Grant Agreement No: 687591

28/12/18

Big Data Analytics for Time Critical Mobility Forecasting

dat Acron

D4.9 Evaluating VA methods in several scenarios and workflows

Deliverable Form	
Project Reference No.	H2020-ICT-2015 687591
Deliverable No.	4.9
Relevant Work Package:	WP 4
Nature:	R
Dissemination Level:	PU
Document version:	1.0.0
Due Date:	2018-12-31
Date of latest revision:	2018-12-28
Completion Date:	2018-12-28
Lead partner:	Fraunhofer IAIS
Authors:	Georg Fuchs, Nathalie Paul
Reviewers:	
Document description:	This deliverable reports on the evaluation of Visual Analytics and visualizations used in corresponding usage scenarios including eye tracking experiments.
Document location:	Documents/datAcron/WP4/Deliverables/Final

HISTORY OF CHANGES

Version	Date	Changes	Author	Remarks
0.1.0	2018-12-05	Initial draft	Georg Fuchs	Document structure
0.1.1	2018-12-05	Eye tracking results	Nathalie Paul	Heatmap & gaze visualizations
0.5.0	2018-12-20	Completed chapter 3	Georg Fuchs	Methodological background
0.6.0	2018-12-21	Completed chapter 4	Georg Fuchs	Eye tracking results
0.7.0	2018-12-27	Added chapter 2	Georg Fuchs	Eye tracking results
1.0.0	2018-12-28	Consolidation & Proofreading	Georg Fuchs	Initial submission version

EXECUTIVE SUMMARY

The Visual Analytics (VA) methods developed in datAcron work package 4, tasks T4.1-T4.4 have been evaluated in corresponding usage scenarios with appropriate categories of professional users. Offline VA methods and tools have primarily been relevant to the aviation use case scenarios and targeted data science specialists focused on batch (offline) data processing and analysis. Since these applications are not time-critical, a pragmatic evaluation of the proposed methods and tools in terms of analysis task coverage has been followed, as outlined in deliverables D6.5 and D6.6. Online, real-time visualizations have mainly been relevant to the maritime use case scenarios, and specifically, have been geared towards supporting scenario assessment at the MSI level (thus leveraging datAcron's LED – SI – CEP/F – T/FLP pipeline), as detailed in D5.5 and D5.6. Since this MSI-supported maritime situational assessment does have a time-critical component, selected scenarios have been subjected by eye-tracking evaluation that provides initial indications towards task performance and problem solving strategies, thereby complementing the results reported in D5.6.

TABLE OF CONTENTS

TERMS & ABBREVIATIONS

ACC	Area Control Center
ADS-B	Automatic Dependent Surveillance - Broadcast
AIS	Automatic Identification System (for maritime vessels)
ANSP	Air Navigation Service Provider
AOI	Area of Interest
ATC	Air Traffic Control
ATCC	Air Traffic Control Center
ATFCM	Air Traffic Flow and Capacity Management
ATM	Air traffic Management
CEF	Complex Event Forecasting
CER	Complex Event Recognition
CTC	Central Trajectory of Cluster
CFMU	Control Flow Management Unit
DDR	Demand Data Repository (EuroControl)
DM	Data Manager
FL	Flight Level
FLP	Future Location Predictor
FM	Flow Management (Scenario)
GIS	Geographic Information System
GPS	Global Positioning [Satellite] System
IFS	InForme de Seguimiento (Spanish: tracking information)
IVA	Interactive Visual Analytics
LED	Low-level Event Detection
MSA	Maritime Situational Awareness
MSI	Maritime Situational Indicator
NOAA	National Oceanic and Atmospheric Administration
OGC	Open Geospatial Consortium
SG	Synopses Generator
SI	Semantic Integrator
STG	State Transition Graph
VA	Visual Analytics
WP	Work Package
WMS	Web Mapping Service

LIST OF FIGURES

List of Figures

LIST OF TABLES

List of Tables

1 INTRODUCTION

1.1 Purpose and Scope

In alignment with datAcron work package WP4, Task 4.5 "Evaluating VA methods in several scenarios and workflows", this deliverable reports on results and findings from the evaluation of the various visual analytics (VA) methods that have been evaluated in corresponding usage scenarios with appropriate categories of professional users. The underlying visual analytics methods and corresponding visualization tools have been developed in datAcron work package 4, tasks T4.1–T4.4 and reported in detail in deliverables D4.5–D4.8, respectively.

Offline VA methods and tools have primarily been relevant to the aviation use case scenarios and targeted data science specialists focused on batch (offline) data processing and analysis. Since these applications are not time-critical, a pragmatic evaluation of the proposed methods and tools in terms of analysis task coverage has been followed, as outlined in deliverables D6.5 and D6.6.

Online, real-time visualizations have mainly been relevant to the maritime use case scenarios, and specifically, have been geared towards supporting scenario assessment at the MSI level (thus leveraging datAcron's LED – SI – CEP/F – T/FLP real-time processing pipeline), as detailed in D5.5 and D5.6. Since this MSI-supported maritime situational assessment does have a time-critical component, selected scenarios have been aligned with limited eye-tracking evaluation that provides initial indications towards task performance and problem solving strategies. The present document reports on the findings from the analysis of the obtained eye tracking data as well as from adjoining participant interviews, thereby complementing the results reported in D5.6.

1.2 Relation to datAcron Objectives

The related objectives of datAcron are:

O.1 Spatio-temporal data integration and management solutions

O.2 Real-time detection and forecasting accuracy of moving entities' trajectories

O.3 Real-time recognition and prediction of important events concerning these entities

O.4 General visual analytics infrastructure supporting all steps of the analysis through appropriate interactive visualisations

O.5 Producing streaming data synopses at a high rate of compression

While the focus of the present deliverable is O.4 most closely related to the activities in work package WP4, the evaluation of any visualization requires availability of the data to be visualized. The successful integration and visualization of datAcron components for trajectory synopses generation and prediction (T/FLP), complex event recognition and prediction (CER/F) therefore was a prerequisite, related to objectives O.1–O.3 and O.5.

1.3 Relation to Other Deliverables

The present deliverable D4.9 relates to the following deliverables in the described way:

- **D4.6** – VA methods for interactive movement detection: this is the final deliverable of WP4, Task 4.2 that reports on types of patterns that can be detected in movement data and corresponding visual analytics methods and visualization tools for exploring, filtering, and manipulating movement data in spatial, temporal, and spatiotemporal domains.
- **D4.7** – VA methods for interactive movement prediction: this is the final deliverable of WP4, Task 4.3 that describes visualization and interaction techniques and interfaces supporting the exploration and evaluation of results of trajectory prediction algorithms.
- **D4.8** – VA methods for building situation overview and situation monitoring: this is the final deliverable of WP4, Task 4.4 that reports on the real-time architecture, map-based visualization, and interactions developed for real-time situation monitoring. It describes the integration of streaming data generated by the datAcron components Trajectory / Future Location Predictor (T/FLP), Complex Event Recognition / Forecast (CER/F), and synopses generation, as well as interaction patterns; this software has been used for conducting eye tracking experiments as reported in the present document.
- **D5.5** – Maritime Prototype Set-up: this deliverable reports on the preparation of scenario-specific data sets and the overall setup of experiments with regard to maritime scenarios and SC1.1 – Collision Prevention in particular; the eye tracking experiments reported in the present document are based on these data and scenarios.
- **D5.6** – Maritime final validation: this deliverable provides a detailed description of evaluation results obtained for the maritime domain; findings from the eye tracking experiments reported in the present document complement D5.6.
- **D6.5** – Aviation Experiments Specification: this deliverable reports on the preparation of scenario-specific data sets and the consecutive experiment setup with regard to aviation scenarios Flow Management FM01–FM03 and Flight Planing FP01–FP10, specifically including evaluation of the offline visual analytics components developed in the context of WP4, Task 4.2 (deliverable D4.6).
- **D6.6** – Aviation Final Validation Report: this deliverable provides a detailed description of evaluation results obtained for the aviation / ATM domain, specifically including evaluation of the offline visual analytics components developed in the context of WP4, task 4.3 (deliverable D4.7).

1.4 Structure of the Deliverable

This deliverable comprises three parts. Section 2 reports on the qualitative evaluation of the offline visual analytics components used as embedded tools in the process of data curation, scenario data preparation and algorithm results visualization and analysis primarily in the aviation /ATM domain (FM and FP scenarios). The main contribution of this deliverable is the report on the setup, and evaluation of the eye tracking experiments that have been conducted as a part of the maritime scenario evaluation. Section 3 first gives a brief overview of the related work and discusses the analytical workflow used to structure the eye tracking experiments in relation to the overall scenario evaluation. Section 4 then reports on the findings from the actual experiments that have been conducted. Section 5 closes with a summary and an outlook on possible future work.

2 OFFLINE VISUAL ANALYTICS

The offline visual analytics suite has been developed as a flexible and extensible toolbox that allows the formulation and execution of user-defined visual analytics workflows using a variety of both algorithmic and visualization tools. Different workflows and tools are applicable to different analysis tasks ranging from data exploration and curation (WP4 task 4.1), pattern detection (task 4.2), prediction modeling (task 4.3). The general system architecture is described in full detail in D1.12. Principal analytical workflows, algorithmic and visual tools related to tasks 4.1–4.3 have been reported in deliverables D4.5–D4.7, respectively.

These functionality are primarily relevant to the aviation domain (WP6), which focuses on pre-tactical and strategic time scales and as such, offline analysis of batch data.

Evaluation criteria are, then, those of general ability to facilitate domain-specific analysis tasks, in terms of flexibility and functional completeness; i.e., qualitative evaluation criteria have been applied. D6.6 further defines usability and responsiveness of the HCI as evaluation criteria for specific evaluation scenarios, in particular, FM01, FM02, FP07, and FP10. It should be noted that the VA suite is a dedicated expert tool targeting expert analysts, not (operative) end users. As such, usability specifically refers to the usability of appropriate means to manipulate data and algorithms as well as to interactively manipulate dynamically linked information visualizations. It does specifically *not* refer to untrained user experience (UX) in the classic sense of end user interface design.

In all cases and for all four scenarios, the offline VA component successfully demonstrated both good usability and at least adequate to good responsiveness, as reported in D6.6. section 3.2.

In addition, WP4 evaluated the overall system independently in terms of its ability to support the workflows of complex analysis settings in the context of three case studies aimed at understanding decision making processes [1]. These reflect various operational environments and problems where decision policies are unknown a priori, and therefore can neither be predicted nor considered for planning purposes. This variety of scenarios illustrates the potential of these techniques, as well as confirming the overall suite indeed facilitates realistic aviation domain-specific analysis tasks.

2.1 Revealing Route Choice Criteria

This study aimed to reveal the criteria used by airlines in choosing particular flight routes from many possible routes connecting a given origin-destination pair. This translates to a significant improvement in terms of predictability during the pre-tactical phase, in particular for routes near local airspace boundaries for which subtle route changes might imply the appearance or disappearance of hotspots.

As a representative example, flights from Paris to Istanbul were considered (Fig. 1). This example provides many flights conducted by multiple airlines, which take diverse routes crossing the air spaces of different European countries whose navigation charges greatly vary. Some airlines may prefer such flight routes that minimize the navigation costs by avoiding expensive airspaces or traveling shorter distances across such airspaces. The primary questions in the study was to check if indeed some airlines are likely to have such preferences. This study has revealed that, while there are flight operators striving to reduce the navigation costs, this is not the main route choice criterion for the majority of operators, who prefer the possibility to fly at higher

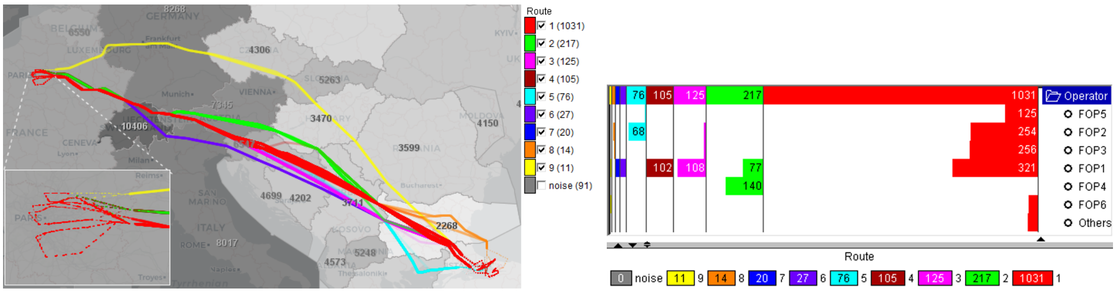

Figure 1: Left: Trajectories according to flight plans have been clustered by route similarity to reveal the major flight routes from Paris to Istanbul. Right: Route choices by six major flight operators FOP1–FOP6.

altitudes as well as higher route stability (i.e., lower deviations), which potentially lead to both better fuel economy and passenger satisfaction due to fewer delays.

The workflow and involved computational and visualization methods, all integrated in the offline VA component, have already been described in D4.6 Section 3.1 and shall not be repeated in detail here.

The objectives of the study were achieved by using filter-aware clustering of trajectories to reveal the major routes, tools for obtaining central trajectories of clusters (CTCs) and for summarizing characteristics of cluster members, and interactive visual displays supporting exploratory analysis of trajectories, clusters, and their attributes (cf. Fig. 1).

This case study was carried out in collaboration with a domain expert from the ATFM domain. The expert attested the offline VA suite enabled this type of data-driven analysis of black-box decision making processes, something not feasible with tools available to him prior [3].

It can therefore be stated that all evaluation criteria: functional completeness, flexibility, usability, and responsiveness have been fulfilled.

2.2 Exploring Separation of Airport Approach Routes

The domain-level goals of this study were three-fold. First, to reconstruct the major approach routes of a major hub in the air transportation system, here specifically, the five airports within Greater London area. Second, to determine which of these approaches may be used simultaneously. Third, to study how the routes that can be used simultaneously are separated in the three-dimensional airspace, i.e., horizontally and vertically. Application of this analysis to TMA allows the understanding of decision making policies, thus facilitating the inclusion of such (previously hidden/implicit) strategies in subsequent prediction modeling.

The workflow and involved computational and visualization methods, all integrated in the offline VA component, have already been described in D4.6 Section 3.2 and shall not be repeated in detail here.

The interactive investigation was able to confirm the general patterns expected by the guiding domain expert [1]: in order to maximize deconfliction, where segments of different approach routes overlap in the horizontal dimension, their altitude ranges overlap as well, and routes intersecting in 2D are separated vertically (Fig. 2).

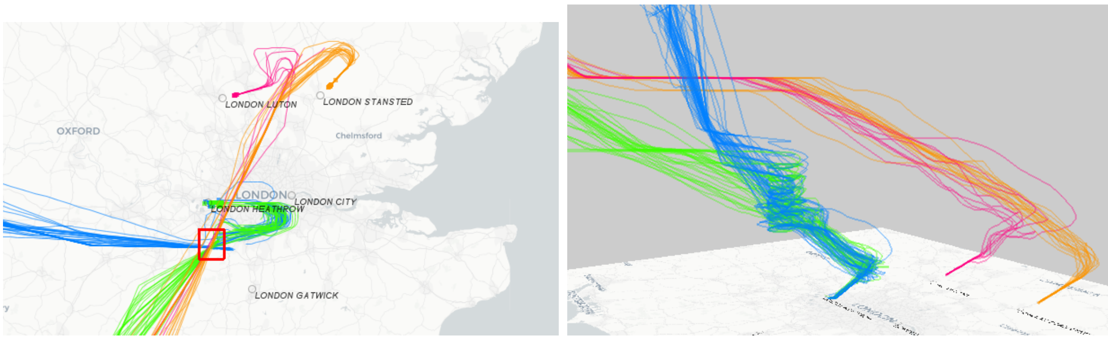

Figure 2: Investigation of the route separation for the five airports within Greater London area.

