correlation with the ideal similarity trend (Figure 20(b)). The two methods did not have a statistically significant correlation (r_s = -0.18, p-value > 0.05).

Figure 20(a) showcases the trends of average string-edit similarity for the between-participant, the within-participant method, and the ideal similarity trend values. The within-participant and the ideal trend followed similar shapes across threshold values, while the between-participant trend remained stable throughout all threshold values. Across the within-participants and ideal scan path sequence trends, one can observe a rapid increase in the similarity values at low thresholds values, followed by a wide period of high and stable similarity values followed by a slow decrease at larger thresholds. On the other hand, the between-participant method not only had much smaller similarity values (all values less than 0.35), it also had a much smaller increase in similarity values at low thresholds (0.287 at 10 °/s to 0.322 at 40 °/s) than the within-participant method (0.632 at

10 °/s to 0.818 at 40 °/s) or the ideal scan path comparisons (0.612 at 10 °/s to 0.809 at 40 °/s) across all threshold values. Furthermore, the ideal trend began to decrease at approximately 310 °/s (0.806) while the within-participant began to decrease at around 300 °/s (0.886). In the case of the between-participants trend, no clear decrease could be identified.

Figure 20. Plot of the string-edit similarity over the range of thresholds evaluated (10 °/s to 400 °/s) for each method and the ideal scan path sequence string-edit similarities, as well as the Spearman correlations between the trends.



(a)

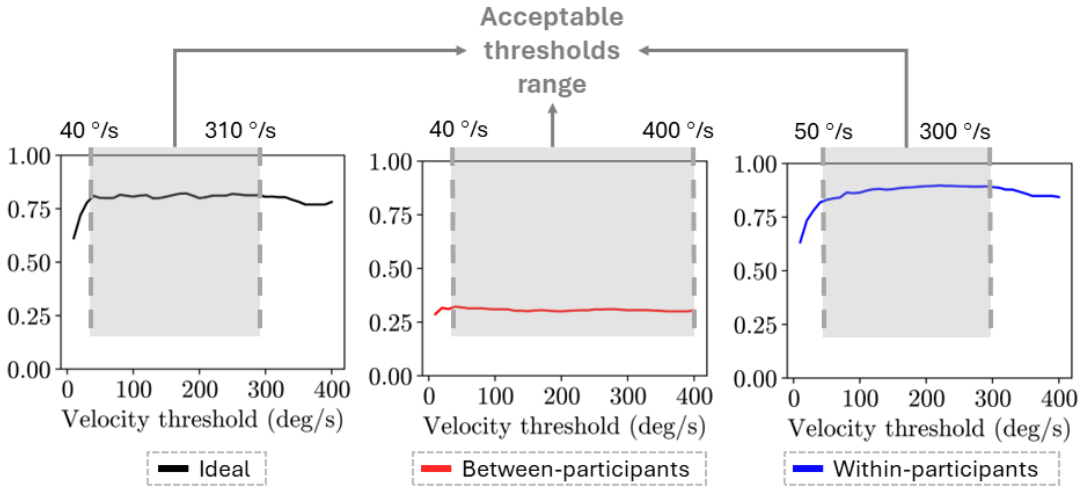
Method	1	2	3
1. Ideal	-		
2. Between-participants	0.36 *	-	
3. Within- participants	0.64 **	-0.18	-

* denotes *p-value* < .05, ** denotes *p-value* < .01

(b)

Figure 21 shows the threshold ranges identified from the between-participant, within-participant, and ideal trends. The threshold range selected for the within-participant threshold range was from 50 °/s to 300 °/s, for the between-participant method was from 40 °/s to 400 °/s, and the threshold range for the ideal similarity values was between 40 °/s to 310 °/s. The threshold range for the between-participant method was 90 °/s larger than the threshold range identified from the ideal trend.

Figure 21. Plot of the string-edit similarity over the range of thresholds values evaluated (10 °/s to 400 °/s) for each method. The accurate threshold ranges identified for each method are highlighted in gray.



Lastly, the threshold ranges identified with the between-participant and within-participant methods had a similarity score of 80% or higher on the ideal trend (Table 4). In more detail, the thresholds range identified by the within-participant had 0.800 (50 °/s) and 0.806 (300 °/s) average similarity scores in the ideal scan path sequence trend, while for the between-participants method, the average similarity values at the bounds of the threshold range identified were 0.809 (40 °/s) and 0.781 (400 °/s) in the ideal scan path sequence trend.

Table 4. String-edit similarity values when comparing to the ideal scan path sequences at the lower and upper bounds of the threshold range identified through the within-participants and between-participants method.

Method	Lower bound threshold (similarity)	Upper bound threshold (similarity)
Between-participants	40 °/s (0.809)	400 °/s (0.781)
Within-participants	50 °/s (0.800)	230 °/s (0.811)

Discussion

We were able to identify and select a range of threshold values for the I-VT algorithm capable of creating accurate scan path sequences for a simple bullseye-target tracking task and a complex local air traffic control task. More specifically, the within-participants and between-participants methods calculated string-edit similarity values at different gaze velocity thresholds of the I-VT algorithm that closely approximated the string-edit similarity values that result from comparisons to ideal scan path sequences. The contribution of these two methods, the novel within-participants method and the expanded between-participants method, might help researchers and practitioners using eye tracking to identify and select threshold values for the I-VT algorithm that can create accurate scan path sequences in their respective applications.

Impact of thresholds on scan path sequence similarity

Our results show that the impact of thresholds values on scan path sequence similarity to an ideal scan path sequence appear to follow a similar trend in both simple and complex environments and task, which match those found in prior research studies. In more detail, prior research using a dot-tracking task on a computer screen (Harężlak & Kasprowski, 2014) and a chess-board memory recall task on a computer screen (Blignaut & Beelders, 2009) have shown that the relationship between threshold values of the Dispertion-Threshold Identification (I-DT) eye movement detection algorithm and similarity values to an ideal scan path sequence tend to increase sharply at low threshold values, after which they remain stable for a period of time prior to decreasing. Therefore, our results expand upon these prior research efforts by identifying a similar trend using the I-VT algorithm on both a simple (bullseye-target tracking task) and complex task (local air traffic control task in a high-fidelity simulator).

The fact that the impact of threshold values on scan path sequence similarity appears to follow a general trend across multiple tasks and environments is important, as these different tasks and environments elicit different eye movements from participants. For example, the bullseye-target tasked used in Experiment 1 contained AOIs that were observed only once by the participant in rapid succession as each AOI only appear once in the environment for 1 second (e.g., the first bullseye-target observed). On the other hand, Experiment 2 contained AOIs that did not change positions in the environment and that participants visited multiple times (e.g., the BRITE radar AOI) or even had multiple transitions between them in order to complete their task. In other words, the two tasks and environments required participants to visually observe the environment differently. Therefore, it might be possible that a similar trend of threshold impact of scan path sequence similarity could be observed in other tasks and environments.

Approximating the impact of thresholds on scan path sequence similarity

The within-participants and between-participants methods approximated the general trend between thresholds values and scan path sequence similarity values when comparing to an ideal scan path sequence. More specifically, the two methods achieved statistically significant very-high-to-high correlations (r_s = 0.77 for within-participants and r_s = 0.99 for between-participants) for the simple bullseye-target tracking task in the Gazebase dataset used in experiment 1, while achieving statistically significant moderate-to-low correlations (r_s = 0.64 for within-participants and r_s = 0.36 for between-participants) in the complex local controller air traffic control task used in experiment 2. To the author's knowledge, these results are present the first effort in the literature to approximate the relationship between threshold values an eye movement detection algorithm and the scan path sequence similarity values.

Although both methods achieved high correlations in the simple task described in Experiment 1, the within-participant method had a much higher correlation and closer similarity values to the ideal scan path similarity trend than the between-participant method in the more complex task used for Experiment 2. The reason behind this result is that the between-participant method compares the scan path sequences of different participants, and thus, it is affected by potential differences that might exist between the scan path sequences of participants. On the other hand, the within-participant method only compares the scan path sequence of a single participant at multiple different thresholds, meaning that is not impacted by the differences in the scan path sequences of participants. Thus, when the scan path sequences carried out by participants might be different from one another, as might be the case in an air traffic control task or other complex tasks and environments, the trend created by the between-participants method might not result in a good approximation of the ideal scan path sequence trend.

Therefore, if a researcher expects the scan path sequences between participants to be different from each other, it might be a better option to apply the within-participants method rather than the between-participants method. On the other hand, if the scan path sequences between participants are likely to be very similar to each other, such as in the bulls-eye target task, then the between-participants method may approximate the ideal scan path sequence trend better than the within-participants method. As shown in the present study, using the appropriate method, we were able to identify and select thresholds capable of creating accurate visual scan paths without the need to manually create nor compare to an ideal visual scan path.

Chapter 5 – Classification framework to identify similar visual scan paths using multiple similarity metrics

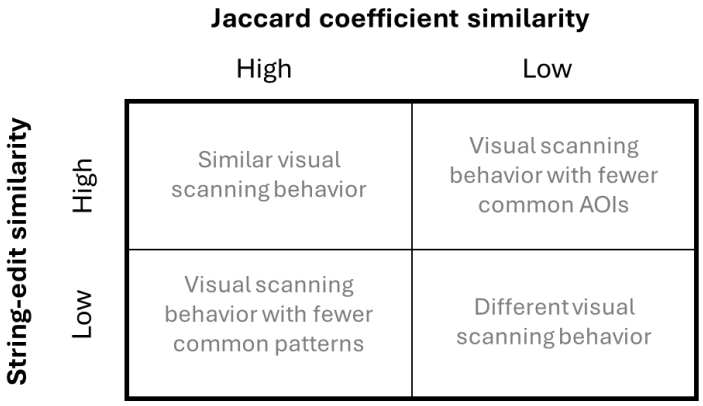
In this chapter, we introduce a classification framework using multiple similarity metrics to identify similar visual scan paths that might contain minor variations, such as slight differences in the order information were observed or what information was observed, but share a common underlying visual scanning strategy.

The chapter is structured in the following manner. First, the proposed classification framework is explained alongside worked examples that highlight how similar visual scan paths with minor variations between them can be identified. Second, the proposed classification framework is applied to the visual scan paths collected from local air traffic controllers issuing clear to take off clearance in a high fidelity simulator. Third, results from the application of the classification framework are presented and discussed.

Proposed approach

The framework applies multiple similarity metrics, the string-edit similarity and the Jaccard coefficient similarity (explained in Chapter 2), in order to quantify how similar two visual scan paths are across multiple dimensions (i.e., what information was observed, in the case of the Jaccard coefficient similarity, or the order information was observed, in the case of the string-edit similarity) and classify them into one of the categories shown in the matrix depicted in Figure 22: (1) similar visual scanning behavior (high string-edit similarity and high Jaccard coefficient similarity); (2) visual scanning behavior with fewer common AOIs (high string edit similarity and low Jaccard coefficient similarity); (3) visual scanning behavior with fewer common patterns (low string-edit similarity and high Jaccard coefficient similarity); (4) different visual scanning behavior (low string edit similarity and low Jaccard coefficient similarity).

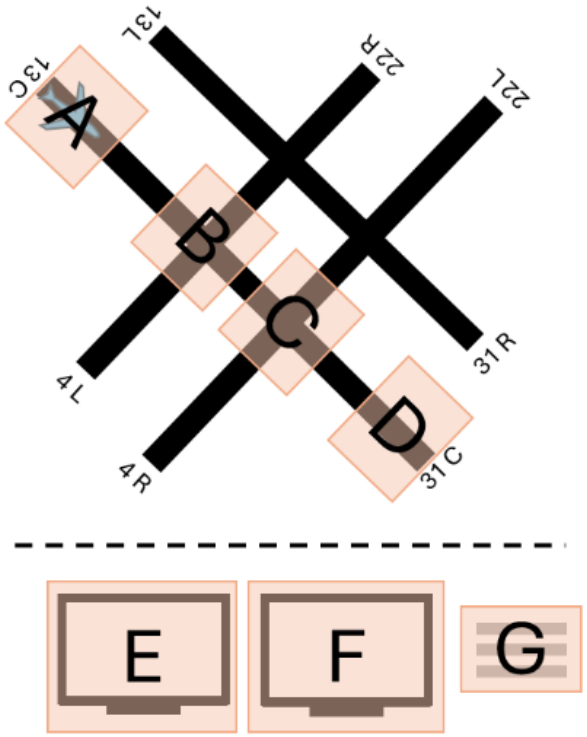
Figure 22. Classification of scan path sequences based on Jaccard coefficient similarity and string-edit similarity values.



