Triangulating User Behavior Using Eye Movement, Interaction, and Think Aloud Data

Tanja Blascheck, Markus John, Steffen Koch, Leonard Bruder, Thomas Ertl*
University of Stuttgart, Germany

Abstract

In information visualization, evaluation plays a crucial role during the development of a new visualization technique. In recent years, eye tracking has become one means to analyze how users perceive and understand a new visualization system. Since most visualizations are highly interactive nowadays, a study should take interaction, in terms of user-input, into account as well. In addition, think aloud data gives insights into cognitive processes of participants using a visualization system. Typically, researchers evaluate these data sources separately. However, we think it is beneficial to correlate eye tracking, interaction, and think aloud data for deeper analyses. In this paper, we present challenges and possible solutions in triangulating user behavior using multiple evaluation data sources. We describe how the data is collected, synchronized, and analyzed using a string-based and a visualization-based approach founded on experiences from our current research. We suggest methods how to tackle these issues and discuss benefits and disadvantages. Thus, the contribution of our work is twofold. On the one hand, we present our approach and the experiences we gained during our research. On the other hand, we investigate additional methods that can be used to analyze this multi-source data.

Keywords: eye tracking, interaction, think aloud, visualization, data synchronization

Concepts: **•Human-centered computing → Empirical studies in visualization;** *Visualization techniques;*

1 Introduction

Evaluation is an important step when developing new visualization techniques. Typically, researchers evaluate new visualization systems calculating completion times or accuracy rates. This gives insights into how fast participants solved a task and if they solved a task correctly. However, if an analyst is more interested in how participants solved a task, why they take a certain path, or what kind of path they took [Brehmer and Munzner 2013], methods that are more sophisticated have to be used. For example, visual strategies of participants are evaluated using eye tracking data. Interaction logs give information about which interactions a participant used, and think aloud data gives insights into cognitive processes of participants. However, so far researchers evaluated all three data sources separately. To analyze what, how, and why participants conducted certain tasks a combination of all three data sources is preferable. Each data source answers one question: eye tracking data what a task pertains to, interaction data how a participant executed a task, and think aloud data why a participant conducted a task.

In this paper, we discuss how our approach, consisting of data collection, synchronization, and analysis, is applied to eye movement data, interaction logs, and think aloud transcripts. Blascheck et al. [2016b] presents the idea of combining different evaluation methods into a visualization approach and first experiences made with this combination. This work extends the previous one, by focusing on underlying ideas, techniques, and methods in greater detail. It emphasizes eye tracking as the data source with the highest temporal resolution and the essential source for understanding the motivation of subsequent interactions a participant carries out. The evaluation data used, has been collected during recently conducted experiments, where we analyzed interactive visualization systems with free exploration tasks. The challenges discussed in Section 5 reflect the lessons learned during these experiments. In addition, we introduce two visualization approaches for analyzing the data. Blascheck et al. [2016b] already presents a first visualization, the Area of Interest (AOI) sequence chart, and we now extend it to represent transparent, overlapping, and dynamic AOIs, as well as interactions influencing multiple AOIs. The second visualization technique is novel and integrates mouse fixations into the analysis, however, without displaying think aloud data explicitly. Furthermore, we apply the Needleman-Wunsch algorithm, in addition to the Levenshtein algorithm to compare strings with temporal binning.

2 Related Work

In visualization research different approaches for evaluating new visualization techniques are applied. Carpendale [2008] presents different methods, for both quantitative and qualitative analysis. She argues that different methods should be combined for a sufficient analysis of a visualization technique. Thus, we propose to combine eye tracking, interaction (mouse movement and interaction events), and think aloud data to evaluate new visualization systems.