It can therefore be stated that all evaluation criteria: functional completeness, flexibility, usability, and responsiveness have been fulfilled for the settings of the second case study as well.

2.3 Understanding Airspace Configuration Choices

A sector configuration is a particular division of an airspace region into sectors, such that each sector is managed by a specific number of air traffic controllers (typically two, Executive and Planning Controllers). The number of active sectors depends, on the one hand, on the expected traffic features (such as number of flights within a time interval and their associated complexity/workload given the traffic complexity) and, on the other hand, on the available number of controllers for that given shift (which depends on the strategical demand forecast, which diverges from actual flights for a set of reason).

On the other hand, often there are multiple ways to divide a region into a given number of sectors. The choice of a particular division depends on the flight routes within the region.

Ideally, configurations are chosen so that the demand for the use of the airspace in each sector does not exceed the sector capacity, while making efficient and balanced use of resources (controllers). In reality, demand-capacity imbalances happen quite frequently for a number of reasons (deviations of actual flights from flight plans, weather conditions, etc...), causing flight regulations and delays. In search for predictive models that might support enhanced pre-tactical planning (i.e., that are able to forecast deviations), researchers need to understand how configuration choices are made by airspace managers. They would also like to find a way to predict which configuration will be used at each time moment during the day of operation, considering uncertainty caused by operational factors in search for a more accurate sector configuration schedule in the day before operation (or earlier), allowing better management of demand-capacity imbalances. However, it is generally unclear what features should be used for building a predictive model.

The third case study addressed these issues by utilizing datAcron's offline VA suite to gain understanding of the configuration system, patterns of change, and probable reasons for preferring one configuration over another. During the study, interactive visual exploration of configurations used in several regions have been performed.

A preliminary version of the workflow and involved computational and visualization methods, have been reported in D4.6 Section 4.3. The evaluated version of the workflow utilizes incremental

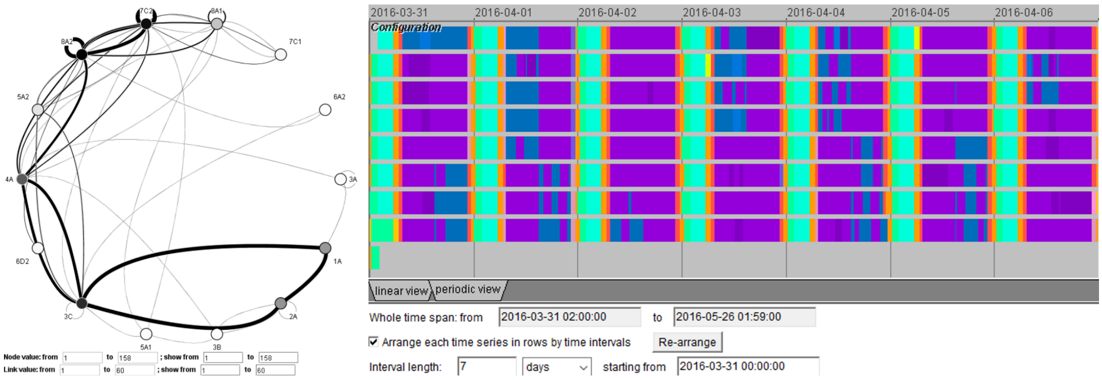

Figure 3: Left: A state transition graph shows changes of airspace configurations in one region during a month. Right: The configurations are represented by differently colored bar segments in a periodic time view.

improvements to interactive data filtering tools, data transformation tools, and visualizations of configuration attributes and patterns in space, time, and state spaces (Fig. 3). The latter are comprehensively described in paper [1].

During the case study, the proposed workflow supported by the VA suite allowed the involved domain expert to confirm regular patterns of airspace configurations and changes between them, as well as hypothesize about a number of outlier configurations that would not be expected for the given time-of-day, day-of-week, and expected traffic volumes [1].

It can therefore be stated that all evaluation criteria: functional completeness, flexibility, usability, and responsiveness have also been fulfilled for the settings of the third case study.

2.4 Summary

Complementing the evaluation metrics for aviation scenarios FM01, FM02, FP07, and FP10 as reported in D6.6 section 3.2, three additional case studies were used to evaluate the offline VA suite's capabilities as a flexible and extensible toolbox. Its potential has been illustrated in that it allows the formulation and execution of user-defined visual analytics workflows using a variety of both algorithmic and visualization tools and applied to diverse operating environment and different data sources. The results have been discussed and validated with domain experts to ensure applicability to operational needs, in particular in terms of predictability. It has demonstrated the value of the integrated technologies (algorithmic & visualization) to identify decision criteria as key aspects of the system, able to feed predictive or analytic models which are then themselves applicable during ATFCM planning. It can be particularly highlighted the power of these techniques to derive results from spatio-temporal patterns. In addition, the case studies also reaffirmed the suite's capability in terms of assessment of data quality from real-life data sources, such as DDR and CFMU, as an indispensable prerequisite to any analysis.

3 EYE TRACKING EXPERIMENTS – THEORETIC FOUNDATIONS

The visual analysis of eye movement data has become an emerging field of research leading to many new visualization techniques in recent years. These techniques provide insight beyond what is facilitated by traditional attention maps and gaze plots, providing important means to support statistical analysis and hypothesis building. There is no single "all-in-one" visualization to solve all possible analysis tasks. In fact, the appropriate choice of a visualization technique depends on the type of data and analysis task.

Experience and corresponding results from datAcron contributed towards a comprehensive taxonomy of analysis tasks that has been derived from literature research of visualization techniques [20]. In addition to the taxonomy, [20] proposes a pipeline model of eye-tracking visualization. This pipeline model is briefly described in section 3.2 as it helps to delineate foraging of eye tracking data and its task-specific visualization – the focus of this deliverable) – from the domain-level reasoning about the obtained results, as outlined in the respective deliverable D5.6 and D6.6 for the maritime and aviation domains.

3.1 Related Work

The application of eye-tracking technology as a means of evaluating human behavior has been established in many different research fields [13]. Due to the interdisciplinary constellation of researchers, the specific analysis tasks may also differ between the fields. While one researcher might be interested in the physiological measures (e.g., eye movement speed [18]), another wants to know in what order specific areas of interest on a visual stimulus were investigated [8]. Despite the differences between the research fields, it is possible to derive a high-level task categorization from a data perspective. Since the structure of the recorded data is usually identical in all eye-tracking experiments, we can categorize the analysis tasks according to three main data dimensions and three elementary analysis operations.

Depending on the research question, a statistical analysis of established eye tracking metrics [17] can be sufficient. However, the more complex the analysis task becomes, the more visual aid is usually required to interpret the data. Regarding the increasing amount of eye-tracking data recorded during experiments [4], it is reasonable to incorporate visual analytics techniques that combine automatic data processing with interactive visualization [2] into the analysis process.

As a starting point, the analysis of eye-tracking data is usually supported by some basic visualization techniques. For statistical measures, the application of statistical plots depicting the changes of a variable over time can already be helpful to interpret the data. In these cases, the visual stimulus is neglected. If the visual stimulus is important for the analysis, additional visualization techniques are usually included in the software suites of the major eye-tracking vendors such as the Tobii Pro X2[1] system that has been available in datAcron.

For many years, gaze plots and attention maps (e.g., Fig. 10ff.) were (and still are) the most popular visualizations that include information about the underlying visual stimulus. However, not all analysis tasks are facilitated by these techniques. For example, it is hard to interpret

[1] https://www.tobiipro.com/product-listing/tobii-pro-x2-30/

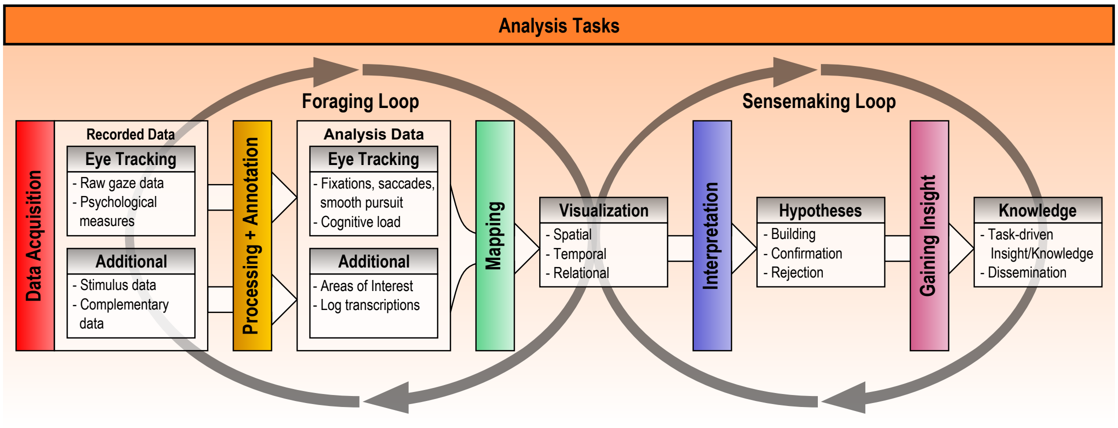

Figure 4: Extended visualization pipeline for eye-tracking data adopted from [20]: The recorded data passes multiple transformation steps before knowledge can be extracted. Each step from data acquisition, processing, mapping, interpretation, to gaining insight is influenced by the analysis task.

changes over time by simply replaying the animation [25]. Therefore, many new techniques have been developed over the last years to address this and many other analysis tasks, summarized by Blascheck et al. [6]. Additionally, as a beneficial but also challenging aspect, apart from the pure eye movement data a wealth of additional data sources can be integrated into an experiment [4]. Such a collection of heterogeneous data sources often impairs a combined analysis by statistical means and makes a visual approach indispensable.

Stemming from these considerations, we define typical analysis tasks when visualization techniques for eye movement data come into play [20]. The proposed high-level categorization is based on data dimensions directly focusing on recorded eye movement data but also on basic analysis operations. As a second goal, [20] discusses for each task category to which degree statistical and visual analysis can be applied to perform the given task, and present the suitable techniques. The provided list of examined visualization techniques in [20] is based on the collection provided in the quite comprehensive state-of-the-art report by Blascheck et al. [6].

3.2 The Eye-Tracking Visualization Pipeline

In [20] we formulate the way from conducting an eye-tracking experiment to gaining insight in the form of a pipeline (Fig. 4) that is an extended version of the generic visualization pipeline[11, 16]. The acquired data consisting of eye movement data and complementary data sources is processed and optionally annotated before a visual mapping, creating the visualization, is performed. By interacting with the data and the visualization, two loop processes are started: a foraging loop to explore the data; and a sensemaking loop to interpret it [22], to confirm, reject, or build new hypotheses from where knowledge can be derived. Since the analysis task plays an important role in all steps of the pipeline, we first discuss the underlying data and how it is processed before we introduce our categorization of analysis tasks.

3.2.1 Data Acquisition

Eye movement data combines several data dimensions of spatio-temporal nature. We distinguish between dimensions directly stemming from the recording of eye movements (raw gaze, physiological measures) and additional data sources serving as complementary data that can help achieve more reliable analysis results when combined with eye movement data. Typically, the displayed stimuli are an additional data source that can usually be included in the analysis process, since they are the foundation of most experiments anyway. Additional data sources provide complementary data such as verbal feedback, electroencephalography (EEG) data, and key press protocols.

The analysis task, or more precisely, the research question, typically defines how the experiment is designed and which data will be recorded. Most scenarios predefine also the visual stimulus. Exceptions are, for example, "in-the-wild" experiments with mobile eye tracking where it becomes much more difficult to control the experiment parameters.

3.2.2 Processing and Annotation

From the time-varying sequence of raw gaze points, more data constructs can be derived in a processing step. We identified fixations, saccades, smooth pursuits, and scanpaths as the most important data constructs [17]. In this processing step, automatic data-mining algorithms can be applied to filter and aggregate the data. Clustering and classification are prominent processing steps: For example, raw gaze points can be clustered into fixations and labeled. As another example, the convex hull of a subset of gaze points can be extracted to automatically identify areas of interest (AOIs). In general, the annotation of AOIs plays an important role in this step.

From the visual content of a stimulus (e.g., a picture or a video), AOIs can be annotated, providing semantic interpretation of the stimulus. With this information, additional data such as transition sequences between AOIs can be derived. Therefore, analysts can either rely on automatic, data-driven approaches to detect AOIs, or define them manually. Basically, there are two approaches: either defining areas or objects by bounding regions on the stimulus and calculating hits with the gaze data, or labeling each fixation individually based on the investigated content. Especially for video sequences, this annotation is a time-consuming step that often takes more effort than the rest of the analysis process.

From the additional data sources, recorded protocols and log files can typically be derived. It should be noted that each additional data source requires a synchronization with the recorded eye movement data, which can be difficult considering different sampling rates and irregularly sampled data (e.g., think-aloud comments) [5]. The processed data can finally be used for the mapping to a visual representation.

The analysis task influences what filters are applied to the data and what AOIs are annotated. For explorative scenarios in the context of visual analytics, visualization and processing are tightly coupled in a foraging loop, where the analyst can identify relevant data artifacts through interaction with the visualization.

3.2.3 Mapping

The mapping step projects the analysis data to a visual representation. According to Blascheck et al. [6], the main categories of state-of-the-art visualization techniques for eye tracking are spatial, temporal, and relational data representations. Therefore, our task categorization follows a similar scheme and appropriate visualizations are selected according to the main data dimension that is required to perform the corresponding task. It may be noted that only a few visualization techniques for eye movement data also take into account the additional data sources for an enhanced visual design in order to explore the data. We think that this is actually noteworthy for future work since those data sources may build meaningful input for sophisticated data analyses if they are combined with the traditional eye movement data.

As mentioned before, the analysis task plays the most important role for the choice of the appropriate visualization technique. In the foraging as well as the sensemaking loop, the visualization has to convey the relevant information and should provide enough interaction supported by automatic processing to adjust the visualization to the specific needs of a certain analysis task.

3.2.4 Interpretation

For the interpretation of the visualization, we can distinguish between two strategies: Applying visualization to support statistical measures and performing an explorative search. In the first case, hypotheses are typically defined before the data is even recorded. Therefore, inferential statistics are calculated on appropriate eye-tracking metrics, providing p-values to either support or reject hypotheses. Here, visualization has the purpose to additionally support these calculations. In the second case, the explorative search, hypothesesmight be built during the exploration process.

Filtering and re-clustering data, adjusting the visual mapping and reinterpreting the visualization can lead to new insights that were not considered during the data acquisition. This explorative approach is particularly useful to analyze data from pilot studies. Building new hypotheses, the experiment design can be adjusted and appropriate metrics can be defined for hypothesis testing in the final experiment.

The interpretation of the data strongly depends on the visualization. With a single visualization, only a subset of possible analysis tasks can be covered. For an explorative search where many possible data dimensions might be interesting, a visual analytics system providing multiple different views on the data can be beneficial. It allows one to investigate the data in general before the analysis task is specified.

3.2.5 Gaining Insight

As a result of the analysis process, knowledge depending on the analysis task is extracted from the data. As discussed before, this knowledge could be insights that allow the researchers to refine a study design or conduct an entirely new experiment. In the cases where visualization has the main purpose to support statistical analysis, it often serves as dissemination of the findings in papers or presentations. In many eye-tracking studies, this is typically the case when inferential statistics are performed on eye-tracking metrics and attention maps are displayed to help the reader better understand the statistical results.