Visual scan path sequences that contain high values on both similarity metrics are considered to represent similar visual scanning strategies, while those that possess low similarity across both similarity metrics are considered to be different visual scanning strategies. When the two similarity metrics differ, two visual scan paths might represent variations of a similar underlying visual scanning behavior. In the case of high Jaccard coefficient similarity and low string-edit similarity, the visual scan path might contain multiple AOIs in common but had uncommon patterns when inspecting the AOIs. When the Jaccard coefficient similarity is low and the string-edit similarity is high, the visual scan path might contain fewer AOIs in common but used common patterns when inspecting the few AOIs in common.

To visualize the application of the proposed classification framework, and its ability to identify similar visual scan paths with slight variations between them, consider the following example visual scan paths contained in Figures 24, 25, 26, and 27 that might be applied by controllers to the airport layout in Figure 23 below.

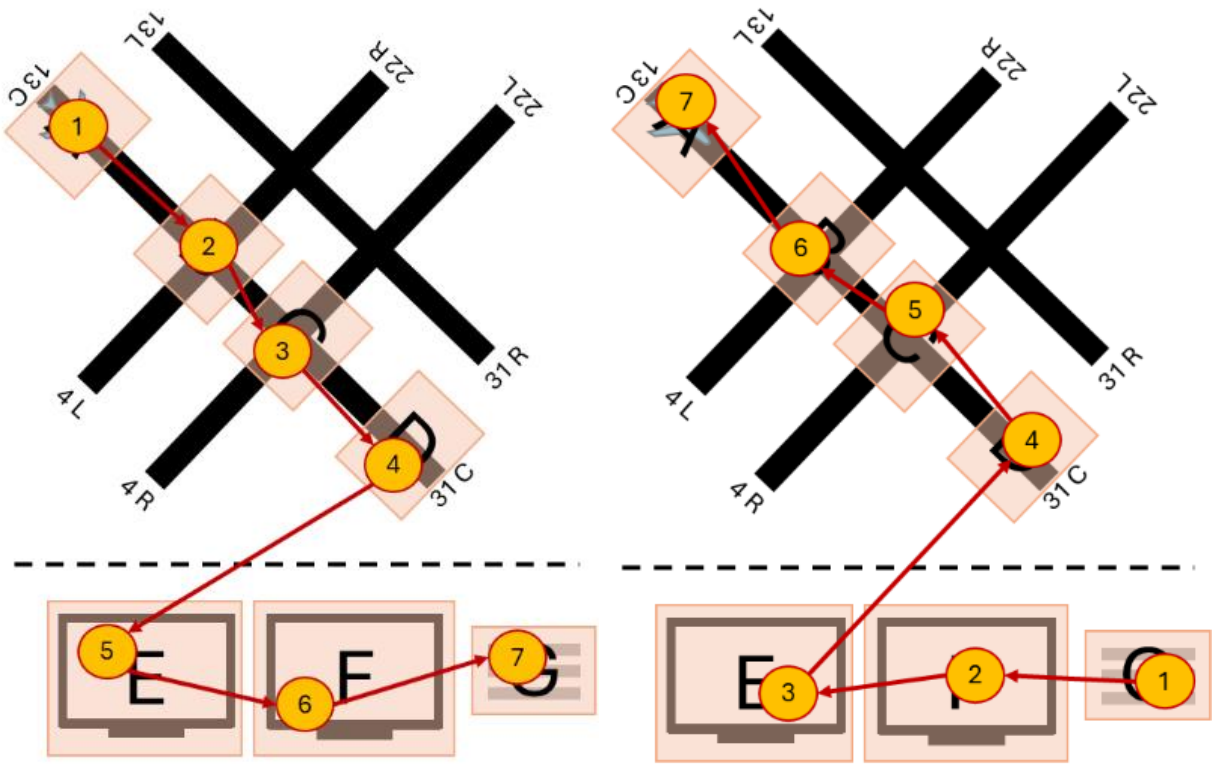
Figure 23. The layout of AOIs in the airport (AOIs A-D) and tower cabin (AOIs E-G), which resembles the Midway International airport.



Note. AOIs were drawn on hotspots in the active runway (Crutchfield et al., 2021), including the beginning and end of the runway (AOIs A and D), as well as intersecting runways (AOIs B and D). In addition, AOIs E and F were drawn to represent radar screens present in the tower environment, while AOI G represents the flight strips, which contain information (e.g., destination airport) regarding the aircraft.

Figure 24 showcases how two controllers inspected all of the same AOIs but in reverse order. Thus, the Jaccard coefficient similarity between these two visual scan paths is high (1), as they inspected all the same AOIs in the airport, but the string-edit similarity is low (0.143), due to the different order in which the AOIs were inspected.

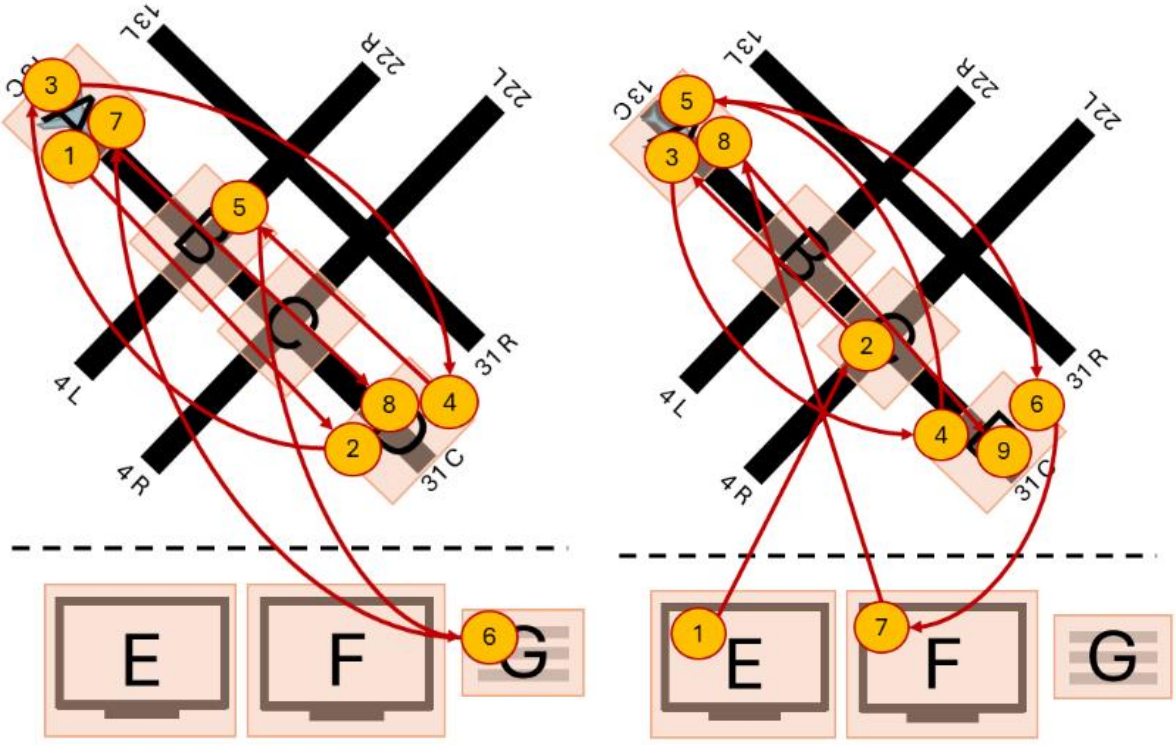
Figure 24. Simplified example of impact two visual scan paths (b) and (c) classified as having high string edit similarity but low Jaccard coefficient similarity.



Note. Figure 24(b) highlights the visual scan path of one controller, creating the scan path sequence ABCDEFG; Figure 24(c) showcases a similar visual scan path of another controller but in the opposite order, creating the scan path sequence GFEDCBA. The yellow circles denote AOIs inspected, and the inscribed numbers the order the AOIs were inspected in. The red lines represent the movement of participants between AOIs.

Figure 25 contains two visual scan paths where controllers inspected different AOIs (B and G in Figure 25(a) and C, E, and F in Figure 25(b)) but had similar patterns among the AOIs they had in common (multiple movements between AOI A and D), resulting in a high string-edit similarity (0.555) but low Jaccard coefficient similarity (0.285).

Figure 25. Simplified example of two visual scan paths (a) and (b) classified as having high string-edit similarity and low Jaccard coefficient similarity.

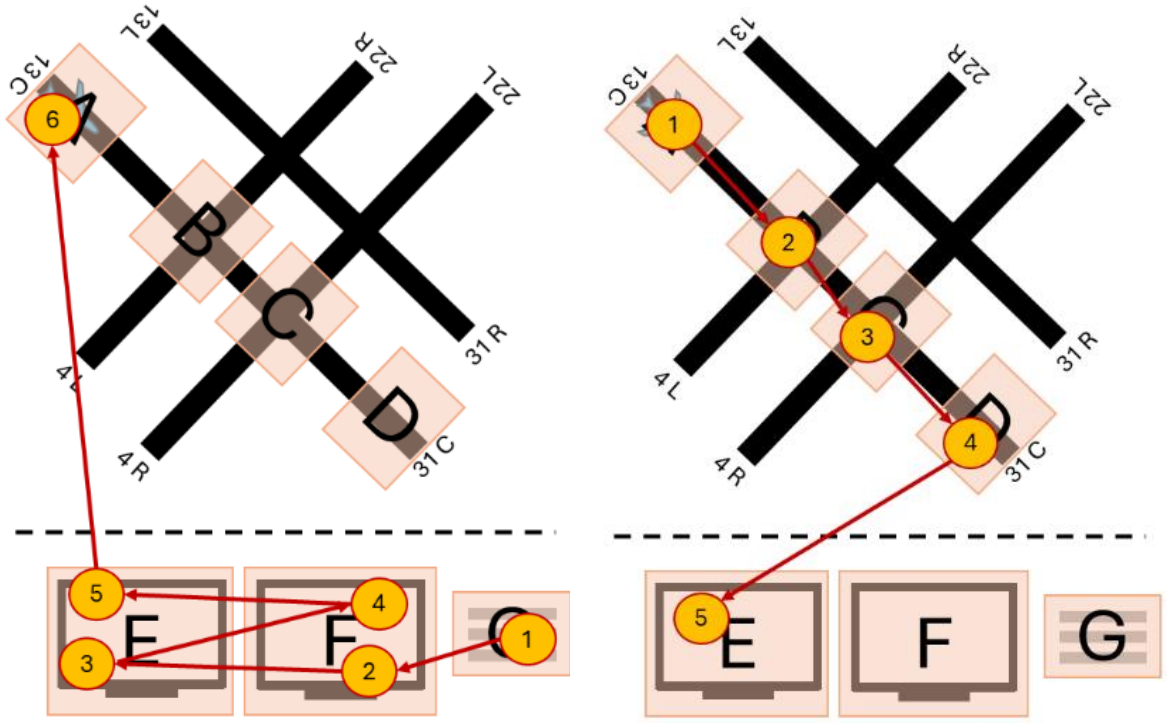


(a) (b)

Note. The yellow circles denote AOIs inspected, and the inscribed numbers the order the AOIs were inspected in. The red lines represent the movement of participants between AOIs.