Typically, eye tracking data is evaluated using statistical methods or visualizations. Statistics based on different eye tracking metrics allow a quantitative evaluation of the data. Holmqvist et al. [2011] present such different metrics for eye tracking data. A qualitative evaluation using visualization approaches permits an analyst to visually analyze and explore data to generate and confirm hypotheses. Blascheck et al. [2014] present an extensive collection of visualizations for eye tracking data. The most widely used methods for visualizing eye tracking data are attention maps and scanpaths. In our approach, we extend two visualization techniques based on AOIs: the AOI sequence chart [Holmqvist et al. 2011, p. 197] and the parallel scanpath visualization [Raschke et al. 2012]. In an AOI sequence chart, AOIs of a stimulus are represented as separate timelines for each individual participant. A rectangle represents each AOI hit by correlating the size of the rectangle to the dwell time in the corresponding AOI. The parallel scanpath visualization is similar to the AOI sequence chart. However, it represents AOI hits of different participants on top of each other. Lines represent AOI hits and transitions between AOIs, similar to a scanpath visualization.

*e-mail:firstname.lastname@vis.uni-stuttgart.de

ETRA 2016, March 14 - 17, 2016, Charleston, SC, USA
ISBN: 978-1-4503-4125-7/16/03
DOI: http://dx.doi.org/10.1145/2857491.2857523

Collecting and analyzing interaction data enabled an analyst to understand how a participant solved a task with a visualization system. Interaction data is also analyzed using different metrics or visualizations. For example, Pohl [2012] represents interaction logs as a temporal sequence where each interaction item is delineated with a colored rectangle to find patterns in the interaction data. The colors of interactions in the log are assigned to different categories. Munzner [2014, p. 58] defined such categories for interaction data in visualization research. We use her categories in our approach to assign all possible interactions to eleven categories. Reda et al. [2014] analyze interaction and think aloud data in combination using transition diagrams. However, including eye tracking data into these transition diagrams would be challenging as they use different states of the system instead of AOIs. Dou et al. [2009] and Lipford et al. [2010] represent interaction logs and think aloud data on parallel timelines to analyze reasoning processes. Both approaches do not include eye tracking data and analyze the interaction logs and think aloud data separately instead of in combination.

In visualization research, think aloud data is collected to analyze gained insights while performing a task [North 2006]. Think aloud data can be collected concurrently or retrospectively [Holmqvist et al. 2011, p. 101ff]. Both approaches have their benefits and drawbacks. Concurrently collecting think aloud data requires participants to verbalize their thoughts while using a system. However, this may influence eye tracking and interaction data. With retrospective think aloud, participants watch a video of their performance and then comment on it. However, in this scenario participants might already have forgotten why they followed a specific path especially if a task took a long time. In either case, think aloud data has to be coded for further analysis. In visualization research, different categorization schemes exist. Saraiya et al. [2005] categorize insights from think aloud data into overview, pattern, group, or detail. Smuc et al. [2009] extended the classification with additional insight categories for data and tool insights. Yi et al. [2008] extended this categorization into: provide overview, adjust, detect pattern, and match mental model. Reda et al. [2014] defined a different set of insight categories, namely observation, hypothesis, and goal. Tenbrink [2014] describes a detailed approach to analyze think aloud data called cognitive discourse analysis. She defines eight steps when using verbal protocols. The first three steps concern the scope of the study, data collection, and preparation of the data. Then, a content analysis is conducted and linguistic features are annotated. These annotations have to be checked for reliability before patterns are found in the data. The last step, Tenbrink proposes, is a triangulation with other data. Our approach follows this scheme, however, the different data sources are interleaved during the complete process.

3 Approach

Our concept aims at combining eye tracking data, interaction logs, and think aloud transcripts, as Blascheck et al. [2014] originally proposed. The pipeline depicted in Figure 1 shows this combination. We discuss each step of this pipeline in the following in more detail. First, the data collection is discussed in Section 4. In our case, we collect eye tracking, interaction, and think aloud data. Here, the challenge is to collect data in a format that can be synchronized afterward. In the second step the different data sources are combined, for example, they are matched based on timestamps and AOIs (cf. Section 5). In the last step, the synchronized data is analyzed (cf. Section 6) requiring new analysis metrics and new visualization approaches to gain insights. We propose a string-based analysis of the data (cf. Section 6.1) and two visualization-based approaches (cf. Section 6.2).