3.3 Categorization of Analysis Tasks

The visualization pipeline for eye-tracking data (Fig. 4) shows the steps in which analysis tasks play an important role. For the experienced eye-tracking researcher, the first two steps—data acquisition and processing—are usually routine in the evaluation procedure. In the context of our chapter, mapping is the most important step in which the analysis task has to be considered. When the analysis task is clear, the chosen visualization has to show the relevant information. In this section, we present a categorization of analysis tasks that aims at helping with choosing appropriate visualizations. We discuss the main properties of the involved data constructs, typical measures for these questions, and propose visualizations that fit the tasks.

To provide a systematic overview of typical analysis tasks, we first derive the three independent data dimensions in eye-tracking data:

- **Where?** For these tasks, space is the most relevant data dimension. Typical questions in eye-tracking experiments consider where a participant looked at.
- **When?** Tasks where time plays the most important role. A typical question for this dimension is: when was something investigated the first time?
- **Who?** Questions that investigate participants. Typical eye-tracking experiments involve multiple participants and it is important to know who shows a certain viewing behavior.

With these three independent dimensions, visualizations can be applied to display dependent data constructs (e.g., fixation durations). Since many visualization techniques may not be restricted to just one of these dimensions but may facilitate different combinations of them, we focus our subsections on techniques where the name-giving dimension can be considered as the main dimension for the visualization.

Additionally, we can derive general analytical operations that can be related to other taxonomies (e.g., the knowledge discovery in databases (KDD) process [14]):

- **Compare:** Questions that consider comparisons within one data dimension.
- **Relate:** Questions that consider the relations between data dimensions and data constructs.
- **Detect:** Questions about summarizations and deviations in the data.

This categorization is based on the survey by Blascheck et al. [6], the work of Andrienko et al. [2], and the work of Kurzhals et al. [19]. The cited publications provide a more in-depth overview of current state-of-the art visualization and visual analytics approaches for the analysis of eye-tracking data. In the following, we briefly recap those aspects relevant to the eye tracking experiments conducted in datAcron as described in Section 4. In particular, since all participants of the conducted experiments were drawn from rahter hommogenous groups of domain aspects, the who dimension is of little relevance. A more detailed discussion of all aspects can be found in [20].

3.3.1 Where? – Space-Based Tasks

Typical questions that consider the spatial component of the data are often concerned with the distribution of attention and saccade properties. Statistical measures such as standard deviations, nearest neighbor index, or the Kullback-Leibler divergence provide an aggregated value about the spatial dispersion of gaze or fixation points. If we define a saccade as a vector from one fixation to another, typical where questions can also be formulated for saccade directions. If AOIs are available, measures such as the average dwell time on each AOI can be calculated and represented by numbers or in a histogram.

If the stimulus content is important for the analysis, attention maps [7] and gaze plots are typically the first visualizations that come to mind. Attention maps scale well with the number of participants and recorded data points, but totally neglect the sequential order of points. With an appropriate color mapping and supportive statistical measures, an attention map can already be enough to answer many questions where participants looked at, if the investigated stimulus is static.

Space-based tasks for dynamic stimuli, such as videos and interactive user interfaces, require a visualization that takes the temporal dimension into account considering also the changes of the stimulus over time. If AOIs are available, we refer to the next section, because in this case, when and where are tightly coupled.

3.3.2 When? – Time-Based Tasks

Eye movement data has a spatio-temporal nature often demanding for a detailed analysis of changes in variables over time. Questions in this category typically have the focus on a certain event in the data (e.g., a fixation, smooth pursuit) and aim at answering when this event happened. Considering the detection of specific events over time, many automatic algorithms can be applied to identify these events. Automatic fixation filtering [24], for example, calculates when a fixation started and ended. For semantic interpretations, combining data dimensions to answer questions when was what investigated, the inclusion of AOIs is common. For statistical analysis, measures such as the time-to-first-hit in an AOI can be calculated.

Timeline visualizations are a good choice to answer questions related to this category. In general, timeline representations depict an additional data dimension or construct, allowing one to combine the data relevant for spatial analysis (e.g., gaze heatmaps) with its temporal progress.

3.3.3 Compare

Comparison in general can be seen as one of the elementary analysis operations performed during the evaluation of eye-tracking experiments. In fact, statistical inference is typically calculated by comparing distributions of a dependent variable. For example, fixation durations between different stimulus conditions can be compared with an ANOVA to find out whether a significant difference between the two distributions exists. However, inferential statistics can only provide the information that a difference exists. To interpret the difference between the conditions, or, if the low number of available participants limits applicability of statistics, a visual comparison is usually a good supplement to the statistical calculations.

Comparison tasks are typically supported by placing several of the visualized data instances next to each other in a side-by-side representation, sometimes denoted as small multiples visualization. Each data instance is visually encoded in the same visual metaphor to facilitate the comparison.

An example of such visual comparison can be found in a seminal eye-tracking experiment conducted by Yarbus [26], with participants investigating the painting "The unexpected visitor". To compare the viewing behavior for different tasks, the resulting eye movement patterns were depicted by rudimentary gaze plots, allowing an easy interpretation of how the task influenced the eye movements. This visualization strategy can be applied to many techniques, for example, to compare investigated stimulus content over time, different distributions of attention on AOIs [9, 12], and the comparison of participants [23].

3.3.4 Relate

In most analysis scenarios, not only a single dimension such as space, time, or participants is in the research focus. A combination of two, three, or even more dimensions and data constructs is included in the analysis to explore the data for correlations and relations between the data dimensions.

Investigating relations between AOIs across participants is an important aspect for analysis tasks in this category. Relations between AOIs are often investigated by transitions between them. They can show which AOIs have been looked at when and in what order. A standard statistical measure is the transition count. Given enough samples (participant sessions), transition matrices or Markov models can give valuable insight into search behavior of a participant [17].

3.4 Detect

Detecting patterns of common viewing behavior is important and often achieved by summarizations or aggregation of the data. Such summarizations can also be applied to find outliers in the data which might either result from a problem of the hardware or from unexpected and potentially interesting behavior of a participant.

Descriptive statistics are often applied to achieve this goal. Calculating the average fixation duration, the variance of saccade amplitudes, or the mean scan path length are some examples. Box plots are typically used to represent these values and additionally depict outliers as a simple-to-understand graph. However, more sophisticated visualization techniques can be utilized to summarize the eye movement data and detect outliers visually. Summaries can be created for the raw data points, for aggregated data using AOIs, or for the participants. One possibility is to depict one dimension of the fixation position plotted against time [15]. This allows investigating the general scanning tendency of a participant with regard to specific, relevant stimuli.

An AOI view facilitates a simple summarizations of eye movement data on the basis of AOIs, and may also be used to find deviations in the data. For example, an AOI may not have been looked at during the complete experiment by one or multiple participants. This may be an indicator that the AOI was not needed to perform the experiment task or participants missed important information to perform the task. AOI time lines can help answer this question (e.g., Fig. 11ff.).

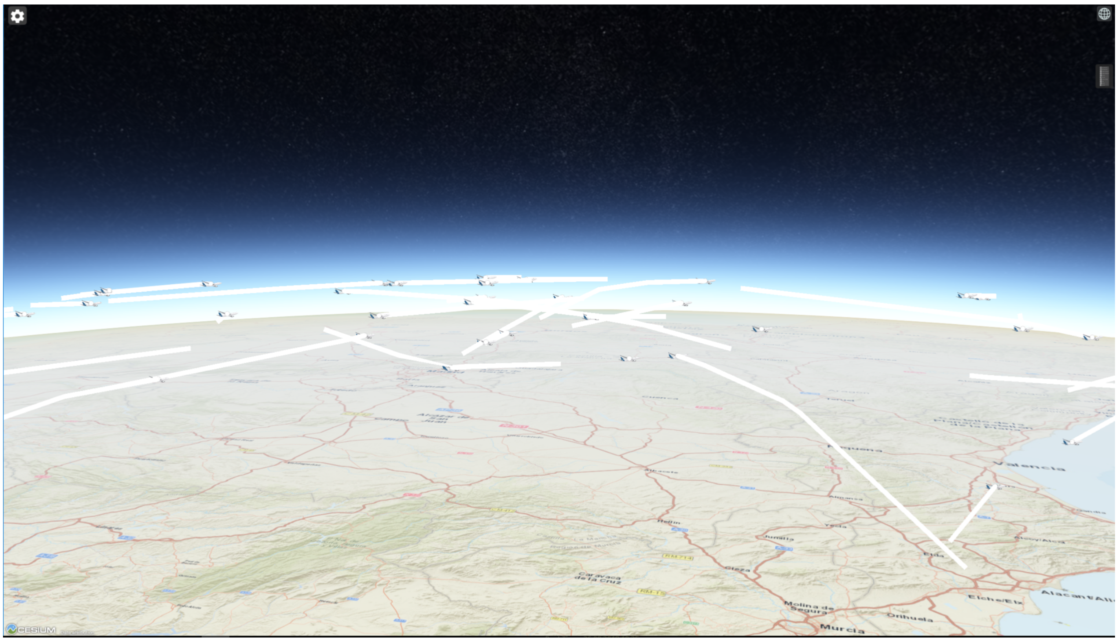

Figure 5: Real-time visualization of aviation data from FP scenarios, in this case, streaming IFS radar data. Note the optional 3D display mode to visualize this 4D (latitude, longitude, altitude, time) data stream.

4 EYE TRACKING EXPERIMENTS – REAL-TIME VISUALIZATION

4.1 Real-Time Visualization in Aviation

While the remainder of this Section does address exclusively the maritime domain and scenarios that are the subject of WP5, it should be noticed that in WP4 efforts have also been spent on supporting real-time visualization of aviation domain data (WP6). The latter relates to the flight prediction scenario FP09 specifically. However as noted in D6.6 section 3.1, FP09 was removed from the Validation plan and postponed for future research. The reason is this scenario is almost identical to FP07 except for using real-time information. The changes needed to manage streaming aviation data – which in all other FM and FP scenarios is collected and analyzed in batches ("offline") – both from architecture (WP1) and algorithms (WP, WP3) perspectives were deemed to be excessive.

As can be seen from Fig. 5, however, the real-time visualization component as of now is already capable of supporting the corresponding data streams so that future work could exploit the datAcron prototype. Details on aviation data visualization using WP4's real-time component are reported in D4.8 section 3.2.

4.2 Real-Time Visualization in the Maritime Domain

Eye tracking experiments have been carried out on real-time visualizations of maritime data using the a modified version of the datAcron integrated prototype. This prototype with its components has been described exhaustively in deliverable D1.12. The design of the overall evaluation scenarios and concrete experiment setups with the respective data sources, data variations, and domain-level events have already been described in datAcron deliverable D5.5. A detailed discussion of the scenario evaluation results, including qualitative assessments and statistical performance measurements, has been reported in D5.6. As such, to avoid redundancy the present document reports on specific information regarding the setup, conducting, and assessment of the eye tracking experiments to complement D1.12, D5.5. and D5.6; the most relevant information from these deliverables is summarized here briefly for context.

4.3 Experiment Settings and Goals

In order to narrow the scope for experiments to those most relevant to the maritime domain, the selection of one scenario amongst the use cases and scenarios defined in D5.3 has been done with the help of experts. Ten experts including military and civilians, from six different countries (Denmark, Finland, France, Germany, Norway, Romania) were requested to evaluate the three maritime use cases that include six scenarios requiring operational monitoring of fishing activities. Based on this feedback from these experts, it had been decided to focus experiments on collision avoidance scenario (SC1.1). The scenario is challenging, of interest and and highly relevant since it occurs very frequently. It is also well-suited to evaluate the overall datAcron architecture and prototype, as the scenario is also aligned with implemented MSIs that, together with the real-time vessel position data, constitute the joint output of datAcron's LED – SI – CEP/F – T/FLP pipeline (cf. D1.12). That is, different MSI relate to either low-level event detection (LED) or complex event recognition/forecast (CER/F) that can be made available as contextual semantic information to the user in the from of icon overlays, cf. Table 1.

In the collision avoidance scenario, the aim of the operator is to prevent and avoid a collision involving fishing vessels. The system could also enhance the situational awareness between vessels, anticipating that a vessel will be required to "give way" to a fishing vessel. In order to prevent a collision of fishing vessels with other ships, the user wants to predict which other vessels (especially larger ones such as cargo ships, tankers, ferries) will cross the trajectory or the areas where the fishing vessels are fishing.

Therefore, the high-level objective of the SC1.1-related experiments can be summarized as to evaluate if, and by how much, the semantic enrichment of the real-time maritime situational picture improves the expert users' capability – in terms of speed, accuracy, and confidence – to detect, verify, and resolve impending collisions in a timely fashion.

Specifically, the following MSIs detection functions (i.e., relating to either low-level event detection LED or complex event recognition/forecast CER/F) have been identified as relevant and have been made available to the users of the real-time visualization:

MSI#02 Vessel within a given area

MSI#03 Vessel on a maritime route

MSI#	Ico	Description
01		Close to a critical infrastructure
02		Within a given area
03		On a maritime route
04		Proximity of other vessels
05		In stationary area (ports or offshore platforms)
06		Null speed
07		Change of speed
08		Mismatch speed area
09		Mismatch speed type of vessel
10		Mismatch speed vessel history
11		Mismatch speed user defined value
12		Change of course
13		Mismatch course vessel destination
14		Mismatch course user defined value
15		No AIS emission/reception
16		AIS emission interrupted
17		Change in AIS static information
18		AIS error detection
19		Under way (using engine or sailing)
20		At anchor or moored
21		Movement ability affected
22		Aground
23		Engaged infishing
24		Tugging (tugged or tugging)
25		In SAR operation
26		Loitering
27		Dead in water, drifting
28		Rendez-vous

Table 1: List of MSIs with associated icons.

MSI#04 Proximity of other vessels

MSI#19 Vessel under way (using engine or sailing)

MSI#23 Vessel engaged in fishing

MSI#26 Loitering

MSI#27 Dead in water, drifting

The real-time visualization component of the prototype has consequently been designed with a strong focus on MSI visualization to support these evaluation objectives, as discussed in the following Section 4.4. For a detailed discussion of the mapping of the domain-level requirements to their corresponding MSIs and the possible data variations facilitating evaluation, refer to D5.5, sections 2 and 3. The eye tracking experiments were conducted aligned with three synthetic scenarios designed by CMRE and NARI, each of 30 minutes duration. For each of the three scenarios, genuine AIS data from the greater Brest area has been used, with target event-specific data added, either shifted in time, shifted in space, or shifted in time and space. Other data were synthesized so that a given event could be created.

4.3.1 Evaluation Scenario 1

For the first scenario, the following modifications were performed:

- Add a collision at minute 20
- Add a rendezvous at minute 24
- Add 4 vessels passing by

The full position data set with the AOIs of the target events are shown in the bottom right panes of Fig. 10 and Fig. 12, respectively.

4.3.2 Evaluation Scenario 2

For the second scenario, the following modifications were performed:

- Add a rendezvous at minute 15
- Add a near-collision at minute 25

The full position data set with the AOIs of the target events are shown in the bottom right panes of Fig. 14 and Fig. 16, respectively.

4.3.3 Evaluation Scenario 3

For the third scenario, the following modifications were performed:

- Add a near-collision at minute 23
- Add a collision at minute 26
- Add 2 vessels passing by

The full position data set with the AOIs of the target events are shown in the bottom right panes of Fig. 18 and Fig. 20, respectively.