Figure 26 includes two controllers applying different visual scanning strategies, where one controller (Figure 26(a)) focused primarily on the radars (AOIs E and F) and flight strip (AOI G), while the other controller scanned the runway hotspots (AOIs A-D) and only one of the radar screens (AOI E). As a result, these two visual scanning strategies have low string-edit similarity (0.166) and low Jaccard coefficient similarity (0.285).

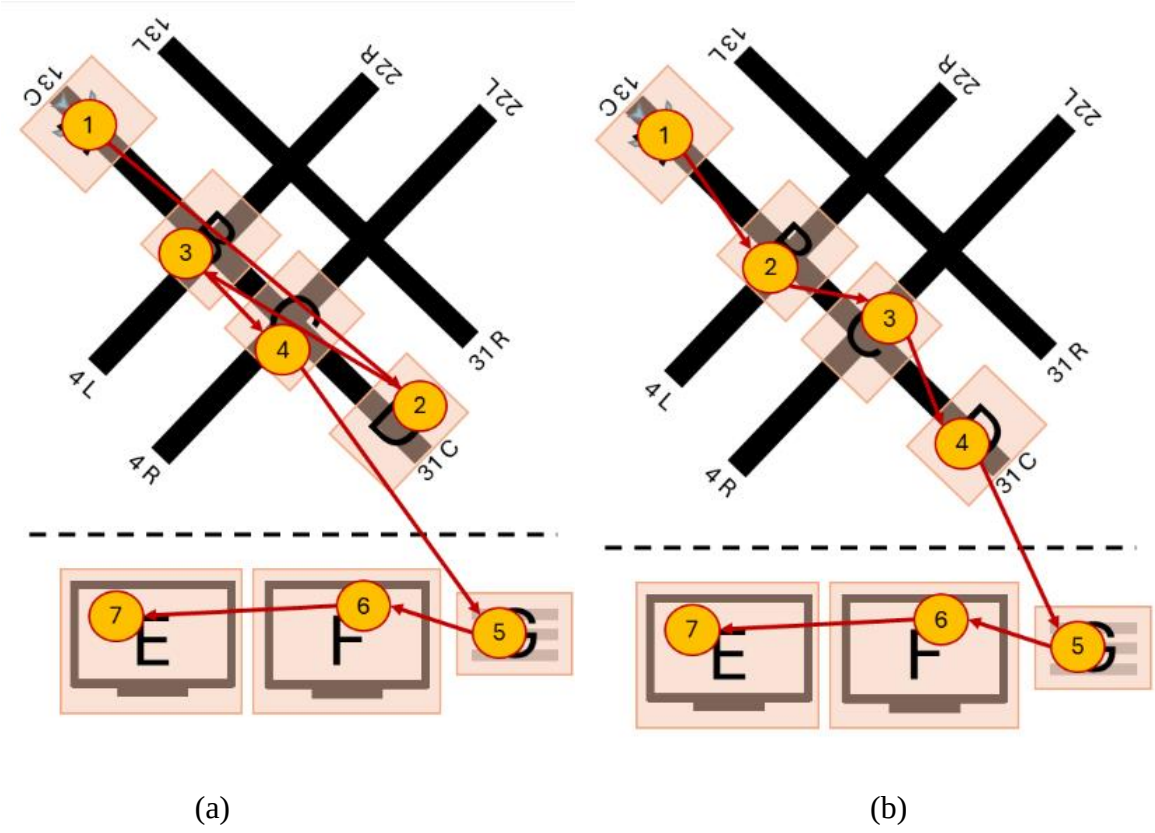
Figure 26. Simplified example of two visual scan paths (a) and (b) classified as having low string-edit similarity and low Jaccard coefficient similarity.



Note. The yellow circles denote AOIs inspected, and the inscribed numbers the order the AOIs were inspected in. The red lines represent the movement of participants between AOIs.

Figure 27 includes two controllers applying similar visual scanning strategies. Both controllers follow a nearly identical order when scanning the active runway hotspots (AOIs A-D), followed by scanning the radars (AOIs E and F) and the flight strips (G). As a result, these two visual scanning strategies have high string-edit similarity (0.714) and high Jaccard coefficient similarity (1).

Figure 27. Simplified example of two visual scan paths (a) and (b) classified as having high string-edit similarity and high Jaccard coefficient similarity.



Note. The yellow circles denote AOIs inspected, and the inscribed numbers the order the AOIs were inspected in. The red lines represent the movement of participants between AOIs.

To identify similar visual scan paths, our proposed framework uses the average Jaccard coefficient similarity and average string-edit similarity values between all visual scan paths as thresh-olds to classify visual scan paths as having “high” and “low” similarity. More specifically, if a visual scan path contains an average string-edit similarity higher than the average string-edit similarity between all participants, then the visual scan path is considered to have a “high” string-edit similarity. On the other hand, a visual scan path with an average Jaccard coefficient similarity lower than the average Jaccard coefficient similarity between all participants is classified as having a “low” Jaccard coefficient similarity. Thus, we can define “high” and “low” similarities based on

the visual scanning behavior contained in each person's visual scan path, without requiring the researcher to define "high" and "low" similarity thresholds based on their subjective judgement.

As an example, consider the following application of the proposed framework to the visual scan paths 4A and 4B represented in Figure 26(a-b) (i.e., 4A is the visual scan path shown in Figure 26(a)) as well as the visual scan paths 5A and 5B showcased in Figure 27(a-b) (i.e., 5A is the visual scan path shown in Figure 27(a)). The string-edit similarity and Jaccard similarity coefficient values between all visual scan path comparisons can be found in Figure 28(a-b) below, as well as the average values for each visual scan path. The average string-edit similarity between all visual scan paths was 0.408, while the average Jaccard similarity coefficient was 0.642. Thus, using the classification matrix shown in Figure 1 above, we can classify the visual scan paths as shown in Figure 28(c).

Figure 28. Example application of the proposed framework to the visual scan paths 25A and 25B shown in Figure 25(a-b) and the visual scan paths 26A and 26B showcased in Figure 26(a-b).

String-edit similarity matrix

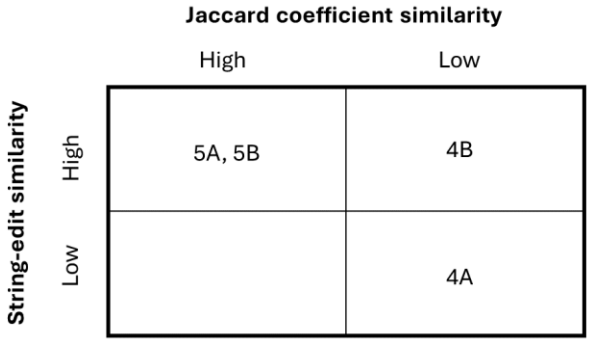
Participant	4A	4B	5A	5B
4A	-			
4B	0.167	-		
5A	0.143	0.571	-	
5B	0.143	0.714	0.714	-
Average	**0.151**	**0.484**	**0.476**	**0.523**

(a)

Jaccard coefficient similarity matrix

Participant	4A	4B	5A	5B
4A	-			
4B	0.286	-		
5A	0.571	0.714	-	
5B	0.571	0.714	1	-
Average	**0.476**	**0.571**	**0.761**	**0.761**

(b)



(c)

Note. (a) contains the similarity matrix for the string-edit distance between all the visual scan paths as well as the average for each visual scan path; (b) contains the similarity matrix for the Jaccard coefficient similarity between all the visual scan paths as well as the average for each visual scan path; (c) contains the classification matrix with each visual scan path classified into one of the four categories based upon its average string-edit similarity and Jaccard coefficient similarity and the average values of the two metrics between all participants.

Experiment

Note that the following contents, such as the participants, apparatus, among others overlap the experiment described in Chapter 4, experiment 2, as they both belong to the same experiment. Nonetheless, key differences exist between the two, such as the number of participants and additional information regarding the specific clearance issued, are highlighted in this section.

Participant and apparatus

A total of 14 retired local controllers with an average of 26 years of experience (range between 10 and 42 years) participated in the experiment. However, data from 5 participants was not included in the present work due to low eye tracking data quality, where eye movements carried out by the participants were not collected by the eye tracker. As a result, only data from 9 participants was analyzed in the present work.

The controllers managed simulated landing and departing air traffic on a high-fidelity Adacel tower simulator used at the FAA's Civil Aeronautical Medical Institute (CAMI) in Oklahoma City,

Oklahoma. Twelve 55” HD (1080p) monitors, wrapped greater than 180° *around the participant,* were used to simulate the out the window view of tower cabin. The simulators included flight strips of the aircraft in the scenarios, as well as working Bright Radar Indicator Terminal Equipment (BRITE) and Airport Surface Detection (ASDE) radar displays. The participants issued verbal clearances to the aircraft using a standard communication headset.

The Tobii Pro Glasses II (100 Hz) were used to capture participant eye movements (equipped with prescription lenses as necessary). In addition, the Tobii Pro Lab software was used to visually observe the recorded video from the participant’s point of view that contained the raw gaze data overlaid in order to identify ideal scan path sequences. In addition, the Tobii Pro Lab software was used to apply the I-VT algorithm in order to identify eye fixations and saccadic movements.

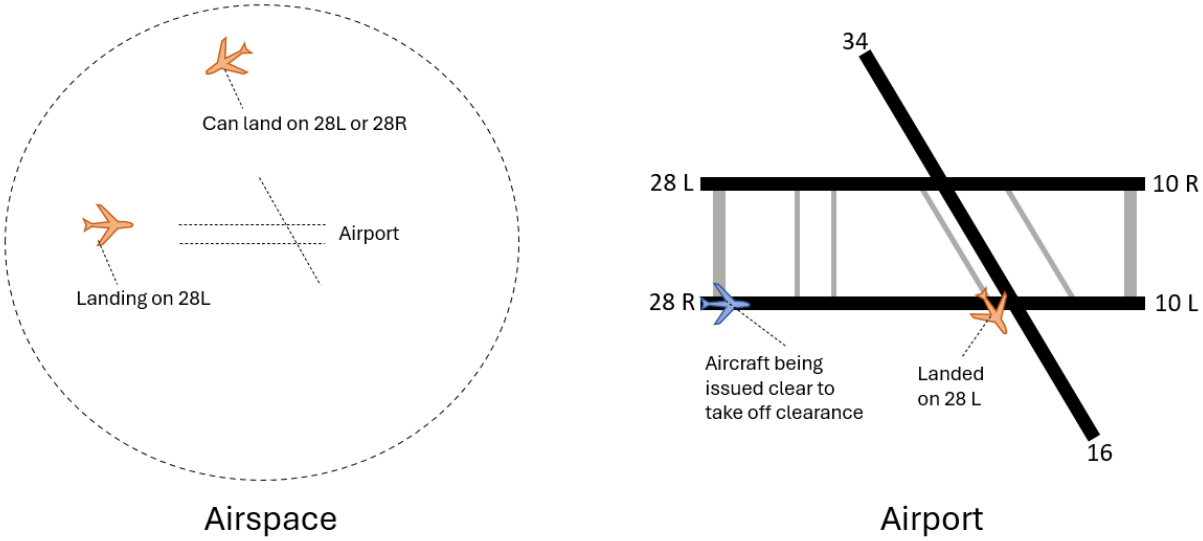
Task and scenario

The participants were tasked with managing landing and departing air traffic during a scenario that lasted approximately 22 minutes. The scenario had high visibility and sunny weather conditions and took place during daylight hours. Furthermore, the scenario contained a total of 33 aircraft, with 19 arriving aircraft and 14 departing aircraft. The scenario ended once the 22-minute mark was reached, which might occur prior to all aircraft present in the environment being issued a clearance.