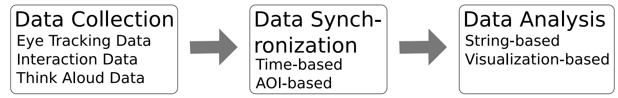
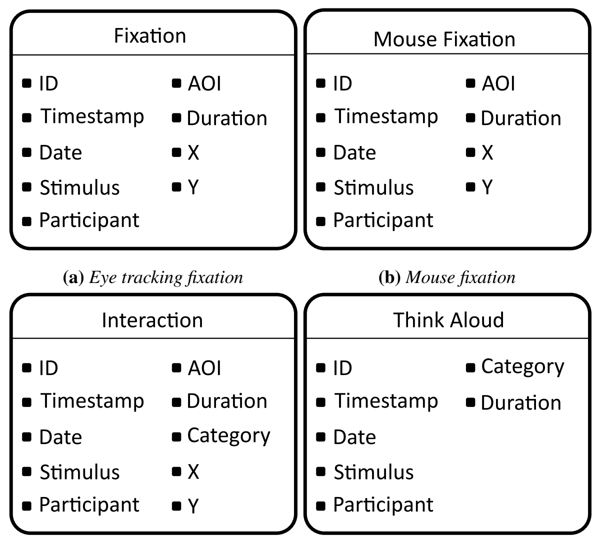


Figure 1: *Our concept is a three step model based on Blascheck et al. [2014]. First, eye tracking, interaction, and think aloud data is collected, then the data is synchronized based on timestamps and AOIs, and last it is analyzed using a string-based and visualization-based analysis.*

4 Data Collection

The first step in our approach is to collect data. Figure 2 shows the data types collected during a typical eye tracking study with a visualization system. Each data type has a unique identifier (ID), timestamp which corresponds to the elapsed time of the data recording, and a date representing the system time. The corresponding stimulus and participant is stored as well. The other attributes vary for each data type.



(a) *Eye tracking fixation* **(b)** *Mouse fixation* **(c)** *Interaction* **(d)** *Think Aloud*

Figure 2: *Data model for eye tracking fixations, mouse fixations, interactions, and think aloud data. Each data type has an ID, a timestamp, date, stimulus, and participant. The other attributes can vary for the different data types.*

4.1 Eye Tracking Data

We collect eye tracking data in our studies using a Tobii T60 XL (sampling rate 60 Hz) with a 24-inch screen and a resolution of 1920×1200 pixels. The data is processed with the fixation algorithm Tobii provides (velocity threshold = 35 pixels/sample, distance threshold = 35 pixels). The attributes for a fixation are listed in Figure 2a. As additional attributes the fixation's screen position, as well as its duration are stored. Fixations are afterwards assigned to AOIs to facilitate a synchronization with the other data types.

4.2 Interaction Data

In our case, we refer to interaction data meaning mouse movements as well as interactions. The analyzed visualization system has to collect mouse movements and interactions itself. We record mouse movements at a sampling rate of 60 Hz, as done for eye tracking data.

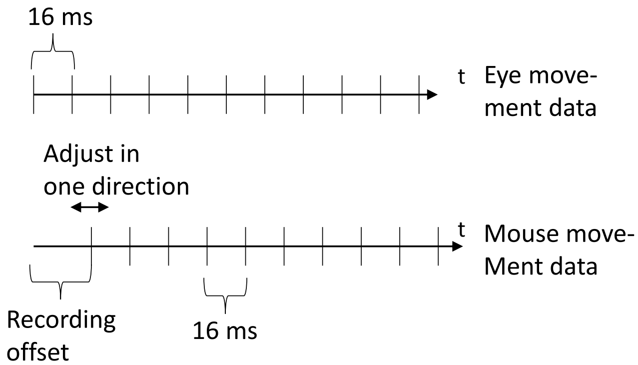


Figure 3: *If different systems record eye tracking and mouse movement data, the data can have a recording offset. Adjusting one data source using the system times of the data solves this challenge.*

Then, the mouse movements are aggregated into mouse fixations (cf. Figure 2b), if the mouse is not moving for a certain time. Thus, if consecutive mouse movements have the same screen position they are aggregated into a mouse fixation. The duration of all fixations is calculated afterwards they are assigned to AOIs.