4.4 Prototype Setup

The datAcron prototype setup for the maritime domain has been realized in two phases. The initial planing and preliminary setup in conjunction with first experiments (phase 1) was held at CMRE at the end of March 2018 (M27) with visiting personnel from NARI and FRHF. It involved three maritime experts and focused on a collision scenario ran on a partial datAcron prototype. The main goal was to test the then-current implementation in terms of functionality and the ability to access and display scenario-specific data. From this initial rounds of experiments two major conclusions were drawn:

- Given the initial prototype stage of the visualization software, overall usability as well as specific functionality for knowledge elicitation in MSA, such as lookup of vessel registry data, was know to be lacking. This was confirmed by the test users. However, given the amount of non-research related effort that would be required to duplicate functionality potentially available in domain-specific commercial software the conscious decision was made to concentrate in phase 2 on supporting the datAcron-specific requirements and objectives, namely, supporting truthful real-time visualization of MSIs in their spatial context provided by datAcron's LED – SI – CEP/F – T/FLP pipeline.
- The relevant part of the overall datAcron prototype that feeds into the real-time visualization component used for monitoring scenarios has been purpose-designed to handle streaming data accumulated in real-time as well. This is at odds with a multi-user evaluation experiment setup in which a scenario-specific data set needs to be re-played to several different participants to facilitate comparisons across users for the selected evaluation metrics. Therefore, the need to develop special ETL tools to better accommodate these very specific requirements within the datAcron architecture was identified.

The refined visualization prototype used in phase two has been further developed during M28–M35 in alignment with these findings. The overall revised experimentation is shown in Fig. 6. It had been conceived in M34 and has been applied in evaluation experiments involving three maritime experts, two cadets from the French navy, and the maritime expert which assessed MSIs along the project. The experiment took place at CMRE, with visiting personnel from NARI and FRHF, from the 5th to the 9th of November 2018 (M35).

4.4.1 Prototype Architecture

The architecture of the overall prototype is described in Fig. 6. It is organized in two parts. The first part concerns data preparation from the reference dataset provided by NARI (1) as starting point, injection of scenario specific patterns and events into the raw data (2), export of this synthetic raw events for integration (3) and processing by datAcron LED – SI – CEP/F – T/FLP components. The results are then stored in scenario-specific event log files (5). Finally, at the time of conducting a specific user session, the server-side visualization component (E) parses these files into a datAcron-compatible data stream using the aforementioned ETL tool (6).

During the experiment (i.e. at run time) a maritime expert in front of the screen executes and visualizes the scenario (F). To do so, the visualization client (A) – a standard web browser – queries the visualization backend (E) comprised of a Node.js/Angular web server (B) and a cascaded Web Mapping Service (WMS) server[2] (C) providing static context information in

[2] WMS is an open standard protocol developed by the Open Geospatial Consortium (OGC) for serving map image requests over HTTP, cf. http://www.opengeospatial.org/standards/wms.

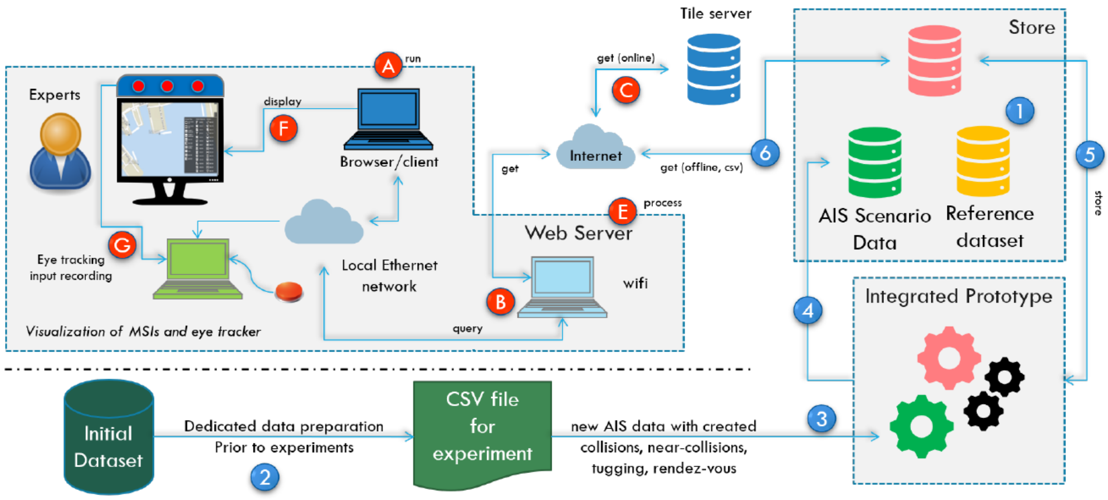

Figure 6: Prototype setup for the eye tracking experiments.

the form of raster map layers (map tiles). The expert's utilization of the visual stimulus (the real-time map visualization of the scenario) in terms of gazes and fixations is captured via the external eye tracker (G).

4.4.2 Visualization Functionality

Figures 7–9 provide a general overview of the user interface to the visualization for maritime situation monitoring. It is centered around the main map visualization that is updating in real-time as data streams are consumed by the visualization system from the datAcron back-end. The map content is comprised of a set of thematic layers that correspond to the different interrelated information produced by the datAcron components.

All other user interface dialogs (Figs. 8, 9) are hidden by default to maximize screen estate for the central visualization content (cf. Fig.7). The map is an interactive display that follows the established "slippy map" convention, that is, panning and zooming the map is achieved via fluid mouse interaction. The level of detail of map layers adjust automatically according to the current zoom level to show more detail when zoomed in where available.

At the top is a selection of available map layers providing geographic context information, such as traffic separation schemes or recommended routes, that may be used to evaluate vessel movement and assess situations (A). For the evaluation and eye tracking experiments, layers OpenSeaMap, fishing areas, traffic separation schemes (TSS), and Natura 2000 protected areas have been available.

Immediately below are toggle switches to enable or suppress the display of raw AIS data ("vessel trajectories") including the length of retained movement traces, compressed information ("datAcron synopses") provided by the datAcron semantic integrator SI (B), and semantic data enrichment provided by datAcron's CER/F and T/FLP components mapped to MSI icon overlays (D).

The vessel movement traces are displayed as polylines comprised of a configurable number most recent trajectory points (Fig. 9, right (B) – 200 by default). This allows the user to observe any vessel's prior movement, which is relevant to assessing unfolding situations, e.g. for impending (near) collisions appearing in the evaluation scenarios.

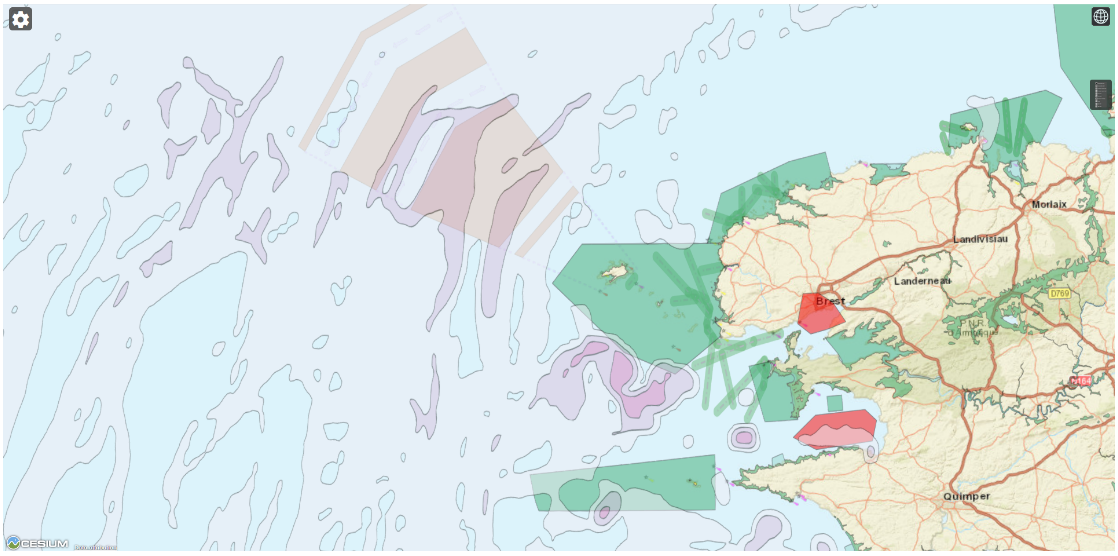

Figure 7: Real-time visualization map display showing all static context layers (WMS layers).

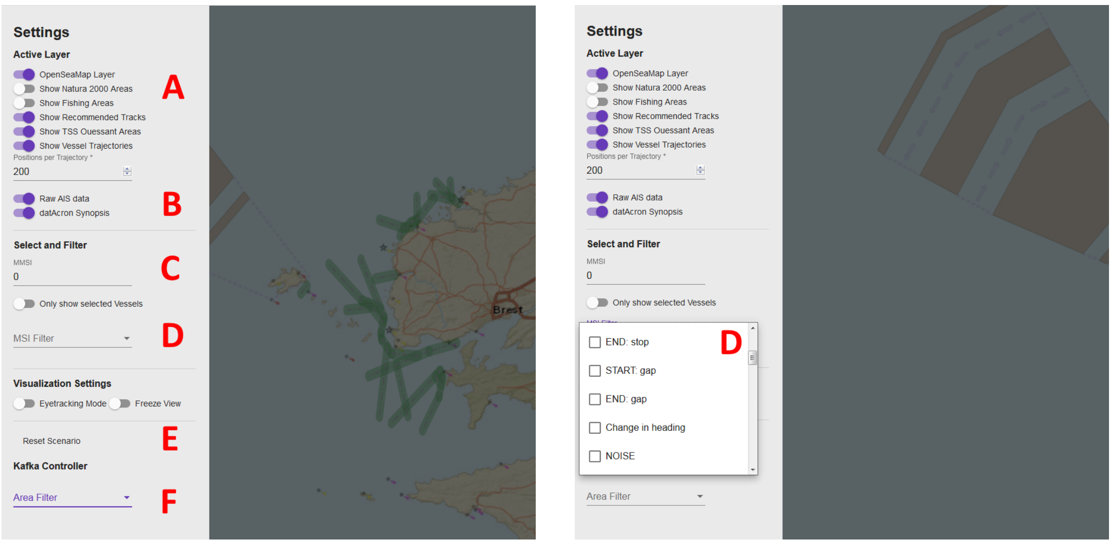

Figure 8: Settings interface of the real-time visualization component used to configure the information displayed on the map display. For the eye tracking experiments, pre-selected settings where used in accordance to the designed data variations (cf. D5.5, D5.6).

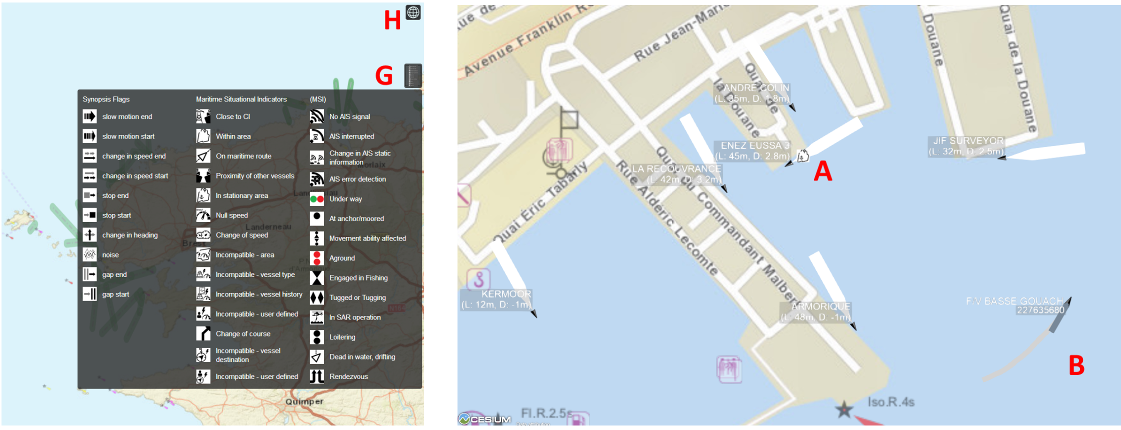

Figure 9: MSI icon legend (left, G) and close-up und vessel information overlays (right) including vessel size and haeding (white outlines), MSI icons (A), and vessel traces (B).

To declutter the map, the number of displayed vessels, i.e., their positions, traces, and MSI overlays can be limited to a selected subset. Vessels of interest can be marked iteratively on the map; however, a specific vessel can be selected directly via its known ID using a filter control (C).

Some complex events supported by the datAcron components can be parametrized by the user at run time. For example, MSI#02 "Vessel within given area" can be tuned to only report events of vessels inside one or several areas of interest via an area ID filter, e.g., the specific fishing areas the maritime surveillance expert is responsible for (F). However, the prepared scenario data already accounted for relevant regions so this option was not used.

For the eye tracking experiments, pre-selected settings were also used for the set of visible context layers, movement traces, and active MSI indicators in accordance to the designed data variations (see D5.5, D5.6). The experiment participants were instructed not modify these settings in order to ensure maximal comparability across different participant sessions.

As mentioned above the main purpose of the evaluation prototype has been to evaluate the MSI concept. This is reflected in the options for fine-grained control over which MSI are displayed or suppressed (Fig. 8 (D)). More importantly, the set of MSI icons is readily available for reference while using the software (Fig. 9). This primarily addresses the fact that there are no established standard icons sets for MSI, thus all icon designs must be considered preliminary proposals, so even the experts subjected to the eye tracking experiments would not have memorized the individual icons.

4.4.3 Eye tracking-specific setup

For conduction the eye tracking experiments, a Tobii Pro X2-60 mobile eye tracker was used. This device is designed to support high-resolution eye tracking (gaze tracking) and fixation-based analysis at 60 Hz temporal resolution. "Mobile" in this case refers to the fact that the actual capturing equipment is a small device fixed near the screen showing the visual stimuli to be analyzed. This device is usually attached to a standard PC (typically, a laptop) executing the eye tracking/device driver software for recording, see Fig. 6 (G). This eye tracking/device driver software suite provided by the hardware vendor. When recording an eye tracking session, the software will integrate the collected gaze position data over time with the corresponding visual stimuli. These stimuli can either be static images, image collections, or pre-rendered videos, or from a live screen capture from a running application. In case of the datAcron experiments, the latter option was used by linking the eye tracking software to the real-time visualization component running in a browser.

This, however, has ramifications on information about the recorded stimuli that are available to the eye tracking software. The gazes are captured as a time series of gazes in pixel coordinates. No additional information, such as which (semantic) image region a given pixel belongs to, are available.

However, as established in Section 3.2.2, AOIs are an essential tool for the subsequent application of automatic data-mining algorithms to filter and aggregate the raw gaze data. Specifically for the eye tracking experiments, the AOI around the target events are highly relevant to assess when, how often and how long the situation had been come to the attention of the expert. The evaluations shown in Figures 10–21 each refer to a target event-specific AOI.

Missing any information on visual content structure and meaning, AOIs can only be defined in the eye tracking software in absolute pixel coordinates. While fine for static images, defining AOI in video sequences is already extremely time-consuming since between-frame translations of AOI must be manually specified. Obviously, if the visual stimulus is then an interactive, real-time dynamic visualization, the task complexity of defining AOI multiplies by an order of magnitude. In case of the datAcron experiments, each participant may have continuously chosen a slightly different zoom level and or pan location on the map, thereby changing AOI pixel coordinates constantly. Thus to ensure comparability between the eye tracking sessions of the different participants within adequate efforts, it is necessary to impose some restrictions on the possible interactions with the visualization.