The present study focuses on the first clear to take off clearance issued by all the participants to account for potential factors that might impact the similarity between controllers (e.g., location of aircraft on the airport). The location of aircraft in the environment can be observed in Figure 29. In the airspace, there are two aircraft present (orange aircraft), one that has already been issued a clear to land clearance by the controller, and a second aircraft that is approaching the airport. On the ground, there are also two aircraft present, one that recently landed and is currently (or close

to) exiting runway 28R (orange aircraft), while another aircraft (blue aircraft) is waiting to be issued a clear to take off clearance by the controller. Note that other aircraft are present in the airport environment, such as those taxiing towards a runway for departure that are under the control of the ground controller.

Figure 29. Visualization of the environmental conditions of the airspace and airport while the clear to take off clearance analyzed was issued.



Note. The aircraft that will be issued the clear to take off clearance is highlighted in blue, while a total of three aircraft of interest to the participant are highlighted in orange. Two of the aircraft of interest in the airspace are arriving at the airport in the next few minutes, with one aircraft already issued a clear to land clearance to runway 28L. The other aircraft of interest has already landed and is exiting the runway momentarily.

Processing eye movement data

The eye movements of participants were analyzed while issuing a single a clear to take off clearance to one aircraft, where controllers are more likely to show focused attention (Manske & Schier, 2015). In this study, the time to issue the clearance began from the moment the aircraft reports to the tower controller that it is ready for departure up until the point when the controller finishes issuing the clearance. Among all 9 participants, the average time to issue the clear to take

off was approximately 25.1 seconds, with the lowest time being 15.7 seconds while the longest was 32.2 seconds.

The I-VT algorithm implemented in Tobii Pro Lab was used to identify eye fixations and saccadic movements. The gaze velocity threshold of the I-VT algorithm was set at 90 °/s, while all other settings were left at their default values (e.g., minimum eye fixation duration of 60 ms) (Olsen, 2012). The 90 °/s gaze velocity threshold was selected through the application of the automated procedure explained in Chapter 2 to the four eye movement metrics collected from the 9 participants: (1) number of eye fixations; (2) percentage of eye fixations in AOIs; (3) percentage of gazes in eye fixations; (4) within-participant similarity (metric described in Chapter 3). The eye fixations and saccades were exported from the Tobii Pro Lab software and used to create the scan path sequences of participants. Lastly, note that the 90 °/s gaze velocity threshold value is close to one of the default gaze velocity threshold values (i.e., 100 °/s) included in the Tobii Pro Lab software.

AOIs in the tower cabin environment were identified and created in collaboration with subject matter experts based on their operational significance, as done in Chapter 4.

Data analysis

The string-edit similarity and Jaccard coefficient similarity between the participants' visual scan paths controller were calculated, as well as the average values between all participants. The average Jaccard coefficient and string-edit similarity values between all scan path sequence comparisons were used to define "high" and "low" similarity values for both similarity metrics. Afterwards the average similarity values of each participant were compared to the average similarity values between all participants were used to classify the visual scan paths following the classification matrix shown in Figure 22. Lastly, the visual scan paths of participants were

visualized by groups based on their assigned classification and compared to identify common and uncommon visual scanning behaviors.

Results

Figure 30 contains the Jaccard coefficient similarity and the string-edit similarity values calculated between the visual scan paths of controllers. On average, the Jaccard coefficient similarity values between participants (0.742) were higher than the average string-edit similarity values (0.322). The highest Jaccard coefficient similarity value achieved between two visual scan paths was 1 while the lowest value was 0.444, while for the string-edit similarity, the highest value calculated was 0.462 and the lowest value was 0.143.

Figure 30. Individual and average (a) string-edit similarity values and the (b) Jaccard coefficient similarity values calculated.

String-edit similarity

Participant	S1	S2	S3	S4	S5	S6	S7	S8	S9
S1	-								
S2	0.400	-							
S3	0.143	0.267	-						
S4	0.154	0.333	0.429	-					
S5	0.364	0.409	0.455	0.364	-				
S6	0.375	0.375	0.375	0.250	0.364	-			
S7	0.318	0.364	0.318	0.273	0.273	0.227	-		
S8	0.462	0.333	0.357	0.273	0.318	0.438	0.273	-	
S9	0.312	0.250	0.375	0.250	0.318	0.250	0.318	0.318	-
Average	**0.316**	**0.341**	**0.339**	**0.290**	**0.358**	**0.331**	**0.295**	**0.338**	**0.290**

(a)

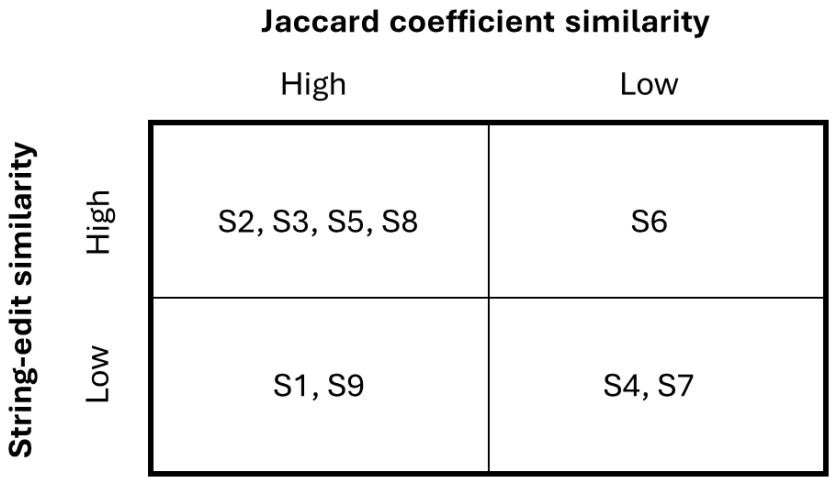
Jaccard coefficient similarity

Participant	S1	S2	S3	S4	S5	S6	S7	S8	S9
S1	-								
S2	1	-							
S3	0.857	0.857	-						
S4	0.714	0.714	0.833	-					
S5	0.857	0.857	1	0.833	-				
S6	0.625	0.625	0.5	0.571	0.5	-			
S7	0.75	0.75	0.625	0.5	0.625	0.444	-		
S8	0.857	0.857	0.714	0.833	0.714	0.714	0.625	-	
S9	0.857	0.857	1	0.833	1	0.5	0.625	0.714	-
Average	**0.814**	**0.814**	**0.798**	**0.728**	**0.798**	**0.559**	**0.618**	**0.753**	**0.798**

(b)

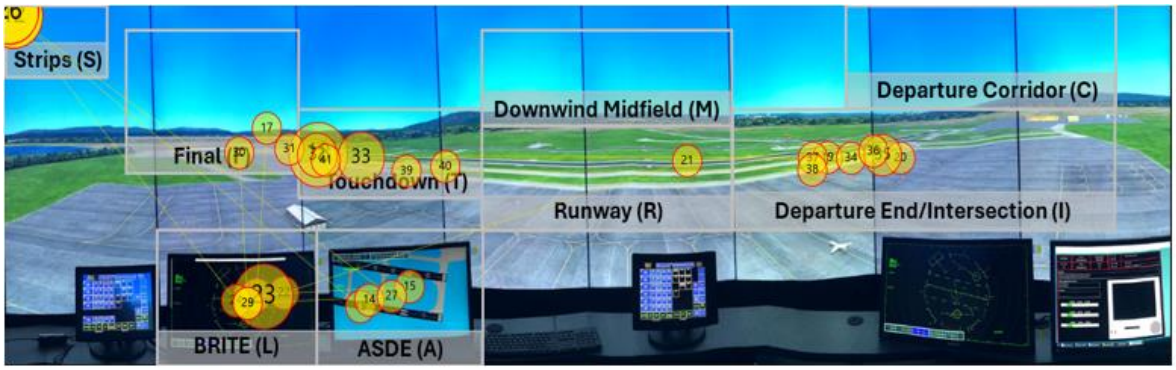
Figure 31 showcases the classification of visual scan paths based on the average string-edit and Jaccard coefficient similarity values between participants. Most visual scan paths of controllers were classified as highly similar in at least one similarity metric. More specifically, four visual scan paths were classified as highly similar across both similarity metrics, one visual scan path was classified as having high string-edit similarity but low Jaccard coefficient similarity, and the remaining two visual scan paths were classified as having low string-edit similarity but high Jaccard coefficient similarity. On the other hand, two visual scan paths were classified as having low similarity across both similarity metrics.

Figure 31. Classification matrix of the 9 visual scan paths of expert tower controllers while issuing a clear to take off clearance based on their average string-edit and Jaccard coefficient similarity values.



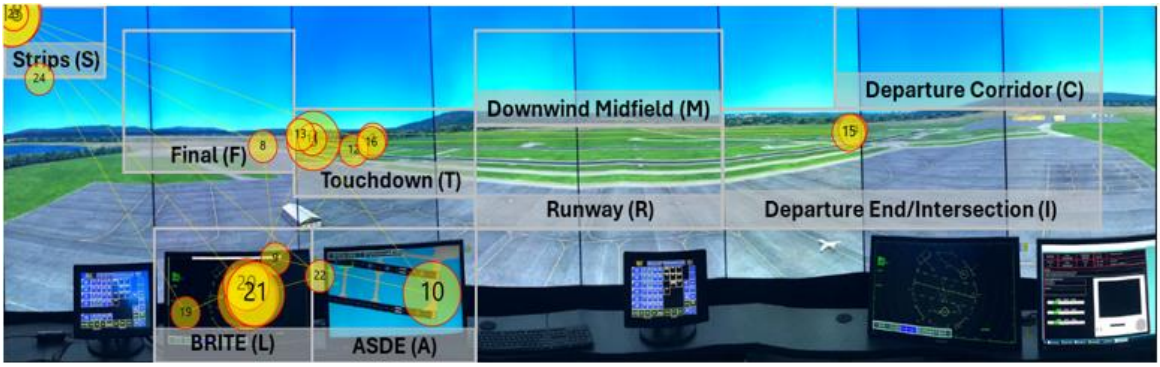
The visual scan paths classified as highly similar across both similarity metrics can be observed in Figure 32. All visual scan paths inspected the two runway hotspots AOIs (T and I), as well as the flight strips AOI (S) and both the BRITE radar (L) and the ASDE radar (A). Two of the four visual scan paths contained the runway AOI (R), while three of the four visual scan paths contained the Final (F) AOI. Furthermore, the scan path sequences showcase common patterns between the controllers. More specifically, two visual scan paths (S5 and S3) had the same pattern TITS to inspect the runways and the flight strips, while the remaining two controllers had a very similar variations, such as TIT (S2) and ITS (S8). Similarly, all visual scan paths contain common patterns between the radar AOIs, such as LA or AL, as well as with the flight strips AOI, such as AS or SA.

Figure 32. Visualization of four visual scan paths (a-d) classified as highly similar across both string-edit and Jaccard coefficient similarities overlaid the AOIs in the airport environment, alongside their scan path sequences.