Interactions are recorded discretely each time they take place. For each interaction (cf. Figure 2c), screen position and type (e.g., selecting, scrolling, zooming). Some interactions (i.e., scrolling) can also have a duration. Afterwards, interactions are assigned to specific categories based on their type. We use the eleven categories Munzner [2014, p. 58] defined for interactions in visualization research. This classification enables to aggregate similar interactions. Interactions are also assigned to AOIs.

4.3 Think Aloud Data

In our studies, we have so far collected think aloud data concurrently. We recorded the verbalizations using Tobii Studio with an external microphone. Afterwards, we hired two students to transcribe the data into written form. The students double-checked each other's work to make sure that the transcriptions were correct. These transcripts contain a duration (cf. Figure 2d). Think aloud data are assigned to categories and not to AOIs. Different categories have been suggested for think aloud analysis in visualization [Saraiya et al. 2005; Smuc et al. 2009; Reda et al. 2014; Yi et al. 2008]. Currently, we are developing a coding-tool which can be used to assign categories manually to think aloud data [Beck et al. 2015].

5 Data Synchronization

The second step in our pipeline is the synchronization of the data. There are different possibilities to synchronize eye tracking, interaction, and think aloud data. We chose a temporal and AOI-based synchronization. However, there are challenges to be considered.

5.1 Recording offset

If different systems are used to record eye tracking and mouse movement data, a challenge is that a recording offset may occur. Mouse movements should be recorded with the same recording rate as the eye tracking data (i.e., 60 Hz). However, if the recording of mouse data is started after the eye tracking data, the data can have a temporal offset (cf. Figure 3). Since we log the system times we can adjust the mouse movement data to overcome this problem. Presently, we are developing a common framework to collect different data sources synchronously.

5.2 Data Source Alignment

There are several options for aligning our three data sources. Theoretically, all three data sources can be used to align the other data sources to afterwards. However, each variant has different challenges. Taking the think aloud data as the main data source enables to analyze data based on insights in the data. Mapping the other two data sources to these insights might be problematic, as insights usually occur at specific points in time. These insights are not necessarily connected to eye movements and interactions directly before an insight occurred [Pike et al. 2009]. Thus, it is hard to tell which fixations and interactions belongs to which think aloud data point.

Interactions change the visual state of a system. Typically, these visual states are contained in the interaction log. Each could be generalized into higher-level states (i.e., layout-preserving or layout-changing [Reda et al. 2014]). These visual states could be used for synchronization purposes. However, different challenges arise with such a synchronization. First, defining visual states is difficult and depends on the system. Second, the status layout-preserving and layout-changing may be too coarse, and if too many conditions are defined an analysis might become challenging. Third, all participants should have the same visual states. However, if participants interact with a system each participant can have a different solution to a task, leading to many different conditions. Furthermore, matching eye tracking data is challenging for analysis purposes. A visual state still includes the complete system as the stimulus and eye fixations would be hard to map or AOIs would still have to be added.

Collecting mouse movements in a study facilitates to analyze correlations between mouse and eye movements. Both are recorded at the same level of detail. This level of detail is the most fine-grained of all data sources. Using either mouse or eye movements for synchronization would allow the most detailed temporal analysis. However, the question remains if mouse or eye movements should be used. In eye tracking research usually fixations are studied rather than saccades. When analyzing mouse data, mouse movements are more interesting than mouse fixations. Eye fixations are the most fine-grained data available having the highest temporal resolution. Additionally, they are an essential source for understanding the motivation of subsequent interactions participants carry out and typically are the cause for an interaction event. Thus, in the following we base the synchronization process on eye fixations and align the other data types to them.