To accommodate for this, the visualization component used for the evaluation experiments was further modified in the time between the first (M27) and second (M35) round of experiments to include an “eye tracking mode” that optionally disables zooming and panning of the map display. This block of settings shown in Fig. 8 (E).

4.5 Execution of Eye Tracking Sessions

The assessment of the scenario level was divided into five experiments which are summarized in D5.6, section 7 (table 50). Experiment 0 represented a rehearsal for experiment 2 and 3. Experiment 1 assessed the perceived relevance of AIS and MSI information in the framework of a serious game. Experiment 2 assessed how the availability of MSIs impacts the expert user assessment of near distance situations with respect to the availability of only AIS messages. Experiment 3 allows the expert user to take full advantage over all implemented functionality of datAcron prototype, both on the interactive visual analytics component (albeit with the interaction restrictions mentioned above) and on the detection capabilities. Experiment 4 consists in an expert-driven in-depth analysis of all detections of datAcron prototype on the data set used for experiments 2 and 3. Experiment 5 compared the impact of displaying “perfect” MSI, i.e., indications derived from ground truth including hindsight, over actual but possibly imperfect MSI, i.e., those computed on the prepared scenario data (but without hindsight) by datAcron components and used in experiments 2 and 3.

The user domain-level task in experiments 2, 3, and 5 that used the real-time visualization component had been to observe the overall maritime picture and to detect, classify and confirm certain situations with similar preludes as quickly as possible: rendezvous maneuvers, near-collisions, and collisions. The detailed discussion of the evaluation criteria, their statistical analysis, and derived findings are reported in deliverable D5.6 and not repeated here. In particular, refer to D5.6 section 5 for a statistical evaluation on the level of MSI perception as well as the level of domain task performance (collision prevention). The present document is intended to give some complementary insights to the indicative findings obtained from eye tracking experiments that had been conducted aligned with these experiments. Specifically, the eye tracking results discussed in the following have been collected during M35 experiment #2 (D5.6 section 5.4) and experiments #3 & #5 combined (D5.6 section 5.5) conducted at CMRE. In addition to gaze data, additional data has been acquired in the form of "think aloud" voice transcripts and key interactions for marking the type of detected situation (cf. Section 3.2.1).

During these experiments, a total of three (3) maritime experts were available as participants for eye tracking sessions. Each participant sat through a total of three scenarios Although we have developed visual analytics methods and workflows for higher-level analysis based on aggregates, e.g., an analysis of search and observation strategies by as proposed in [2, 8], unfortunately this very low number of study participants (due to limited availability) does not warrant such analyses for lack of data points.

Another important point to note is that as described in Section 4.3, all three scenarios had a 30-minutes duration. This is in line with the time frames it takes for the injected situations/events to unfold in a realistic fashion. However, this would also have meant to sit each expert for a minimum of 90 minutes in front of the eye tracker as a participant once in each scenario, which was hardly feasible. Therefore, the ETL tool (cf. Section 4.4) implements the option to accelerate the replay of scenario data at a fixed increased rate. Given that the scenario data have been condensed to the vessels creating the target situations plus selected "confounding" vessel movements, compressed time was deemed an acceptable trade-off between scenario time and session time requirements.

For all recordings a replay acceleration factor of x3 was used. Indicated recording times in seconds therefore range from 0 seconds (scenario start) to $\tilde{6}00$ seconds or 10 minutes, corresponding to 30 minutes of scenario time. For the same reason, the cumulative gaze heatmaps capture one minute of session time but 3 minutes of scenario time leading up to the event.

4.6 Evaluation Results

However, even for only three participants (who have significantly different levels of experience) a basic analysis in space and time with respect to the key AOI (see Sections 3.2.2, 3.2.3) some interesting indicative findings can be derived. Additionally, interviews conducted with the participants well as analysis of the "think-aloud" recordings allowed to confirm certain known limitations of the current prototype, but also identified previously unknown issues that could be improved in future work.

4.6.1 Gaze Analysis

In all following images 10–21 the cumulative gaze heatmaps and temporal dynamics of AOI fixations, respectively, are shown for expert #1 at top left, for expert #2 at top right, and for expert #3 (the most experienced one) at bottom left. All participant were male.

All heatmap images show the cumulative sum of gazes at pixel resolution for the indicated one-minute time time periods mapped to a green-to-red continuous color scale. Thus, green regions of the image attracted a minimum number of gazes, while red regions collected the maximum cumulative sum of gazes. Areas not looked at will not show a color overlay at all.

All AOI fixation time plots show a spatio-temporal aggregation of this raw gaze data: first, consecutive gazes on the area of the corresponding AOI are transformed into fixation events, i.e., the starting time and duration for which the participant ha his gazes inside that region. Then, this series of fixation events is aggregated into a time series with second resolution that conveys an approximate degree of attention on the AOI. Times when the participant did not fixate the respective AOI thus show up as periods with zero fixation length.

The following set of figures present a side-by-side comparison of both the gaze heatmaps and the corresponding AOI fixations of the three participants (P1–P3) for all three scenarios (Sc1–Sc3) and the contained two critical events each. In all figures, the heatmap plots show the cumulative gaze density for one minute of eye tracking recording prior and leading up to the event, i.e., leading up to the respective critical event. By contrast, the fixation plots show fixations only for the AOI corresponding to that critical event, as indicated in the lower-right panel of the respective figure. In all of these plots, a red vertical bar marks the occurrence of the corresponding event.

As a general observation, while the attention distribution is quite similar for all participants in each scenario with some exceptions, the most experienced third participant in general has more efficient attention distribution patterns in almost all scenarios and situations.

Scenario #1, Collision Event. For the collision situation in scenario #1, the heatmaps in Fig. 10 show very similar distribution of attention. P2 appears slightly more distracted by a non-critical traffic situation inside the Ouessant TSS to the north.

However, the fixation time graphs in Fig. 11 highlight a difference in the evolution of attention: P1 and P2 devote almost their entire attention on the unfolding situation, frequently coming back to observe vessel movement within the event AOI proper. P3 also notices the potential for a collision occurring very early (first peak in the plot), but then only comes back after some time to check if the situation has resolved, which in the scenario it did not so his attention stays on the situation until the onset of the aftermath.

It should be noted here again, in a real situation the participants would have alerted the vessels by contacting them after establishing the risk for collision existed. Obviously, the participants had no agency to change the unfolding situation during the eye tracking sessions. As recorded in the post interview (see D5.6), all participants remarked that they would have initiated contact at some point before the collision, and observed the movement of one or both vessels after the event confirmed to them at least one vessel was adrift in the aftermath, which was also apparent from the gaze data (not shown here).

Scenario #1, Rendezvous Event. In this situation as shown in Fig. 12, P3 is again very efficient in his distribution of attention. By contrast, P1 and P2 are still somewhat distracted by the aftermath of the collision that occurred 4 minutes earlier (recall this corresponds to 1:20 minutes of recorded session time). P1 and P2 concentrate most of their remaining attention on the Ouessant TSS and the rendezvous, which could also be a potential collision; whereas P3 scans more broadly to also monitor other traffic.

The post-interview indicated that P3 was quite sure early on this was a purposeful approach and not a pending collision. This expectation is also reflected in the fixation time series of P3 vs. P1, P2 as shown in Fig. 13.

Scenario #2, Rendevous Event. As can be seen from Fig. 14, all participants focus almost exclusively on the two approaching vessels. P2 and P3 also scan other vessels inside the harbor.

However, the fixation time series in Fig.15 reveal that P1 and P2 do come back to the target AOI over the entire duration of the scenario, while P3 again apparently follows a different scanning strategy and does only come back once after the situation has resolved.

Scenario #2, Near-collision Event. Fig. 16 shows similar attention distribution for P1 and P3. Interestingly, P2 also fixates the previous rendezvous situation that occurred almost 10 minutes (or roughly 2.5 minutes session time) before, splitting his attention.

This is also very apparent fro the fixation time plots in Fig. 17 which show a long but fluctuating succession of fixations on the near-collision related AOI. This is again in stark contrast with the early recognition followed by two concentrated peaks of attention shortly before the actual rendezvous exhibited by P3.

Scenario #3, Near-collision Event. Again, Fig. 18 shows very similar attention distributions for all three participants. P2 devotes a little more attention to non-critical traffic further to the north.

The corresponding fixation time plots in Fig. 19 reveal a very interesting pattern: all participants now follow a strategy of periodically checking the potentially dangerous situation as it develops. However, while the attention of P2 and P3 on the AOI is apparent immediately before closest point of approach, P1 appears to not pay any attention for a significant time before, and some time after, the point of closest approach.

Scenario #3, Collision Event. Fig. 20 again reveals very similar attention distribution in space. All three participants P1–P3 distribute their attention between the target event and AOI and the near-collision event that occurred 3 minutes (i.e., one minute of session time) earlier. It should be noted this effect is likely in part due to the recording time frames of Fig, 18 and Fig 20 being adjacent, in combination with the 3x accelerated data replay.

However, despite this potential effect the temporal dimension of AOI attention dynamics depicted in Fig. 21 again reveal two different strategies. On the one hand there is again the periodic checking on the situation during its prelude until the situation can not be averted, as apparently followed by P1 and P3, versus constant revisiting of the area as followed by P2. All participants frequently monitored the aftermath of the collision.

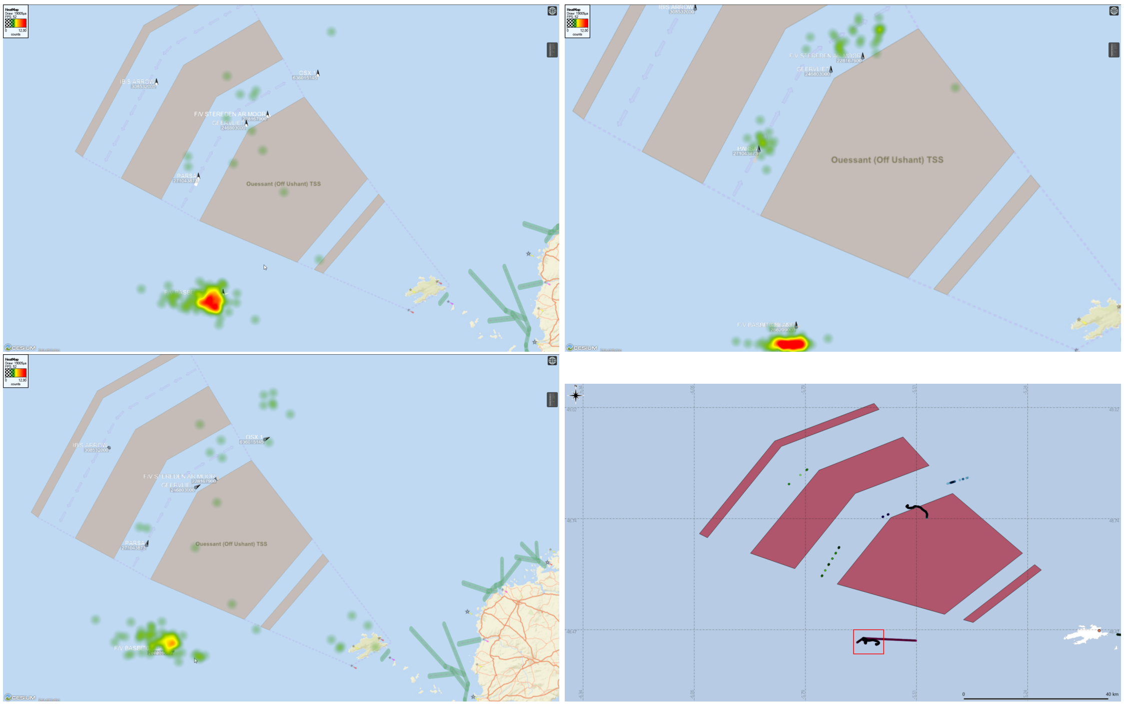

Figure 10: Gaze heatmaps of the three participants, first scenario, aggregated for one minute prior and up to the actual collision event (i.e., covering scenario minutes 18–20). These maps show the cumulative attention of the respective participant to specific regions of the visualization. The area of collision is shown in the bottom right map for reference (red rectangle).

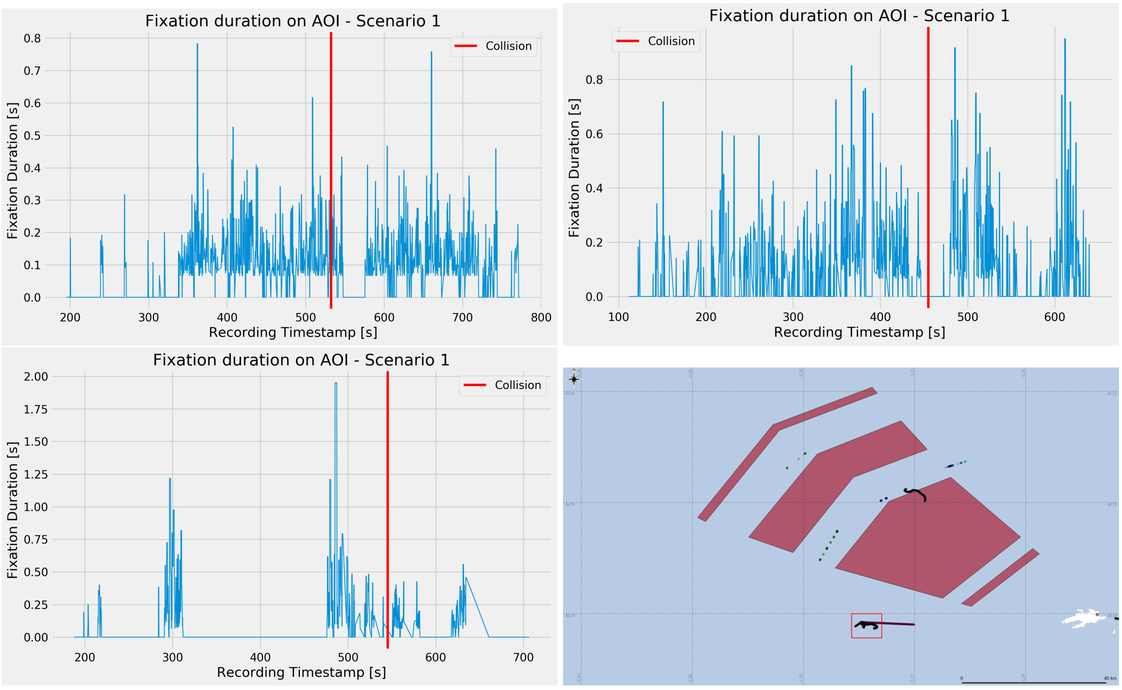

Figure 11: AOI fixations of the three participants during the entire first scenario, with the AOI on the area of collision (bottom right map, red rectangle). X axis shows the relative time in seconds since scenario start, Y axis the fixation duration.