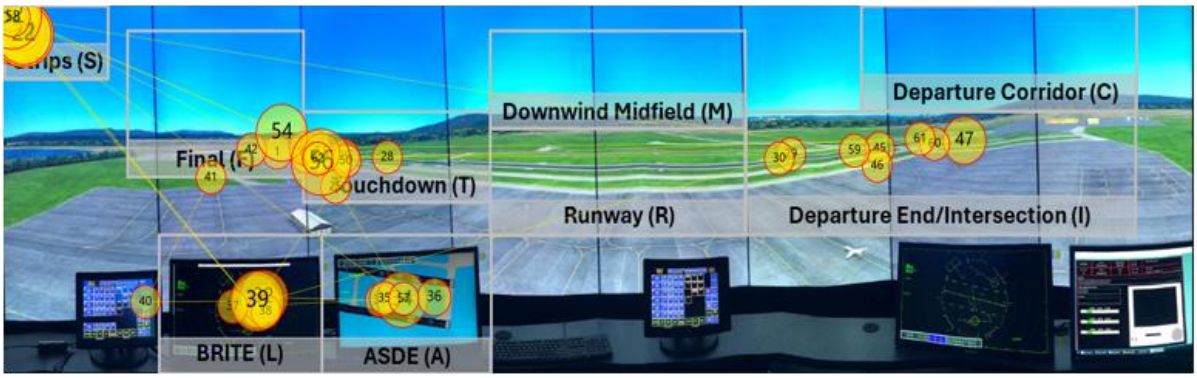
S2 - Scan path sequence: SALFTIRLSALFTIT

(a)



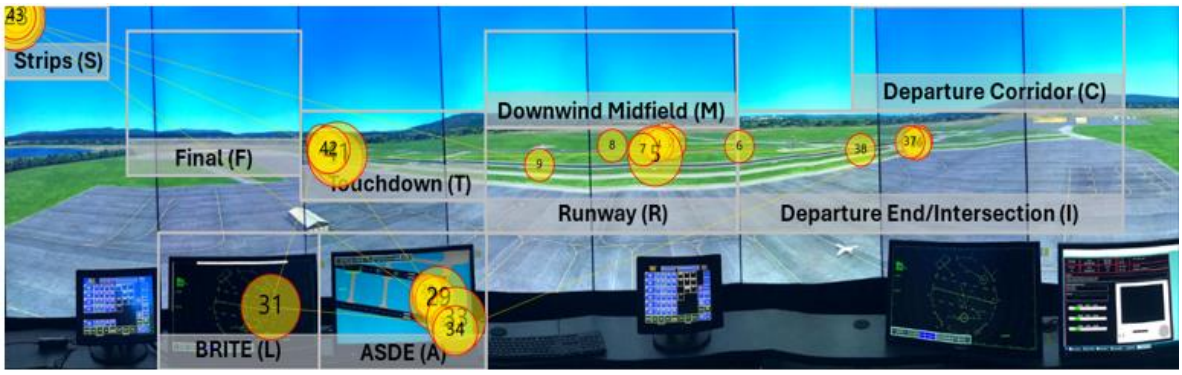
S3 - Scan path sequence: LTLSFLATITSLAS

(b)



S5 - Scan path sequence: FSLSLTILALFTITSTFTASIT

(c)



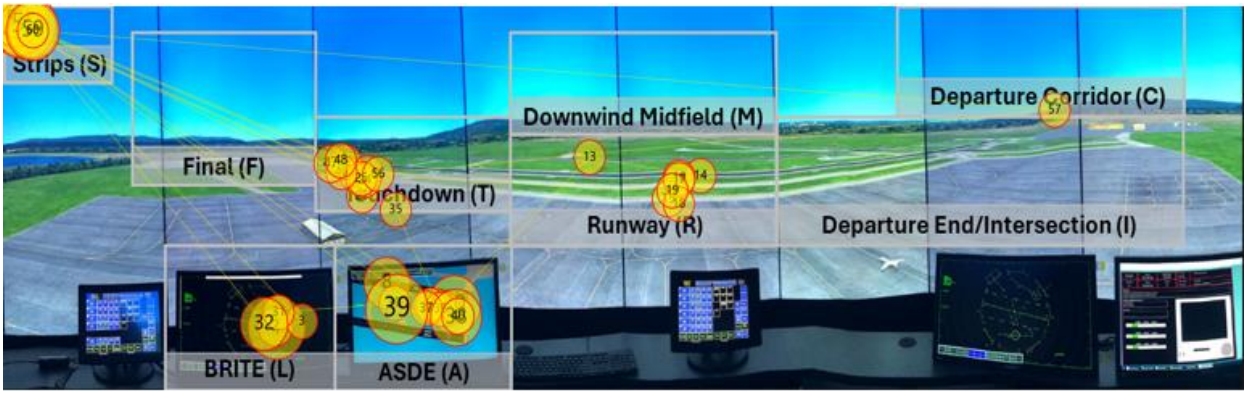
S8 - Scan path sequence: RSATLAITS

(d)

Note. The yellow circles denote eye fixations, and the inscribed numbers the order the said eye fixations took place. The yellow lines represent the saccadic movement of participants between eye fixations.

Figure 33 visualizes the sole visual scan path classified as having high string-edit similarity but low Jaccard coefficient similarity. The visual scan path did not contain two commonly inspected AOIs, the Final AOI (F) as well as the Intersection AOI (I). Furthermore, it is the only visual scan path that fixated on the Departure Corridor AOI (C). On the other hand, the visual scan path contained multiple commonly used patterns between the BRITE radar (L), the ASDE radar (A), and the flight strips (S) AOIs, such as SAL, SA, and AS, that can be observed in other visual scan paths (e.g., S2, S8, and S5). In addition, it also contained the pattern TS between the touchdown AOI (T) and the flight strips (S) AOI present in S3, S5, and S8.

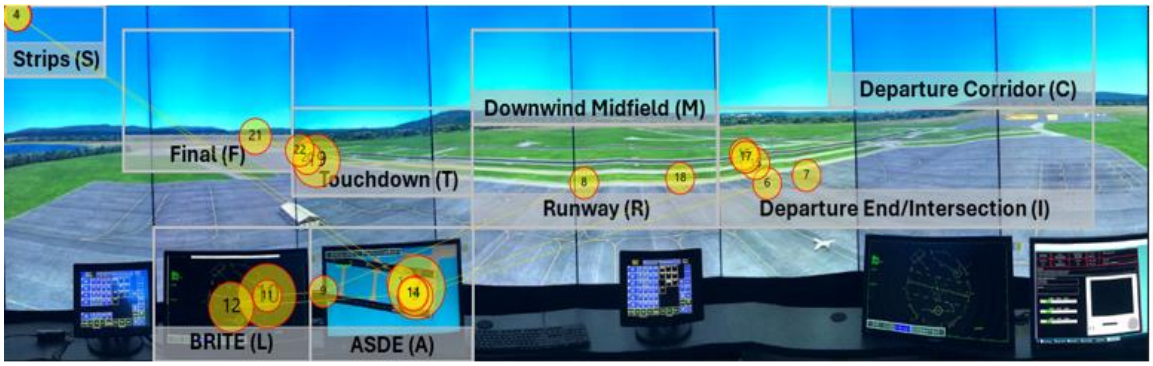
Figure 33. Visualization of the visual scan path classified as having high string-edit similarity but low Jaccard coefficient similarities overlaid the AOIs in the airport environment.



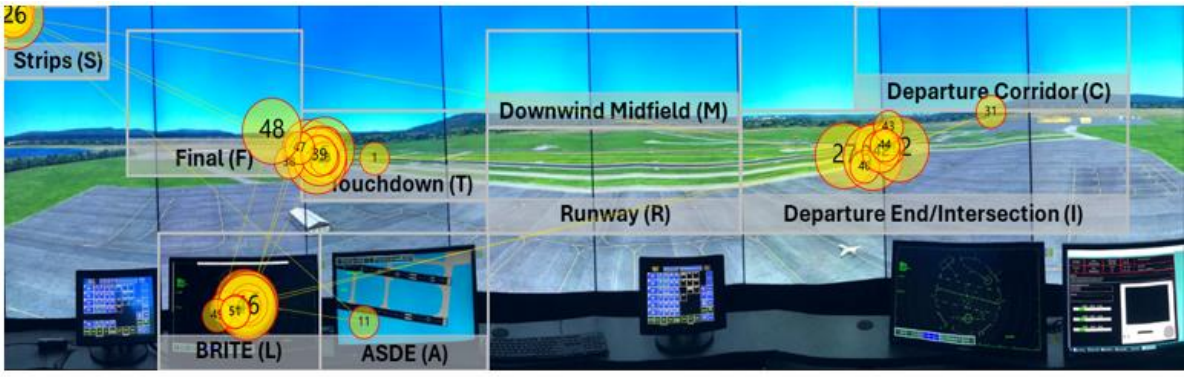
Note. The yellow circles denote eye fixations, and the inscribed numbers the order the said eye fixations took place. The yellow lines represent the saccadic movement of participants between eye fixations.

Figure 34 showcases the two visual scan paths classified as having low string-edit similarity but high Jaccard coefficient similarity. Although the visual scan paths contain the Intersection AOI (I) and the Touchdown AOI (T), as well as both radar AOIs (L and A) and the flight strips AOI (s), they were inspected in less common patterns. For example, S1 was the only visual scan path to use the pattern AIR, which was applied twice. S9 was the only participant to apply the pattern ITI when inspecting the runways.

Figure 34. Visualization of the two visual scan paths (a-b) classified as having low string-edit similarity but high Jaccard coefficient similarities overlaid the AOIs in the airport environment.



(a)



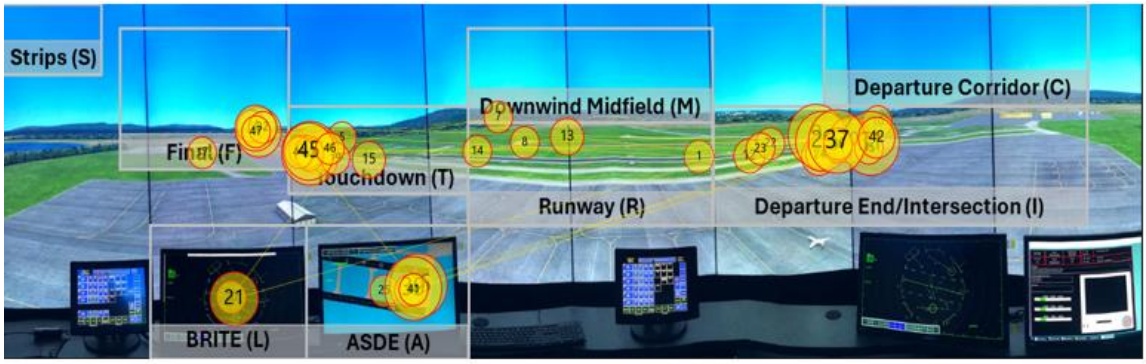
S9 - Scan path sequence: TLASTSITILFTILFL

(b)

Note. The yellow circles denote eye fixations, and the inscribed numbers the order the said eye fixations took place. The yellow lines represent the saccadic movement of participants between eye fixations.

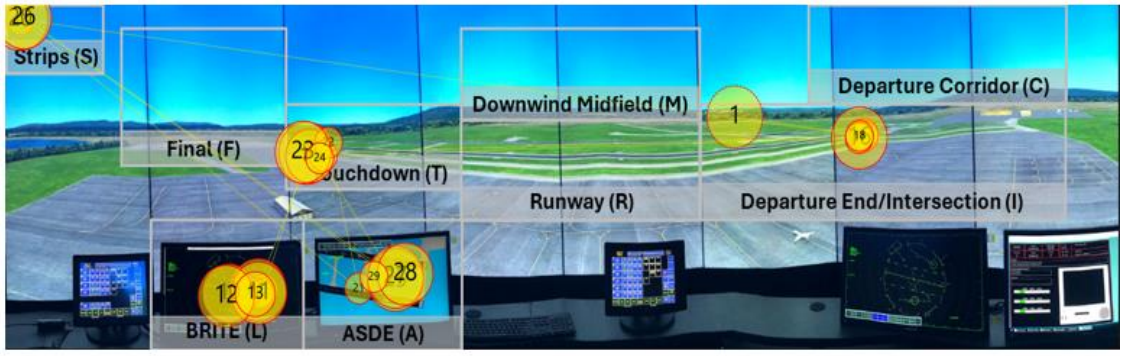
Figure 35 visualizes the two visual scan paths classified as having low similarity across both string-edit similarity and Jaccard coefficient similarity metrics. Both visual scan paths do not contain commonly inspected AOIs, such as the flight strips AOI (S) in the case of S7, as well as the Runway AOI (R) and the Final AOI (F) in the case of S4. In addition, S7 was the visual scan path to fixate on the Downwind Midfield AOI (M). Furthermore, neither visual scan path contains common AOI patterns or variations between the ASDE radar (A), the flight strips (S), and the BRITE radar (L), such as SAL or LSA, observed in other visual scan paths. On the other hand, they apply uncommon patterns, such as IAI, only applied by S4, as well as RIR, which can only be observed in S7.