5.3 Time-based Synchronization

We synchronize our data on a temporal basis. We consider two possibilities: linearly ordered time sequences [Aigner et al. 2011, p. 50f] and temporal binning [Cristino et al. 2010]. Linearly ordered time sequences ignore the duration of the data, whereas temporal binning considers the actual times. Depending on which approach is used, different synchronization results are possible. In our previous publication, we only considered linearly ordered sequences [Blascheck et al. 2016b]. However, if time periods are rather long it can be helpful to show the duration of an event, to make a comparison between participants easier. We think a longer viewing of a specific AOI indicates a different behavior than if a participant just quickly focused on an AOI.

Linearly Ordered Time Sequences. Figure 4 shows how the data sources are synchronized based on linearly ordered time and eye fixations. For each fixation, the start and end timestamp is considered. The start timestamp of mouse fixations, interactions, and think aloud data is used to match them to fixations. For example, mouse fixation M1, interaction I1, and think aloud TA1 would all be assigned to fixation F1. Interaction I2 would be assigned to fixation

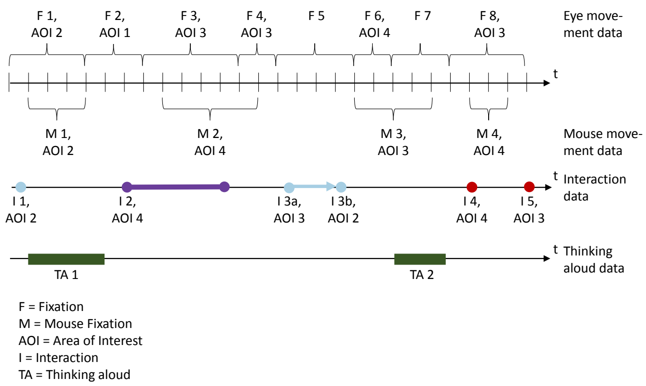


Figure 4: *Eye movements and mouse fixations last for a certain duration and are assigned to AOIs. Interactions are discrete (e.g., interaction I1) or continuous (e.g., interaction I2). Interactions can also influence multiple AOIs (e.g., interaction I3). Think aloud data can be represented as linearly ordered events based on start and end timestamps of a video recording.*

F2, and so on. However, with fixations lasting for some time, it can emerge that multiple data points of the same data source are assigned to the same fixation. In this case, interactions I4 and I5 are both assigned to fixation F8. This becomes a problem, depending on how the data is later represented. If one single visual element represents each fixation, this could lead to a duplication of the fixation as the case with the AOI sequence chart approach presented in Section 6.2. In the radial AOI timeline in Section 6.2, each data type has its own timeline and thus a duplication of fixations becomes superfluous.

Temporal Binning. Temporal binning considers fixation durations. The duration can be represented, for example, using bins of 50 ms allowing an accurate sampling of large variabilities of fixation durations [Cristino et al. 2010]. Analyzing data based on temporal aspects may be an important indicator how long participants stayed in one AOI. If one participant looked at an AOI quickly, this should be considered as different behavior than a participant inspecting an AOI for a long time. The simplest way is to divide each fixation into bins of a certain duration. If a fixation has a duration of 105 ms this would lead to three bins. However, in this case it might be more useful to create only two bins. We can solve this problem by defining a threshold for bins, for example, durations $\geq$25 ms lead to a new bin.

Considering the duration of each data source, with or without temporal binning, would lead to a different synchronization of the data. In Figure 4, fixations F1 and F2 would overlap with think aloud TA1. Mouse fixation M2 would be assigned to fixation F3 and F4. In this case, the same data points are assigned to multiple fixations. This again might lead to problems in the visualization if an AOI-based approach is used. Fixations belonging to different AOIs would require that data is represented multiple times.