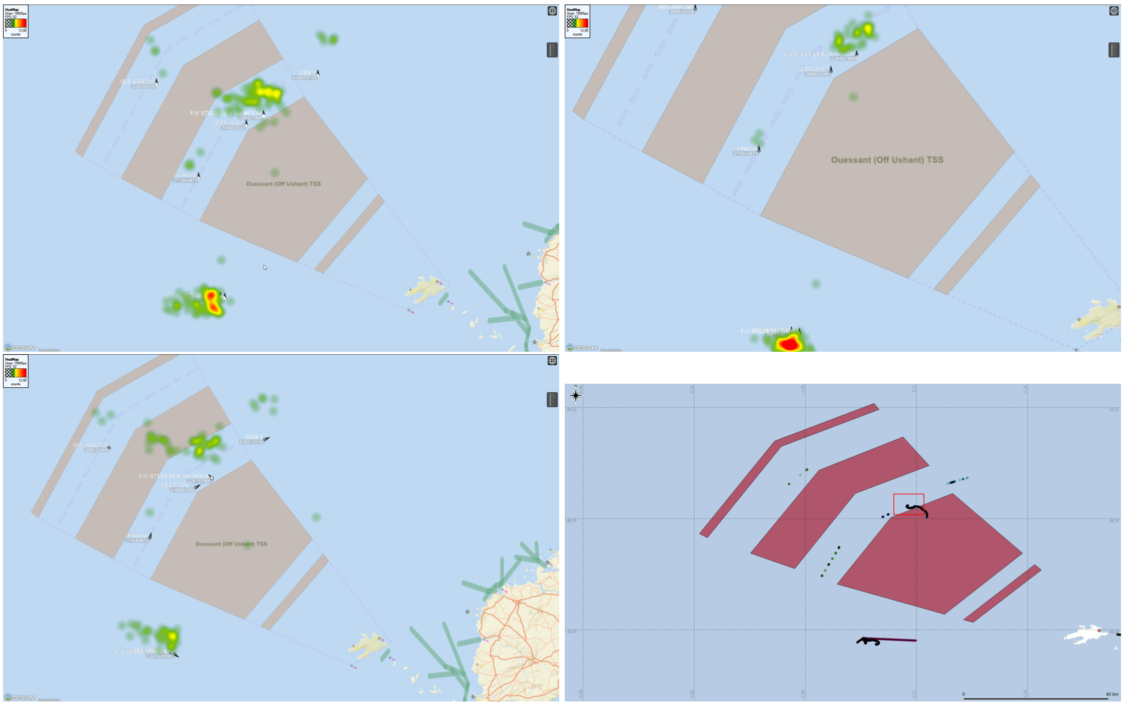

Figure 12: Gaze heatmaps of the three participants, first scenario, aggregated for one minute prior and up to the rendezvous event (i.e., covering scenario minutes 22–24), analog to Fig. 10. The area of rendezvous is shown in the bottom right map for reference (red rectangle).

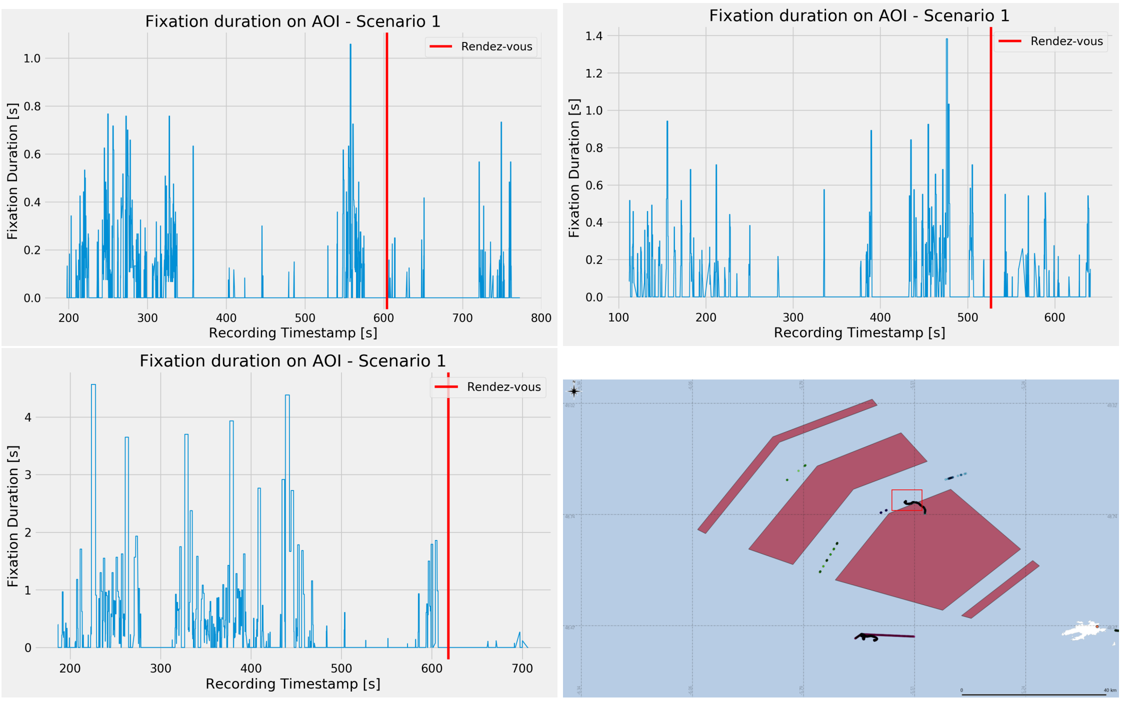

Figure 13: AOI fixations of the three participants during the entire first scenario, with the AOI on the area of rendezvous (bottom right map, red rectangle). X axis shows the relative time in seconds since scenario start, Y axis the fixation duration.

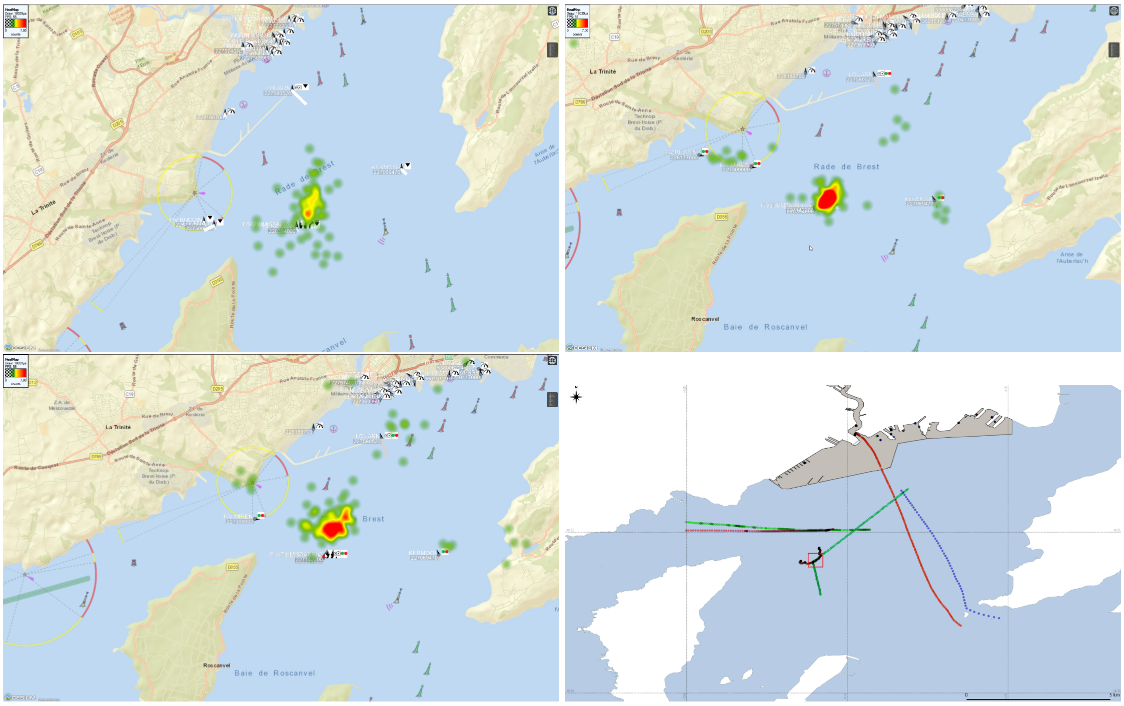

Figure 14: Gaze heatmaps of the three participants, second scenario, aggregated for one minute prior and up to the rendezvous event (i.e., covering scenario minutes 13–15), analog to Fig. 10. The area of rendezvous is shown in the bottom right map for reference (red rectangle).

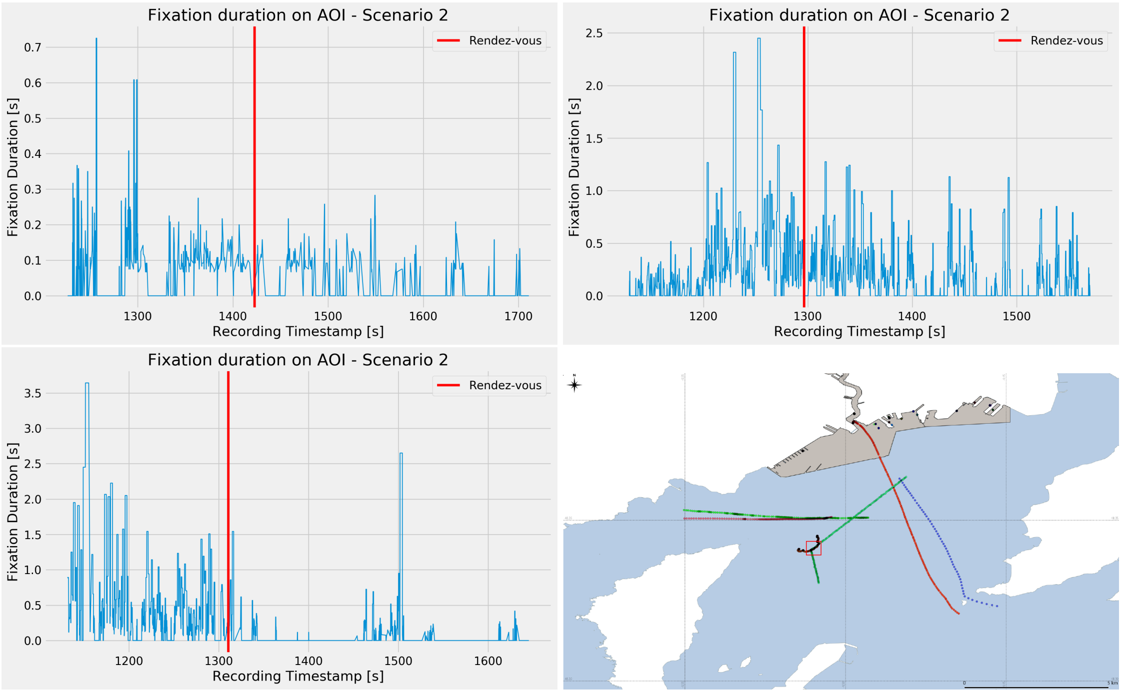

Figure 15: AOI fixations of the three participants during the entire second scenario, with the AOI on the area of rendezvous (bottom right map, red rectangle). X axis shows the relative time in seconds since scenario start, Y axis the fixation duration.

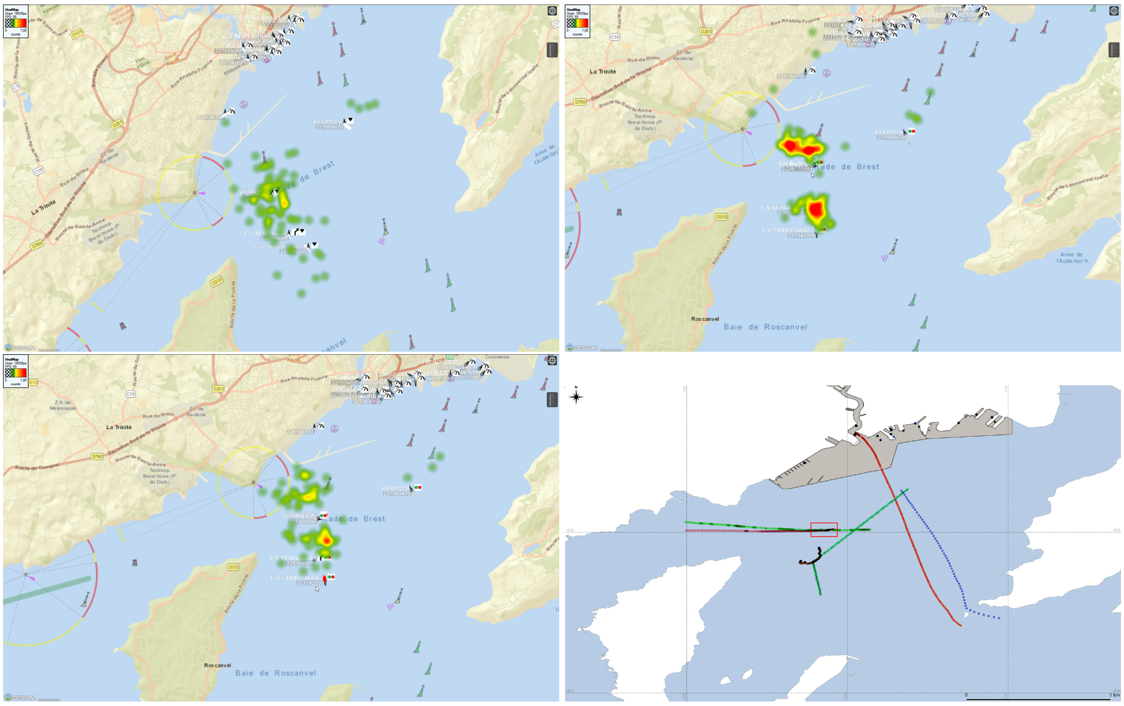

Figure 16: Gaze heatmaps of the three participants, second scenario, aggregated for one minute prior and up to the near-collision event (i.e., covering scenario minutes 23–25), analog to Fig. 10. The area of near-collision is shown in the bottom right map for reference (red rectangle).

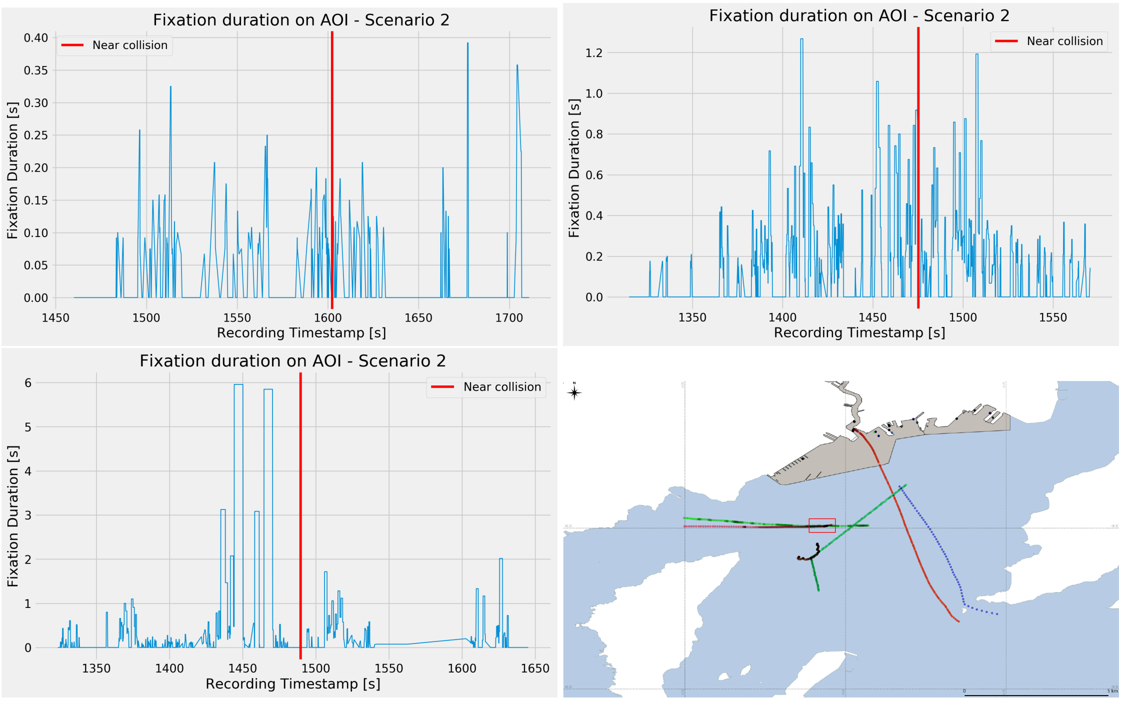

Figure 17: AOI fixations of the three participants during the entire second scenario, with the AOI on the area of near-collision (bottom right map, red rectangle). X axis shows the relative time in seconds since scenario start, Y axis the fixation duration.