Figure 35. Visualization of the two visual scan paths (a-b) classified as having low similarity across both similarity metrics overlaid the AOIs in the airport environment.



S7 - Scan path sequence: RFTMRIRTFTLIATIFTIAITF

(a)



S4 - Scan path sequence: ITLTISATASA

(b)

Note. The yellow circles denote eye fixations, and the inscribed numbers the order the said eye fixations took place. The yellow lines represent the saccadic movement of participants between eye fixations.

Discussion

We were able to identify similar visual scan paths carried out by expert tower controllers while issuing clear to take off clearances using our proposed framework. More specifically, our proposed approach classified the visual scan paths of tower controllers based upon their average Jaccard coefficient similarity, which accounts for the areas inspected by the controllers, as well as by their

average string-edit similarity, which considers the order in which the controllers inspected the areas. The contribution of the proposed framework might help researchers and practitioners interested in identifying similar visual scanning behaviors, particularly in cases where multiple variations of a similar behavior may exist.

Identifying similar visual scanning strategies when multiple variations are possible

The use of two complimentary similarity metrics allowed us to identify participants with similar visual scan paths that would otherwise be classified different by using only one metric. For example, had we only applied the string-edit similarity, we might have interpreted that S9 had a different visual scanning behavior even though they have a higher-than-average Jaccard coefficient similarity. In other words, throughout the visual scan path, the controller inspected many of the same AOIs as other the other participants but applied a different order. In addition, if we had only used the Jaccard coefficient similarity, we might have erroneously considered S6 to have a different visual scanning behavior even though they share many similar patterns between AOIs with other participants, as shown by their higher-than-average string-edit similarity. Therefore, by combining multiple similarity metrics that complement each other we were able to correctly identify similar visual scan paths while accounting for potential variations that might exist between them.

In addition, average Jaccard coefficient similarity and string-edit similarity values calculated between the expert tower controllers were higher than those reported in other tasks and environments. More specifically, we calculated an average Jaccard coefficient similarity of 0.742 and a string-edit similarity of 0.322. On the other hand, in a prior study in which participants observed images, such as paintings, reported a string-edit similarity of 0.28 between participants,

as well as an average similarity value of 0.54 based upon the common AOIs that were observed by participants (Priviteva & Stark, 200). Another study where participants were instructed to visually fixate on letters or numbers in a pre-determined order reported an average string-edit similarity of 0.23 and an average similarity value of 0.47 based upon the common AOIs that were observed by participants (Duchowski et al., 2010). One potential reason behind these differences was briefly discussed by Duchowski et al (2010), which we also share, is that the task participants are assigned with completing might impact how they carry out their eye movements, and as a result, the similarity values that can be calculated. Nonetheless, we also believe that the expertise of tower controllers in our experiment might have contributed to larger similarity values. The expert controllers, due to their years of experience and familiarity with the tower cabin environment, know what information they need prior to issuing an aircraft to take off, as well as where that information is located in the environment. Note that these prior studies did not specifically use the Jaccard coefficient similarity, however, they did calculate the similarity of visual scan paths based upon the AOIs they share, and thus, the resulting values should be very closely related.

Overall, researchers might need to account for and consider how Jaccard coefficient similarity and string-edit similarity values might be impacted by the task, environment, as well as the participants' expertise when comparing similar visual scanning patterns in order to identify similar ones. Furthermore, using multiple similarity metrics simultaneously might be particularly useful in tasks where participants can have multiple variations of similar visual scanning behaviors.

Similarities in the visual scan paths of tower controllers when issuing clear to take off clearances

In our small sample, most of the visual scan paths of tower controllers (6 out of 9 visual scan paths) had an above average Jaccard coefficient similarity. All visual scan paths contained both radar AOIs (A and L), runway AOIs (T, R, and I), some of the most commonly inspected AOIs reported in prior research. For example, Manske & Schier (2015) found that, on average, AOIs in the tower cabin environment such as the flight strips, radar displays, and runways have a higher probability of being observed by tower controllers while issuing a clear to take off than other AOIs such as the departure corridor AOI. Nonetheless, these results might not be unexpected, as controllers have to ensure that there are no hazards on the runway while issuing a clear to take off to, for example, avoid runway incursions (Federal Aviation Administration, 2022). To ensure there are no hazards on the runway would be to inspect AOIs associated with the runways, such as the intersection AOI (I) and the touchdown (T), as well as the ASDE radar AOI (A).

Controllers in our study gathered the information required to issue a clear to take off from multiple AOIs that provided the same or similar information. Prior studies have described that, depending on the situation, controllers might look directly at the runway or the ground radar while issuing a clear to land clearance (Svensson, 2015). In other words, two areas of the tower cabin environment that provide the same information (i.e., the runway is clear) to the controller. A similar behavior might be applied by controllers when issuing a clear to take off clearance. For example, one visual scan paths did not contain the flight strips AOI (S7) and another visual scan path did not contain the runway intersection AOI (S6), two of the most commonly observed AOIs by the controllers. However, these controllers were able to gather the information required to issue the clearance by observing other AOIs, such as the ASDE radar (A), which contains the callsign of the aircraft that was ready for departure, rather than looking at the flight strips, as well as to determine that there were no hazards present on the runway for the aircraft. Furthermore, participants observed AOIs

that contained the same information multiple times, such as the runway hotspot AOIs (T and I) as well as the ground radar AOI (A).

In addition, some controllers (5 out of 9 controllers) applied similar visual search patterns when gathering information from the tower environment to issue a clear to take off clearance (i.e. had an above average string-edit similarity). One of the common patterns applied by these controllers was to inspect the radar AOIs (L and A) one after the other (i.e., AL or LA), after which they scanned scan the hotspot AOIs (T and I) back and forth (i.e., TIT). These can be seen particularly in the case of S5 and S3, where S5 contains the sequence "LATITS" while S3 contains the sequence "LALFTITS", which have a string-edit similarity of 0.75 between each other. The other controllers had similar variations of these patterns, with S8 applying "LAITS" and S2 using the pattern "ALFTIT".

Although future research is needed, one potential reason behind these patterns might be that inspecting radar AOIs first, followed by runway hotspot AOIs might serve as a way for controllers to build "the picture" (i.e., their situational awareness) (Niessen & Eyferth, 2001) of the airport environment, and immediately validating the picture with new sources of information. Thus, it might be possible that in some circumstances the order the expert inspects AOIs might not be as important as long as the required information is gathered in a timely manner. Furthermore, if any mismatch exists were to occur between the mental picture built by the controller and the information they are gathering to validate it, the controller has an opportunity to quickly instruct the pilots to disregard the clearance without affecting the movement or safety of aircraft.

Further future research could include identifying and using commonly applied visual search strategies exhibited by expert tower controllers, who can generally identify and extract information from the environment faster than novices (Cokely et al., 2018), as a potential aid to the training of

novice tower controllers. Prior studies in en-route air traffic control have described how showing the visual scan paths of experts to novices as an intervention can lead to improved performance on an en-route air traffic control task (Kang & Landry, 2014). Furthermore, other research has shown that eye movement modeling examples (EMMEs), such as when the eye movement of an expert are shown alongside their explanation and rationale regarding how they completed the task, might lead to positive effects on learning and task performance (Emhardt et al., 2022). Thus, novices the visual search strategies of experts while they verbally describe their decision-making process as they complete the task might help novices learn the decision-making process applied by experts. The verbal input provided by the experts might highlight portions of their heuristic deliberation, a driver of superior-decision making in Skilled Decision Theory (Cokely et al, 2018), while the visual scan paths could showcase what informed their heuristic deliberation.

Chapter 6 – Conclusions

The present dissertation set out to address two challenges that researchers encounter when conducting eye tracking research: (1) identifying and selecting accurate thresholds for eye movement detection algorithms; (2) identifying similar visual scan paths when variations of a common underlying visual scanning behavior exist.

First, identifying and selecting accurate thresholds is crucial when using eye movement detection algorithms, as thresholds with poor performance may fail to identify eye fixations that took place or even combine multiple eye fixations into a singular eye fixation that never took place. As mentioned previously, current approaches to determine accurate thresholds require the researcher to know the eye movements of participants ahead of time, limiting their applicability to certain experimental designs or environments, or are largely manually laborious tasks that can be time consuming.

Chapters 3 and 4 introduced new approaches to tackle this challenge and aid researchers in identifying and selecting accurate thresholds automatically for their respective applications.

Chapter 3 presented an automated procedure that analyzes how threshold values impact eye movement trends in order to determine and suggest a range of thresholds to the researcher – extending upon the original work of Blignaut (2009). The automated procedure identified elbow/knee points using the Kneedle algorithm in eye movement metric trends by determining changes in concavity and slope direction that occur as a result of the impact of threshold values without user intervention. Our results showed that the proposed automated procedure identified accurate thresholds. More specifically, the number of participants with string-edit similarity to the ideal visual scan path above 80% from 252 at 30 °/s, one of the default recommended values in Tobii Pro Lab, to 287 at both recommended thresholds (50 °/s and 120 °/s). In other words, the

automated procedure identified thresholds that increased the performance above 80% for 12.5% of participants. Furthermore, 24 participants had a string-edit similarity of 40% or less at 30 °/s, while only 2 participants had a string-edit similarity of 40 % or less at 50 °/s, and only 3 participants at 120 °/s.

Chapter 4 introduced two approaches, within-participant and between-participant comparisons, to quantify the impact of thresholds on string-edit similarity. The within-participant method compared the visual scan path at one threshold to the visual scan paths created at every other threshold, while the between-participant methods compared the visual scan paths of participants at each threshold to each other. These two methods were capable of approximating the relationship between string-edit similarity values of an ideal visual scan path and thresholds reported in prior literature. As a result, we were able to use these two procedures to identify accurate thresholds for eye movements of participants in two different experiments, one involving a simple task where participants followed the movement of a bullseye target as instructed, and the other a more complex air traffic control task in a high-fidelity simulator. For the first experiment, the between-participant comparisons were a closer approximation of the relationship between string-edit similarity values of an ideal visual scan path, allowing us to determine a threshold range between 50 °/s and 180 °/s with an average similarity above 90%. For the second experiment, the within-participant comparisons were a closer approximation, allowing us to identify a threshold range between 50 °/s and 230 °/s with an average similarity above 80%.

Second, identifying similar visual scan paths using solely one similarity metric might not be sufficient, as participants applying variations of a common visual scanning behavior could appear to be more different than they really are. Participants may change the order in which they observe the same information, or observe different sources of information in the environment but follow

very similar patterns between the few sources of information they had in common. All these potential variations influence the resulting similarity scores when using metrics such as string-edit similarity and the Jaccard coefficient similarity.

Chapter 5 addresses this challenge by introducing a classification framework that combines multiple similarity metrics in order to identify similar visual scan paths, as well as potential variations that may exist. More specifically, the framework used the average string-edit similarity and Jaccard coefficient similarity values between all participants to classify participants into high and/or low similarity values. The classification framework was applied to an experiment in which air traffic controllers in a high fidelity tower simulator were managing air traffic and issuing clear to take off clearances. Using multiple similarity metrics allowed us to identify 3 controllers that had an above average similarity to other controllers in at least one metric (i.e., they observed the same information or used very similar patterns) that would have been considered as different if only one similarity metric was used.

Limitations & future research

Given the importance of analyzing and comparing visual scan paths, multiple similarity metrics have been proposed and employed in eye tracking research that address some limitations faced by string-edit similarity and Jaccard coefficient similarity. For example, metrics such as MultiMatch (Jarodzka, Holqmvist, & Nyström, 2010) or ScanMatch (Cristino et al., 2010) consider the duration or shape of the visual scan path when comparing two visual scan paths. Incorporating these metrics into the procedures and approaches introduced in the present dissertation could potentially provide additional information with which to identify similar visual scan paths in our classification framework (Chapter 5) or enable new within-participant or between-participant comparisons across different similarity metrics (Chapter 4).