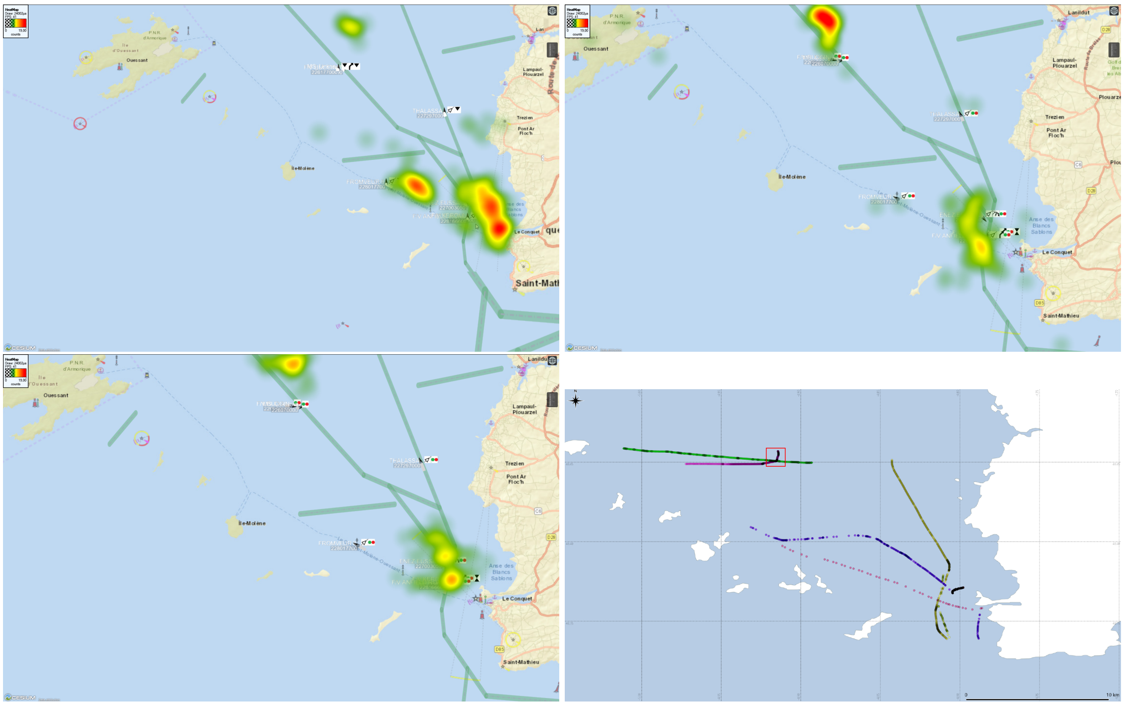

Figure 18: Gaze heatmaps of the three participants, third scenario, aggregated for one minute prior and up to the near-collision event (i.e., covering scenario minutes 21–23), analog to Fig. 10. The area of near-collision is shown in the bottom right map for reference (red rectangle).

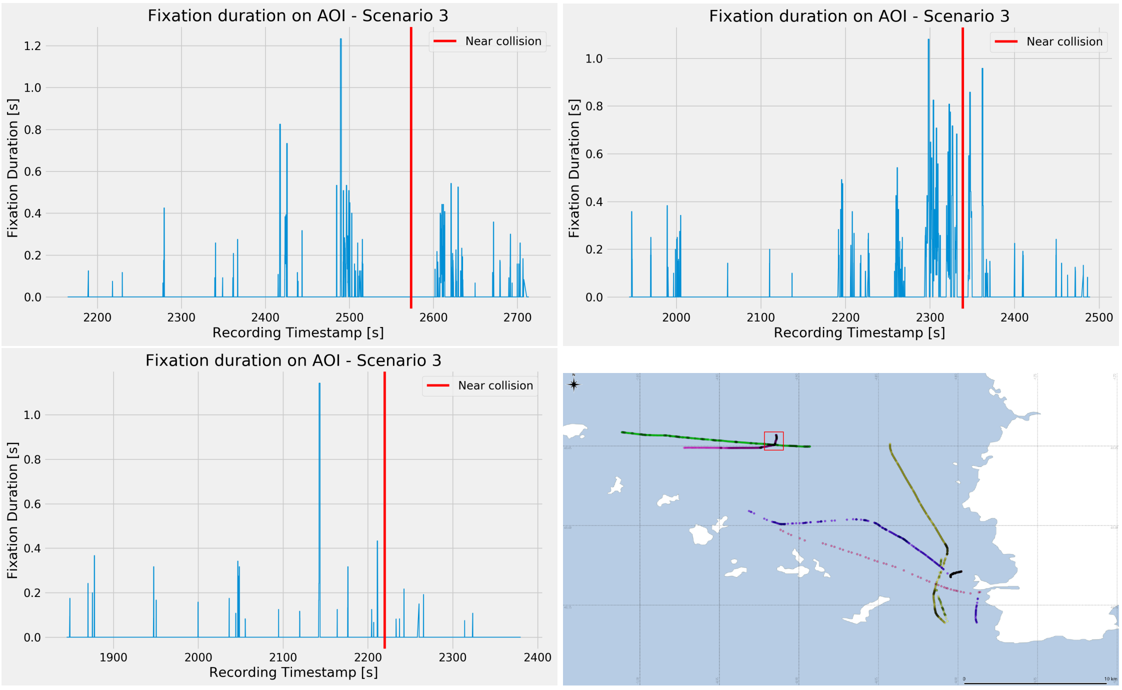

Figure 19: AOI fixations of the three participants during the entire third scenario, with the AOI on the area of near-collision (bottom right map, red rectangle). X axis shows the relative time in seconds since scenario start, Y axis the fixation duration.

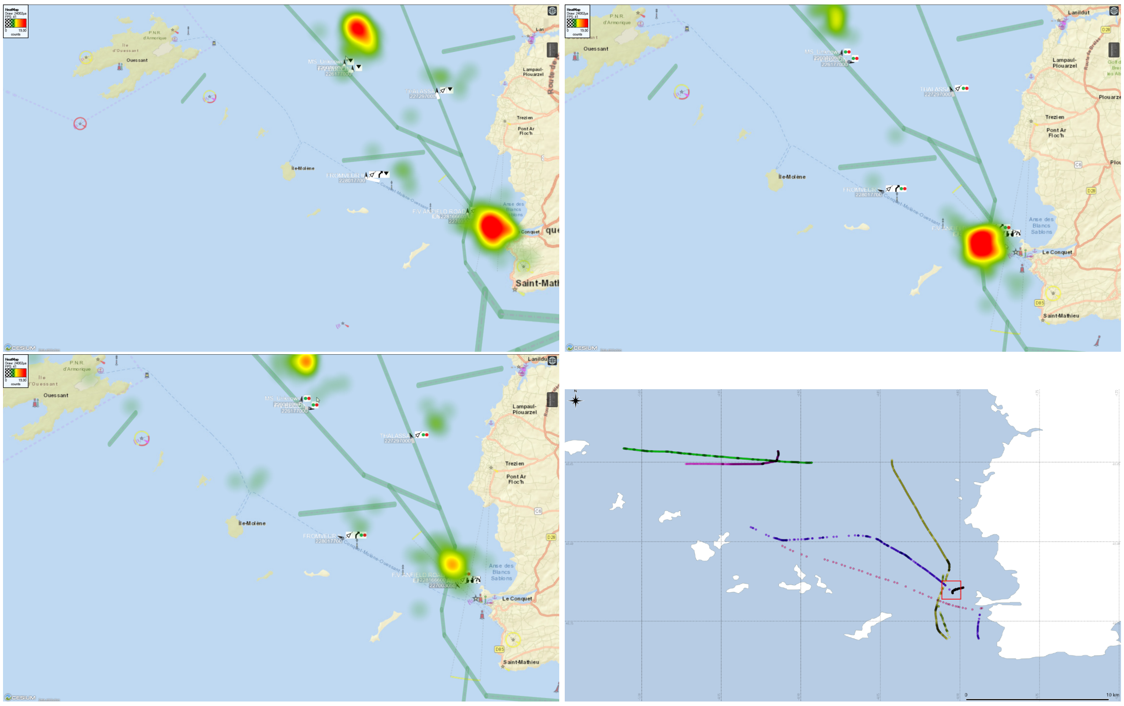

Figure 20: Gaze heatmaps of the three participants, third scenario, aggregated for one minute prior and up to the actual collision event (i.e., covering scenario minutes 24–26), analog to Fig. 10. The area of near-collision is shown in the bottom right map for reference (red rectangle).

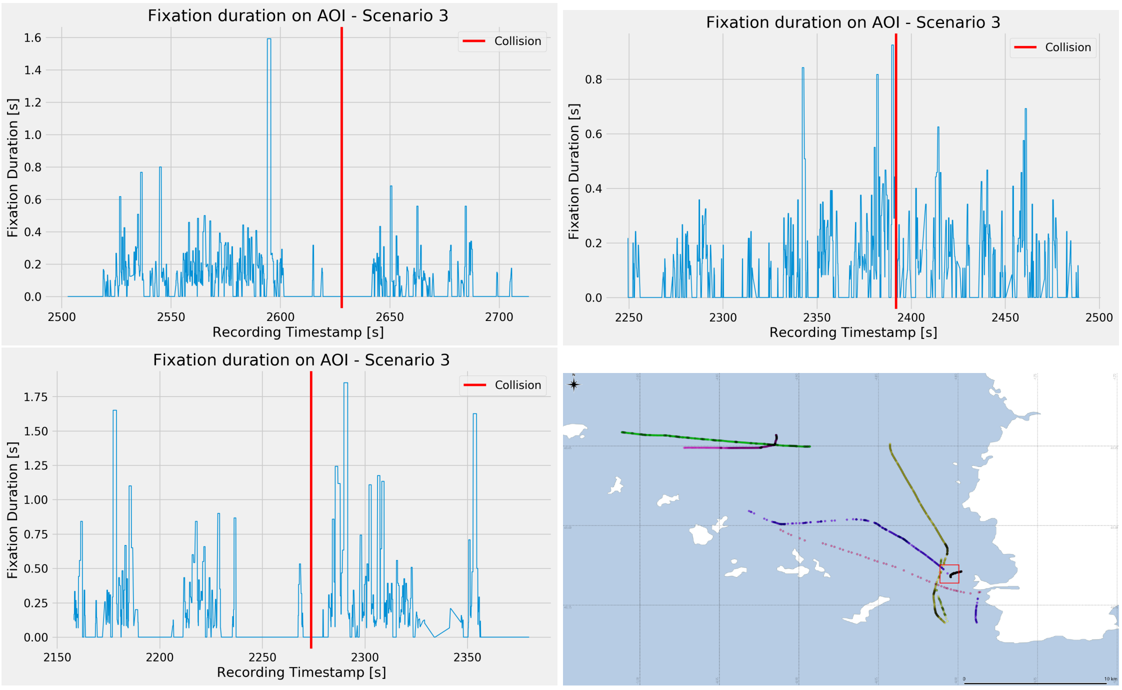

Figure 21: AOI fixations of the three participants during the entire first scenario, with the AOI on the area of collision (bottom right map, red rectangle). X axis shows the relative time in seconds since scenario start, Y axis the fixation duration.

4.6.2 MSI Visualization and System Usability

Some additional findings could be derived from the "think-aloud" recordings during the sessions as well as from the post-interviews with the participants. Reported here are issues brought up by the participants that relate to the specific issue of MSI detection and their visual icon-based representation in real-time displays, as well as on overall usability.

As a result of the focused development process geared towards the support of MSI evaluation pointed out in Section 4.4 and limited resources, the current software is missing some functionality that would be expected from a mature product. Therefore, no attempt to evaluate the usability, in the classical sense, of the overall software in its prototype state has been made.

That said, from the scenario design phase the involved partners were aware of specific functionality that should be prioritized during any future development. These expectations were largely confirmed by the expert users, as listed below. The points raised can mostly be seen as "quality of life" improvements. Specifically, they do not invalidate the principle approach of an MSI-enriched real-time map display. Confirmatory comments were received on the following possible improvements:

Speed and Distance Information. The indication of speed, together with a proper map scale, are imminently important (not only) for collision prevention tasks. All participants confirmed they would expect easy-to-use, in-display measuring tools for taking distances (and ideally, approach rates) between vessels similar to corresponding tools available in GIS spatial analysis toolboxes. For the indication of speed, experts confirmed the use of an explicit visual encoding, despite its additional screen space requirement, in a similar fashion to the established "lead line" standard (Fig. 22, left).

Contextual Information. The ability to look up contextual data on vessels of interest, i.e., those involved in a potential collision, has been pointed out by all participants. While some static vessel information is available in the AIS data stream, other facets requires interfacing the visualization component with a (potentially global) database with curated vessel information, which is why it was omitted from the prototype setup. Of note here is the fact that the experts confirmed different sets of information are most relevant depending on tasks. Specifically for collision prevention, information on length, draft, and speed that help to assess vessel maneuverability would be most relevant, whereas for the analysis of suspicious behavior information on vessel type, typical routes and last ports of call become more relevant. A possible avenue for future work is therefore the exploration of novel ways to directly encode the most relevant context information in a task-sensitive fashion on the primary display, which would go significantly beyond the interactive pop-up window technique that is the current state-of-the-art in commercial systems (cf. Fig. 22, right).

Configuration Options. Expanded configuration settings to optimize display information density from the beginning of sessions; specifically, availability of task- and user-specific presets for active icons/MSIs and context layers have been commented as being "nice to have" to "relatively important" for actual operative deployment.