Furthermore, the present dissertation focused on the I-VT algorithm and the gaze velocity threshold it uses to identify eye fixations and saccadic movements from eye movement data collected by an eye tracker. Nonetheless, eye movement detection algorithms tend to use multiple thresholds simultaneously – in this dissertation, we used 60 ms as the threshold for the minimum eye fixation duration. Future research should seek to expand the approaches discussed in both Chapter 3 and Chapter 4 to consider multiple thresholds simultaneously, providing researchers with all the thresholds needed to accurate model visual scan paths in their respective applications, as well as different eye movement detection algorithms, such as I-DT. In addition, other algorithms are capable of identifying eye movements beyond eye fixations and saccadic movements, such as smooth pursuits (Komogortsev & Karpov, 2013), which could be investigated in future research.

References

Anderson, N. C., Anderson, F., Kingstone, A., & Bischof, W. F. (2015). A comparison of scanpath comparison methods. *Behavior research methods*, *47*, 1377-1392. https://doi.org/10.3758/s13428-014-0550-3

Andersson, R., Larsson, L., Holmqvist, K., Stridh, M., & Nyström, M. (2017). One algorithm to rule them all? An evaluation and discussion of ten eye movement event-detection algorithms. *Behavior research methods*, *49*, 616-637. https://doi.org/10.3758/s13428-016-0738-9

Arvai, K. (2020). kneed (Version 0.7.0) [Computer software]. https://doi.org/10.5281/zenodo.6496267

Beelders, T. R., & du Plessis, J. P. L. (2016). Syntax highlighting as an influencing factor when reading and comprehending source code. *Journal of Eye Movement Research*, *9*(1). https://doi.org/10.16910/jemr.9.1.1

Besta, M., Kanakagiri, R., Mustafa, H., Karasikov, M., Rätsch, G., Hoefler, T., & Solomonik, E. (2020). Communication-efficient jaccard similarity for high-performance distributed genome comparisons. In *2020 IEEE International Parallel and Distributed Processing Symposium* (IPDPS) (pp. 1122-1132). IEEE. https://doi.org/10.1109/IPDPS47924.2020.00118

Birawo, B., & Kasprowski, P. (2022). Review and evaluation of eye movement event detection algorithms. *Sensors*, *22*(22), 8810. https://doi.org/10.3390/s22228810

Blignaut, P. (2009). Fixation identification: The optimum threshold for a dispersion algorithm. *Attention, Perception, & Psychophysics*, *71*, 881-895. https://doi.org/10.3758/APP.71.4.881

Blignaut, P., & Beelders, T. (2009). The effect of fixational eye movements on fixation identification with a dispersion-based fixation detection algorithm. *Journal of eye movement research*, *2*(5). https://doi.org/10.16910/jemr.2.5.4

Borji, A., & Itti, L. (2014). Defending Yarbus: Eye movements reveal observers' task. *Journal of vision*, *14*(3), 29-29. https://doi.org/10.1167/14.3.29.

Brunyé, T. T., Drew, T., Kerr, K. F., Shucard, H., Weaver, D. L., & Elmore, J. G. (2020). Eye tracking reveals expertise-related differences in the time-course of medical image inspection and diagnosis. *Journal of Medical Imaging*, *7*(5), 051203-051203. https://doi.org/10.1117/1.JMI.7.5.051203

Burch, M., Kumar, A., Mueller, K., Kervezee, T., Nuijten, W., Oostenbach, R., ... & Smit, G. (2019). Finding the outliers in scanpath data. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications* (pp. 1-5). https://doi.org/10.1145/3317958.3318225

Cristino, F., Mathôt, S., Theeuwes, J., & Gilchrist, I. D. (2010). ScanMatch: A novel method for comparing fixation sequences. *Behavior research methods*, *42*, 692-700. https://doi.org/10.3758/BRM.42.3.692

Crutchfield, J., Kang, Z., Fraga, R. P., & Mandal, S. (2021). Applying eye-tracking technology to explore the visual scanning practices of air traffic control tower controllers. In *21st International Symposium on Aviation Psychology* (p. 60). https://corescholar.libraries.wright.edu/isap_2021/11

Cokely, E. T., Feltz, A., Ghazal, S., Allan, J. N., Petrova, D., & Garcia-Retamero, R. (2018). Skilled decision theory: From intelligence to numeracy and expertise. In K. A. Ericsson, R. R. Hoffman, A. Kozbelt, & A. M. Williams (Eds.), The Cambridge handbook of expertise and expert performance (2nd ed., pp. 476–505). Cambridge University Press. https://doi.org/10.1017/9781316480748.026

Davies, A. R., Vigo, M., Harper, S., & Jay, C. (2018). Using simultaneous scanpath visualization to investigate the influence of visual behaviour on medical image interpretation. *Journal of Eye Movement Research*, *10*(5). https://doi.org/10.16910/jemr.10.5.11

DeAngelus, M., & Pelz, J. B. (2009). Top-down control of eye movements: Yarbus revisited. *Visual Cognition*, *17*(6-7), 790-811. https://doi.org/10.1080/13506280902793843

Dewhurst, R., Nyström, M., Jarodzka, H., Foulsham, T., Johansson, R., & Holmqvist, K. (2012). It depends on how you look at it: Scanpath comparison in multiple dimensions with MultiMatch, a vector-based approach. *Behavior research methods*, *44*, 1079-1100. https://doi.org/10.3758/s13428-012-0212-2

Duchowski, A. T., Driver, J., Jolaoso, S., Tan, W., Ramey, B. N., & Robbins, A. (2010). Scanpath comparison revisited. In *Proceedings of the 2010 symposium on eye-tracking research & applications* (pp. 219-226). https://doi.org/10.1145/1743666.1743719

Duchowski, A. T., Jörg, S., Allen, T. N., Giannopoulos, I., & Krejtz, K. (2016, March). Eye movement synthesis. In *Proceedings of the ninth biennial ACM symposium on eye tracking research & applications* (pp. 147-154). https://doi.org/10.1145/2857491.2857528

Emhardt, S. N., Kok, E., van Gog, T., Brandt-Gruwel, S., van Marlen, T., & Jarodzka, H. (2023). Visualizing a task performer's gaze to foster observers' performance and learning—A systematic literature review on eye movement modeling examples. *Educational Psychology Review*, 35(1), 23. https://doi.org/10.1007/s10648-023-09731-7

Eraslan, S., Yesilada, Y., & Harper, S. (2016). Eye tracking scanpath analysis techniques on web pages: A survey, evaluation and comparison. *Journal of Eye Movement Research*, 9(1). https://doi.org/10.16910/jemr.9.1.2

Fahimi, R., & Bruce, N. D. (2021). On metrics for measuring scanpath similarity. *Behavior Research Methods*, *53*, 609-628. https://doi.org/10.3758/s13428-020-01441-0

Federal Aviation Administration. (2022). Runway Incursions. U.S. Department of Transportation. https://www.faa.gov/airports/runway_safety/resources/runway_incursions

Federal Aviation Administration. (2023). *Air Traffic By The Numbers.* U.S. Department of Transportation.

https://www.faa.gov/air_traffic/by_the_numbers/media/Air_Traffic_by_the_Numbers_2023.pdf

Ferdous, R. (2009). An efficient k-means algorithm integrated with Jaccard distance measure for document clustering. In *2009 first asian himalayas international conference on internet* (pp. 1-6). IEEE. https://doi.org/10.1109/AHICI.2009.5340335

Fletcher, S., & Islam, M. Z. (2018). Comparing sets of patterns with the Jaccard index. *Australasian Journal of Information Systems*, *22*. https://doi.org/10.3127/ajis.v22i0.1538

Goldberg, J. H., & Helfman, J. I. (2010). Scanpath clustering and aggregation. In *Proceedings of the 2010 symposium on eye-tracking research & applications* (pp. 227-234). https://doi.org/10.1145/1743666.1743721

Goldberg, J. H., & Helfman, J. I. (2010). Identifying aggregate scanning strategies to improve usability evaluations. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 54, No. 6, pp. 590-594). Sage CA: Los Angeles, CA: SAGE Publications. https://doi.org/10.1177/154193121005400610

Griffith, H., Lohr, D., Abdulin, E., & Komogortsev, O. (2021). GazeBase, a large-scale, multi-stimulus, longitudinal eye movement dataset. *Scientific Data*, *8*(1), 184. https://doi.org/10.1038/s41597-021-00959-y

Guralnik, V., & Srivastava, J. (1999). Event detection from time series data. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 33-42). https://doi.org/10.1145/312129.312190

Hahn, L., & Klein, P. (2022). Eye tracking in physics education research: A systematic literature review. *Physical Review Physics Education Research*, 18(1), 013102. https://doi.org/10.1103/PhysRevPhysEducRes.18.013102

Harężlak, K., & Kasprowski, P. (2014). Evaluating quality of dispersion based fixation detection algorithm. In *Information Sciences and Systems 2014: Proceedings of the 29th International Symposium on Computer and Information Sciences* (pp. 97-104). Springer International Publishing. https://doi.org/10.1007/978-3-319-09465-6_11

Hermens, F., Flin, R., & Ahmed, I. (2013). Eye movements in surgery: A literature review. *Journal of Eye movement research*, *6*(4). https://doi.org/10.16910/jemr.6.4.4

Holmqvist, K. (2016). Optimal settings for commercial event detection algorithms based on the level of noise. [Unpublished manuscript]

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.

Hooge, I. T., Niehorster, D. C., Nyström, M., Andersson, R., & Hessels, R. S. (2018). Is human classification by experienced untrained observers a gold standard in fixation detection?. *Behavior Research Methods*, *50*, 1864-1881. https://doi.org/10.3758/s13428-017-0955-x

Hosp, B., Eivazi, S., Maurer, M., Fuhl, W., Geisler, D., & Kasneci, E. (2020). RemoteEye: An open-source high-speed remote eye tracker: Implementation insights of a pupil-and glint-detection algorithm for high-speed remote eye tracking. *Behavior research methods*, *52*, 1387-1401. https://doi.org/10.3758/s13428-019-01305-2

Hsu, C. W., Chang, C. C., & Lin, C. J. (2003). A practical guide to support vector classification.

Huang, A. (2008). Similarity measures for text document clustering. In *Proceedings of the sixth new zealand computer science research student conference* (NZCSRSC2008), Christchurch, New Zealand (Vol. 4, pp. 9-56).