Visual Representation. Improved icon placement. Currently, vessel name and ID are placed to the left of the vessel symbol, and icons are placed in variable-lengths rows to the top-right of the vessel symbol (for synopses flags), and to the bottom-right for MSIs (cf. Fig. 9, right). This will result in occlusion of other vessel icons, labels, and associated MSI indications when two or more vessels are in close proximity for a given map scale. To preserve the expressiveness of the visual representation, a placed label (i.e., text or icon) should neither occlude other labels nor visual representatives (e.g., ship and MSI icons, speed and trace

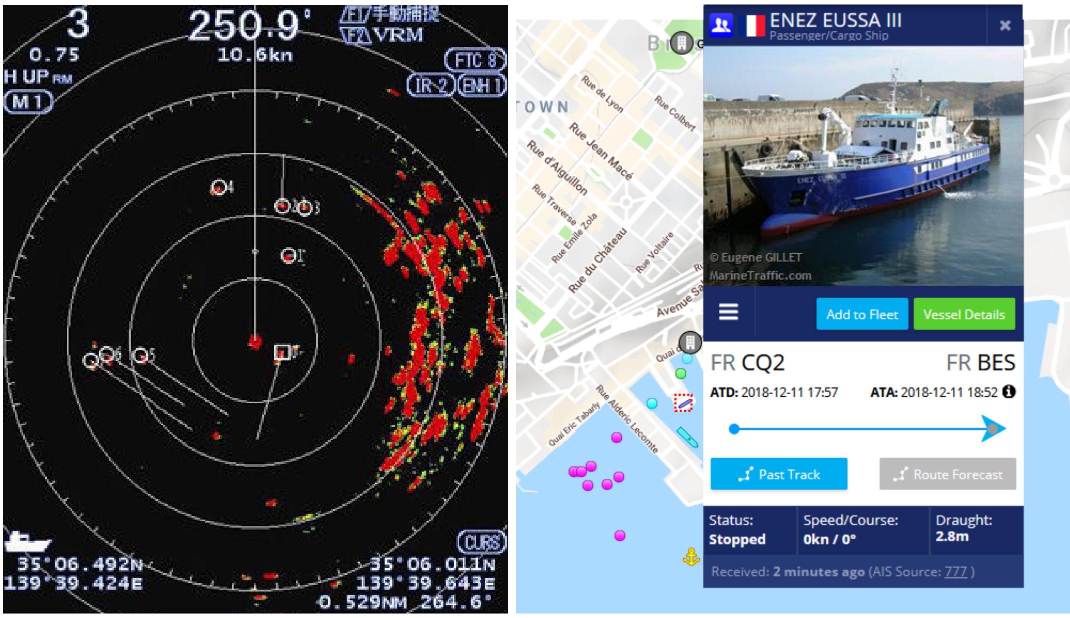

Figure 22: Left: Example of the established way of explicit visual encoding of vessel speed by "lead lines" in a MDC-2040 ship radar display (©Koden Electronics). Right: Example of a localized on-demand information pop-up for a selected vessel (dashed red rectangle) offering access to nested information dialogs such as "past track" (©Marine Traffic).

lines) that communicate crucial information. Although optimal, non-overlapping labeling is an NP-hard problem, there exist heuristic, real-time capable algorithms to this problem such as [21].

However, some interesting view points and observations were also made by the experts both while thinking aloud and during post interviews. These should be taken into considerations during future research and development of the MSI-based approach. Specifically, experts commented on:

Continuity of icons. A conscious prototype design decision has been to visualize MSIs as they are signaled or not by the underlying datAcron enrichment pipeline as accurately as possible. This results in some MSI being only shown extremely briefly (i.e., signaled as positional attribute for only one or a few positions), resulting in a potentially unstable display ("flickering icons"), or the user missing an indication shown only very briefly. Comments almost uniformly recommended to show any icon for a minimum amount of time before fading again. This raises the interesting question how to design the visualization to deviate from algorithmic CER output to accommodate for these opposed design goals. Specifically, retention times of icons corresponding to no longer signaled MSI need to be balanced between perceptibility vs display clutter; more challenging still is how to balance suppression of spurious indications – i.e., a single trajectory point annotated with a given MSI – with display delay, as without foresight, a visualization would need to suppress display of a given indicator until at least a number of successive trajectory points (above the threshold for spurious indication) have been received. This is a general consideration that might further need to be adaptable to different situations and operator needs. It is irrelevant whether this filter is implemented visualization-side or as part of the CER/F component.

False positive vs. false negative MSI. Related to the previous issue, expert comments offered divided opinions on the preferred way to handle the display of potentially incorrect MSIs, i.e., those where algorithm classification confidence is below a threshold. A False positive MSI in this context means an MSI that is displayed although the vessel's actual state does not warrant it; conversely, a false negative is an MSI *not* displayed although it should to accurately reflect the vessel's current state or situation. Two experts voiced a preference for minimizing false positives at the expense of increased false negative rates – i.e., rather too few MSI icons displayed if uncertain – on the grounds that MSIs as additional cues should only be provided if classification is almost certain. The third expert on the other hand opinionated in favor of false positives (and therefore, more MSI icons displayed overall), stating that "it should alert you to look closer". With regard to the previous aspect, it should be noted that retaining icons for a minimum time even after the MSI itself is no longer indicated in the data stream constitutes a false positive indication, whereas suppressing spurious MSI annotations creates a false negative situation, so both aspects should be jointly examined in future work.

Time dependence of MSI relevance. One experts further commented on the fact that specific MSI will have their relevance determined not only by the task at hand but that it will also be a function of time. For example, MSI # 19 "under way" is relevant and useful only at the very beginning when the task is to quickly assess and assimilate a complex situational picture comprised of many vessels. The utility of the indication then declines and actually becomes redundant as a real-time visualization by principle dual-encodes the same information simply by the fact that a ship under way is moving on the screen. And in our version of the visualization, depending on the setting, is also trailing its movement history trace so the under way MSI actually represent a triple encoding – the icon next to a moving vessel symbol trailing its trace line. This concept of time-variable icon relevance is an interesting aspect in future exploration of display deluttering strategies, see next point.

Clutter of current icon sets. In line with the remarks regarding MSI relevance and utility, participants also remarked on the degree of clutter the current icon set creates. This of course is to a large degree due to the sub-optimal placement currently (see above), and compounded by the fact that by default, all MSI icons are show. However, we also received comments to "only show the most important icons to remove clutter", with P3 going so far as to state that "with proper declutter/icon prioritization measures in place, vessel filtering [as it is implemented now] would be not that important" (additions in square brackets for context). This, then, implies that a context-dependent order of importance (i.e., based on task, situation, elapsed time, and potentially user) should be established in future research and the findings potentially encoded as display rules.

Icon Design (Standard). When designing the icons for maritime real-time visualization, the team was surprised to find that to the best of their knowledge, no established standard for MSA picture symbology, e.g., similar to NATO APP-6C, appears to exists, neither for civilian nor military applications. All participants correctly pointed out that the current, preliminary icon set is not optimal yet as it contains too many fine details that exacerbate the clutter problem (and overall readability) significantly. Moreover, operator task performance is expected depend heavily on MSI icon memorization by discarding the need for frequent legend look-ups. Therefore, a clear recommendation for future work in this direction is to seek out exchange with corresponding regulatory entities to align MSI icon design with relevant existing or developing standards.

5 CONCLUSION AND FUTURE WORK

In alignment with datAcron work package WP4, Task 4.5 “Evaluating VA methods in several scenarios and workflows”, this document reported on results and findings from the evaluation of the various visual analytics (VA) methods that have been evaluated in corresponding usage scenarios with appropriate categories of professional users. The underlying visual analytics methods and corresponding visualization tools have been developed in datAcron work package 4, tasks T4.1–T4.4 and reported in detail in deliverables D4.5–D4.8, respectively.

Offline VA methods and tools have primarily been relevant to the aviation use case scenarios and targeted data science specialists focused on batch (offline) data processing and analysis.

Complementing the evaluation metrics for aviation scenarios FM01, FM02, FP07, and FP10 as reported in D6.6 section 3.2, three additional case studies were used to evaluate the offline VA suite’s capabilities as a flexible and extensible toolbox. For this, qualitative evaluation criteria have been defined to assess the general ability to facilitate domain-specific analysis tasks, in terms of flexibility and functional completeness. The VA suite’s potential has been successfully illustrated along these criteria in that the offline VA suite allows the formulation and execution of user-defined visual analytics workflows using a variety of both algorithmic and visualization tools and applied to diverse operating environment and different data sources. The results have been discussed and validated with domain experts to ensure applicability to operational needs. The VA suite has demonstrated the value of the integrated technologies (algorithmic & visualization) to identify decision criteria as key aspects of the system, able to feed predictive or analytic models which are then themselves applicable during ATFCM planning. In addition, the case studies also reaffirmed the suite’s capability in terms of assessment of data quality from real-life data sources, such as DDR and CFMU, as an indispensable prerequisite to any analysis.

As noted in D6.6 section 3.1, scenario FP09 was removed from the validation plan. As explained in Section 4.1, WP4 contributions enable corresponding experiments already now so that future work could revisit this scenario.

That said, the real-time visualization component has so far been developed primarily in direct support of the MSI- and scenario-level evaluation in the maritime domain. Eye tracking experiments have been carried out successfully in conjunction with a number of evaluation scenarios as reported in D5.5 and D5.6.

While only three available participants each sitting through an equal number of relatively restricted scenario sessions unfortunately does not provide enough data for deeper analysis of visual search or scanning strategies (e.g., as applied in [8]). Likewise, the missing agency of the participants to influence the scenario as it unfolded meant the visualization was passively consumed rather that used as an analysis or interaction tool, thus missing elements of visually supported problem solving strategies that could be analyzed [2]. It is important to note that these limitations where a matter of practicality rather than oversight, as mentioned above.

Still, some interesting indicative findings can be derived from the collected gaze data. One definite conclusion is that in all scenarios and for all events, participants did divert non-trivial amounts of attention to the relevant AOI, both during the prelude as the situation developed and during the actual event. This confirms that all participants where able to successfully detect and assess the target situations in advance in all cases. Therefore, the performance scores reported in D5.6 sections 5.4 and 5.5 are indeed measuring the variance/effects of datAcron MSI for an enriched presentation of maritime situational pictures; missing of critical events by participants was not a factor.

As pointed out above, an interesting finding beyond the intended MSI- and scenario-level performance reported in D5.6 was the differences between the temporal dynamics of attention distribution between participating maritime experts of varying expertise. It stands to reason that the apparently more efficient division of attention exhibited by P3 (the most experienced participant) during most of the target events would also scale better in settings were visual/cognitive load becomes a factor, i.e., when the real-time visualization displays a larger number of contacts with many more potential situations (including false positives). This dimension has been determined to be out of scope for the primary objective of MSI evaluation and left out of the current set of experiments by design. However, exploration of this dimension should be interesting for future work.

As also mentioned a tradeoff had to be found between realistic time frames for maritime situations, which measure in several minutes, versus practical session lengths for the individual participants. The 3x data replay acceleration factor has been debated among the involved partners during the design phase of the scenarios. The eye tracking results indicate the compression of situational development did not negatively impact the participants' ability to perform their tasks for this particular set of tasks and scenarios; this aspect will need to be revisited for more complex and interactive evaluation scenarios in future work, however.

The real-time visualization component has been focus-designed to support the evaluation of the utility and effectiveness of MSI. As such, it does lack functionality that would be expected in an operative system, which has also been remarked by the participants. Nonetheless, besides the expected limitations a number of suggestions for future improvements could be gathered that can inform the design of MSA tools of the next generation, as discussed in Section 4.6.2.

6 REFERENCES

References

[1] Gennady Andrienko, Natalia Andrienko, Georg Fuchs, Stefan Rueping, Jose-Manuel Garcia, David Scarlatti, George Vouros, Ricardo Herranz, and Rodrigo Marcos. Visual analytics of flight trajectories for uncovering decision making strategies. In *Eighth SESAR Innovation Days*, Salzburg AT, 3rd – 7th December 2018.

[2] G.L. Andrienko, N.V. Andrienko, M. Burch, and D. Weiskopf. Visual analytics methodology for eye movement studies. *IEEE Trans. Visual. Comput. Graph.*, 18(12):2889–2898, 2012.

[3] N. Andrienko, G. Andrienko, J.-M. Garcia, and D. Scarlatti. Analysis of flight variability: a systematic approach. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):54–64, Jan 2019.

[4] T Blascheck, M Burch, M Raschke, and D Weiskopf. Challenges and perspectives in big eyemovement data visual analytics. In *Proceedings of the 1st IEEE International Symposium on Big Data Visual Analytics*. IEEE, Piscataway, 2015.

[5] T Blascheck, M John, K Kurzhals, S Koch, and T Ertl. Va^2: a visual analytics approach for evaluating visual analytics applications. *IEEE Trans. Visual. Comput. Graph*, 22(01):61–70, 2016.

[6] T. Blascheck, K. Kurzhals, M. Raschke, M. Burch, D. Weiskopf, and T. Ertl. State-of-theart of visualization for eye tracking data. In *EuroVis–STARs*,, pages 63–82. The Eurographics Association, 2014.

[7] A. Bojko. *Human- Computer Interaction. New Trends*, chapter Informative or misleading? Heatmaps deconstructed, pages 30–39. LNCS. Springer, Berlin, 2009.

[8] M Burch, G L Andrienko, N V Andrienko, M Höferlin, M Raschke, and D Weiskopf. Visual task solution strategies in tree diagrams. In *Proceedings of the IEEE Pacific Visualization Symposium*, pages 169–176. IEEE, Piscataway, 2013.

[9] M Burch, A Kull, and D Weiskopf. Aoi rivers for visualizing dynamic eye gaze frequencies. *Comput. Graph. Forum*, 32(3):281–290, 2013.

[10] Ed Huai-hsin Chi and John T. Riedl. An Operator Interaction Framework for Visualization Systems. In *Proceedings IEEE Symposium on Information Visualization 1998*, pages 63–70, 1998.

[11] E.H. Chi. A taxonomy of visualization techniques using the data state reference model. In *IEEE Symposium on Information Visualization*, pages 69–75. IEEE Computer Society, Los Alamitos, 2000.

[12] E.C. Crowe and N.H. Narayanan. Comparing interfaces based on what users watch and do. In *Proceedings of the Symposium on Eye Tracking Research & Applications*, pages 29–36, New York, 2000. ACM.

[13] A Duchowski. *Eye Tracking Methodology: Theory and Practice.* Science+Business Media. Springer, 2nd, 2007.

[14] U Fayyad, G Piatetsky-Shapiro, and P Smyth. The kdd process for extracting useful knowledge from volumes of data. *Commun. ACM*, 39(11):27–34, 1996.

[15] J.H. Goldberg and J.I. Helfman. Visual scanpath representation. In *Proceedings of the Symposium on Eye Tracking Research & Applications*, pages 203–210, New York, 2010. ACM.

[16] R. B. Haber and D. A. McNabb. Visualization idioms: A conceptual model for scientific Visualization Systems. In B. Shriver, G. M. Nielson, and L. Rosenblum, editors, *Visualization in Scientific Computing*, pages 74–93, Los Alamitos, 1990. IEEE Computer Society Press.

[17] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. an de Weijer. *Eye Tracking: A Comprehensive Guide to Methods and Measures.* Oxford University Press, Oxford, 2011.

[18] h. Kirchner and S.J. Thorpe. Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. *Vision Research*, 46(11):1762–1776, 2006.

[19] K. Kurzhals, B.D. Fisher, M. Burch, and D. Weiskopf. Evaluating visual analytics with eye tracking. In *Proceedings of the Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, pages 61–69, New York, 2014. ACM.

[20] Kuno Kurzhals, Michael Burch, Tanja Blascheck, Gennady Andrienko, Natalia Andrienko, and Daniel Weiskopf. A task-based view on the visual analysis of eye-tracking data. In M. Burch, L. Chuang, B. Fisher, A. Schmidt, and D. Weiskopf, editors, *Eye Tracking and Visualization. Foundations, Techniques, and Applications*, Mathematics and Visualization, pages 3–22. Springer, 2017.

[21] M. Luboschik, H. Schumann, and H. Cords. Particle-based labeling: Fast point-feature labeling without obscuring other visual features. *IEEE Transactions on Visualization & Computer Graphics*, 14:1237–1244, 10 2008.

[22] P Pirolli and S Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of the International Conference on Intelligence Analysis*, volume 5, pages 2–4, 2005.

[23] M Raschke, D Herr, T Blascheck, M Burch, M Schrauf, S Willmann, and T Ertl. A visual approach for scan path comparison. In *Proceedings of the Symposium on Eye Tracking Research & Applications*, pages 135–142, New York, 2014. ACM.

[24] D.D. Salvucci and J.H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the Symposium on Eye Tracking Research & Applications*, pages 71–78, New York, 2000. ACM.

[25] B. Tversky, J.B. Morrison, and M. Bétrancourt. Animation: can it facilitate? *Int. J. Hum. Comput.Stud.*, 57(4):247–262, 2002.

[26] A.L. Yarbus. *Eye Movements and Vision.* Plenum Press, New York, 1967.