Jaccard, P. (1901). Etude comparative de la distribution florale dans une portion des Alpes et du Jura. *Bulletin de la Société's Vaudoise des Sciences Naturelles*, 37(1): 547–579

Jarodzka, H., Holmqvist, K., & Nyström, M. (2010). A vector-based, multidimensional scanpath similarity measure. In *Proceedings of the 2010 symposium on eye-tracking research & applications* (pp. 211-218). https://doi.org/10.1145/1743666.1743718

Jeong H., Kang. Z., and Liu Y. (2019). Driver glance behaviors and scanning patterns: Applying static and dynamic glance measures to the analysis of curve driving with secondary tasks. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 29(6). https://doi.org/10.1002/hfm.20798

Jones, S. R. (2019). Students' application of concavity and inflection points to real-world contexts. *International Journal of Science and Mathematics Education*, *17*, 523-544. https://doi.org/10.1007/s10763-017-9876-5

Josephson, S., & Holmes, M. E. (2002). Visual attention to repeated internet images: testing the scanpath theory on the world wide web. In *Proceedings of the 2002 symposium on Eye tracking research & applications* (pp. 43-49). https://doi.org/10.1145/507072.507081

Kang, Z., & Landry, S. J. (2014). An eye movement analysis algorithm for a multielement target tracking task: Maximum transition-based agglomerative hierarchical clustering. *IEEE Transactions on Human-Machine Systems*, *45*(1), 13-24. https://doi.org/10.1109/THMS.2014.2363121

Kang, Z., & Landry, S. J. (2014). Using scanpaths as a learning method for a conflict detection task of multiple target tracking. *Human factors*, *56*(6), 1150-1162. https://doi.org/10.1177/0018720814523066

Kang, Z., Crutchfield, J., Fraga, R. P., & Mandal, S. (2021). Spatial-temporal cluster approach to discover visual scanning behaviors in virtual reality. In *21st International Symposium on Aviation Psychology* (p. 66). https://corescholar.libraries.wright.edu/isap_2021/12

Kang, Z., Mandal, S., Crutchfield, J., Millan, A., & McClung, S. N. (2016). Designs and algorithms to map eye tracking data with dynamic multielement moving objects. *Computational intelligence and neuroscience*, *2016*. https://doi.org/10.1155/2016/9354760

Komogortsev, O. V., & Karpov, A. (2013). Automated classification and scoring of smooth pursuit eye movements in the presence of fixations and saccades. *Behavior research methods*, *45*, 203-215. https://doi.org/10.3758/s13428-012-0234-9

Komogortsev, O. V., Gobert, D. V., Jayarathna, S., & Gowda, S. M. (2010). Standardization of automated analyses of oculomotor fixation and saccadic behaviors. *IEEE Transactions on biomedical engineering*, *57*(11), 2635-2645. https://doi.org/10.1109/TBME.2010.2057429

Kumar, A., Timmermans, N., Burch, M., & Mueller, K. (2019). Clustered eye movement similarity matrices. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications* (pp. 1-9). https://doi.org/10.1145/3317958.3319811

Larsson, G. (2010). Evaluation methodology of eye movement classification algorithms [Master's Thesis, Royal Institute of Technology].

Leube, A., Rifai, K., & Wahl, S. (2017). Sampling rate influences saccade detection in mobile eye tracking of a reading task. *Journal of eye movement research*, *10*(3). https://doi.org/10.16910/jemr.10.3.3

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady* (Vol. 10, No. 8, pp. 707-710).

Llanes-Jurado, J., Marín-Morales, J., Guixeres, J., & Alcañiz, M. (2020). Development and calibration of an eye-tracking fixation identification algorithm for immersive virtual reality. *Sensors*, *20*(17), 4956. https://doi.org/10.3390/s20174956

Mandal, S. and Kang. Z. (2018). Using eye movement data visualization to enhance training of air traffic controllers: A dynamic network approach. *Journal of Eye Movement Research, 11*(4), 1-20. https://doi.org/10.16910/jemr.11.4.1

Mandal, S., & Kang, Z. (2015). Eye Tracking Analysis Using Differently Shaped Areas of Interest to Represent Multi-Element Moving Objects. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 59, No. 1, pp. 1515-1519). Sage CA: Los Angeles, CA: SAGE Publications. https://doi.org/10.1177/1541931215591328

Manske, P. G., & Schier, S. L. (2015). Visual scanning in an air traffic control tower–A simulation study. *Procedia Manufacturing*, 3, 3274-3279. https://doi.org/10.1016/j.promfg.2015.07.397

McClung, S. N., & Kang, Z. (2016). Characterization of visual scanning patterns in air traffic control. *Computational intelligence and neuroscience*, *2016*. https://doi.org/10.1155/2016/8343842

Meyer, L., Josefsson, B., Vrotsou, K., Westin, C., & Lundberg, J. (2021). Evaluation of an AOI mapping and analysis tool for the identification of visual scan pattern. In *2021 IEEE/AIAA 40th Digital Avionics Systems Conference* (DASC) (pp. 1-8). IEEE. https://doi.org/10.1109/DASC52595.2021.9594500

Naeeri Ph D, S. M., Kang Ph D, Z., & Palma Fraga MS, R. (2022). Investigation of Pilots' Visual Entropy and Eye Fixations for Simulated Flights Consisted of Multiple Take-Offs and Landings. *Journal of Aviation/Aerospace Education & Research*, *31*(2), 2. https://doi.org/10.15394/jaaer.2022.1920

Navarro, J., Lappi, O., Osiurak, F., Hernout, E., Gabaude, C., & Reynaud, E. (2021). Dynamic scan paths investigations under manual and highly automated driving. *Scientific reports*, *11*(1), 3776. https://doi.org/10.1038/s41598-021-83336-4

Niessen, C., & Eyferth, K. (2001). A model of the air traffic controller's picture. Safety Science, 37(2-3), 187-202. https://doi.org/10.1016/S0925-7535(00)00048-5

Nodine, C. F., & Kundel, H. L. (1987). Using eye movements to study visual search and to improve tumor detection. *Radiographics*, *7*(6), 1241-1250. https://doi.org/10.1148/radiographics.7.6.3423330

Nyström, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior research methods*, *42*(1), 188-204. https://doi.org/10.3758/BRM.42.1.188

Olsen, A. (2012). The Tobii I-VT fixation filter. *Tobii Technology*, *21*, 4-19.

Olsen, A., & Matos, R. (2012). Identifying parameter values for an I-VT fixation filter suitable for handling data sampled with various sampling frequencies. In *proceedings of the symposium on Eye tracking research and applications* (pp. 317-320). https://doi.org/10.1145/2168556.2168625

Onumanyi, A. J., Molokomme, D. N., Isaac, S. J., & Abu-Mahfouz, A. M. (2022). AutoElbow: An automatic elbow detection method for estimating the number of clusters in a dataset. *Applied Sciences*, *12*(15), 7515. https://doi.org/10.3390/app12157515

Orquin, J. L., & Holmqvist, K. (2018). Threats to the validity of eye-movement research in psychology. *Behavior research methods*, *50*, 1645-1656. https://doi.org/10.3758/s13428-017-0998-z

Ouzts, A. D., & Duchowski, A. T. (2012, March). Comparison of eye movement metrics recorded at different sampling rates. In *Proceedings of the symposium on eye tracking research and applications* (pp. 321-324). https://doi.org/10.1145/2168556.2168626

Palma Fraga, R., Kang, Z., Crutchfield, J. M., & Mandal, S. (2021). Visual search and conflict mitigation strategies used by expert en route air traffic controllers. *Aerospace*, *8*(7), 170. https://doi.org/10.3390/aerospace8070170

Privitera, C. M., & Stark, L. W. (2000). Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *IEEE Transactions on pattern analysis and machine intelligence*, *22*(9), 970-982. https://doi.org/10.1109/34.877520

Raza, M. A., Kiran, R., Ghazal S., Kang, Z., Salehi, S., Cokely, E., Jeon, J. (2023). An eye tracking based framework for safety improvement of offshore operations. *Journal of Eye Movement Research*, *16*(3), 1-17. https://doi.org/10.16910/jemr.16.3.2

Robinski, M., & Stein, M. (2013). Tracking visual scanning techniques in training simulation for helicopter landing. *Journal of Eye Movement Research*, *6*(2). https://doi.org/10.16910/jemr.6.2.3

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications* (pp. 71-78). https://doi.org/10.1145/355017.355028

Satopaa, V., Albrecht, J., Irwin, D., & Raghavan, B. (2011). Finding a "kneedle" in a haystack: Detecting knee points in system behavior. In *2011 31st international conference on distributed computing systems workshops* (pp. 166-171). IEEE. https://doi.org/10.1109/ICDCSW.2011.20

Savitzky, A., & Golay, M. J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, *36*(8), 1627-1639.

Shi, C., Wei, B., Wei, S., Wang, W., Liu, H., & Liu, J. (2021). A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm. *Eurasip Journal on Wireless Communications and Networking*, *2021*(1), 1-16. https://doi.org/10.1186/s13638-021-01910-w

Shic, F., Scassellati, B., & Chawarska, K. (2008). The incomplete fixation measure. In *Proceedings of the 2008 symposium on Eye tracking research & applications* (pp. 111-114). https://doi.org/10.1145/1344471.1344500

Špakov, O., Siirtola, H., Istance, H., & Kari-Jouko, R. (2017). Visualizing the reading activity of people learning to read. *Journal of Eye Movement Research*, *10*(5). https://doi.org/10.16910/jemr.10.5.5

SR Research. (2018). *EyeLink 1000 Plus* [Brochure] https://www.sr-research.com/wp-content/uploads/2018/01/EyeLink-1000-Plus-Brochure.pdf

Startsev, M., & Zemblys, R. (2023). Evaluating eye movement event detection: A review of the state of the art. *Behavior Research Methods*, *55*(4), 1653-1714. https://doi.org/10.3758/s13428-021-01763-7

Stewart, J., Clegg, D. K., & Watson, S. (2020). *Calculus: early transcendentals*. Cengage Learning.

Strohmaier, A. R., MacKay, K. J., Obersteiner, A., & Reiss, K. M. (2020). Eye-tracking methodology in mathematics education research: A systematic literature review. *Educational Studies in Mathematics*, 104, 147-200. https://doi.org/10.1007/s10649-020-09948-1

Svensson, Å. (2015). Air traffic controllers' work-pattern during air traffic control tower simulations: A eye-tracking study of air traffic controllers' eye-movements during arrivals. [Master's Thesis, Linköping University].

Tang, H., Day, E., Kendhammer, L., Moore, J. N., Brown, S. A., & Pienta, N. J. (2016). Eye movement patterns in solving science ordering problems. *Journal of eye movement research*, *9*(3). https://doi.org/10.16910/jemr.9.3.6

Tobii. (2020). *User's manual Tobii Pro Glasses 2* [Brochure] http://andrewd.ces.clemson.edu/courses/cpsc881/manuals/Tobii/Tobii-Pro-Glasses-2-User-Manual.pdf

Trabulsi, J., Norouzi, K., Suurmets, S., Storm, M., & Ramsøy, T. Z. (2021). Optimizing fixation filters for eye-tracking on small screens. *Frontiers in Neuroscience*, *15*, 1257. https://doi.org/10.3389/fnins.2021.578439

Transportation Security Administration. (2024). *2023 Year in Review.* U.S. Department of Homeland Security. https://www.tsa.gov/sites/default/files/17228_info_tsa_2023_year_in_review.pdf

Underwood, G. (2007). Visual attention and the transition from novice to advanced driver. *Ergonomics*, *50*(8), 1235-1249. https://doi.org/10.1080/00140130701318707

Van der Gijp, A., Ravesloot, C. J., Jarodzka, H., Van der Schaaf, M. F., Van der Schaaf, I. C., van Schaik, J. P., & Ten Cate, T. J. (2017). How visual search relates to visual diagnostic performance: a narrative systematic review of eye-tracking research in radiology. *Advances in Health Sciences Education*, *22*, 765-787. https://doi.org/10.1007/s10459-016-9698-1

Van der Lans, R., Wedel, M., & Pieters, R. (2011). Defining eye-fixation sequences across individuals and tasks: the Binocular-Individual Threshold (BIT) algorithm. *Behavior research methods*, *43*, 239-257. https://doi.org/10.3758/s13428-010-0031-2

Waite, S., Scott, J., Gale, B., Fuchs, T., Kolla, S., & Reede, D. (2017). Interpretive error in radiology. *American Journal of Roentgenology*, *208*(4), 739-749. https://doi.org/10.2214/AJR.16.16963

Wolfe, J. M. (2010). Visual search. *Current biology*, *20*(8), R346-R349. https://doi.org/10.1016/j.cub.2010.02.016

Wolfe, J. M. (2020). Visual search: How do we find what we are looking for?. *Annual review of vision science*, *6*, 539-562. https://doi.org/10.1146/annurev-vision-091718-015048

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

Yoo, S., Jeong, S., & Jang, Y. (2021). Gaze behavior effect on gaze data visualization at different abstraction levels. *Sensors*, *21*(14), 4686. https://doi.org/10.3390/s21144686