5.4 AOI-based Synchronization

Eye tracking data is typically evaluated using AOIs. AOIs are specific regions on a stimulus or views and windows in a visualization system. Different approaches exist how to create AOIs. A grid, attention maps, or semantic information of a stimulus can be used [Holmqvist et al. 2011, p. 206ff]. Typically, the semantic information is used for analyzing interactive visualization systems. Depending on the evaluation requirements either AOIs cover a whole stimulus or only parts. In the first case, an AOI for those parts of a stimulus without semantic meaning may be defined. Thus, this approach ensures that all data is mapped to an AOI. Eye fixations, mouse fixations, and interactions are all assigned to the AOI they occurred in. If only parts of a stimulus are covered, it can occur that certain data is not represented or evaluated. If a participant interacted in an AOI while looking at a part of a stimulus no AOI covers, the fixation is not be mapped. In Figure 4, fixation F5 is not assigned to an AOI, however, interaction I3a and I3b would have to be mapped to this fixation. Thus, the interactions have to be visualized without their corresponding fixation.

An AOI-based approach for aligning eye tracking data, mouse fixations, and interaction data has some issues. Depending on the system evaluated, a large number of AOIs may have to be defined. Thus, an analysis on different levels using a hierarchical organization of AOIs, is an option. A hierarchy can be defined based on the semantic information of a stimulus [Blascheck et al. 2016a]. In the stimulus in Figure 5, AOI 1 contains AOIs 2 and 3. When analyzing this data, either all AOIs or only AOIs 1 and 4 could be used for an evaluation.

AOIs are also problematic if a system has overlapping transparent views, for example, when using focus-and-context techniques. In this case, it is not always possible to define which AOI contains which data point. Thus, data has to be mapped to multiple AOIs, leading to duplicate data. Another option is to assign uncertainty values to a fixation-AOI mapping. In Figure 5 the orange participant looked at AOI 3 (fixation F3) and then moved to a part which is overlapping (fixation F4). In this case, Fixation F4 could be assigned to AOI 2 or 3. However, it is more likely that the participant gazed at AOI 3 rather than AOI 2. Thus, this mapping could be assigned a higher value than the mapping to AOI 2.

Another challenge with interactive visualization systems are dynamic AOIs. Dynamic AOIs could be a window or view which a participant moves, removes, adds, or changes the content. Several authors [Blascheck et al. 2014; Kurzhals et al. 2014; Holmqvist et al. 2011, p. 209] have already discussed how to handle dynamic AOIs. In our case, dynamic AOIs require that they have to be logged during the data collection and thus, have to be defined before a study. Each window position, add/remove interaction, and the content of an AOI has to be stored. This can be done using the system, which also collects mouse movements and interactions. A listener could be attached to each UI element and if a UI element changes its position or its visibility this can be logged. In a timeline-based approach the timeline can be drawn with an interruption as depicted in Figure 6 (participant 02, AOI 3). If a dynamic AOI is transparent and represented on top of another AOI, the two AOIs can be drawn temporally aligned (participant 01, Fixation F4 in AOI 2 and 3).

In some cases, an interaction may influence a complete stimulus. For example, if parameters are changed in a program (one AOI) and the result of this change influences a different AOI, an analysis has to include this information. This requires an analyst to have semantic information about the evaluated system to take such dependencies into account. In section 6.2 we present two visual approaches, one where interactions are represented dependent on AOIs and one where they are independent of specific AOIs. Depending on the system evaluated one or the other approach can be used for an evaluation.

6 Data Analysis

An analyst may evaluate the aligned data applying different approaches. In eye tracking, calculating scanpath similarities is one approach. Additionally, a visual analysis is helpful in finding patterns in the data to generate and confirm hypotheses about participant behavior. We discuss two visualization approaches to combine the three data sources for an evaluation.

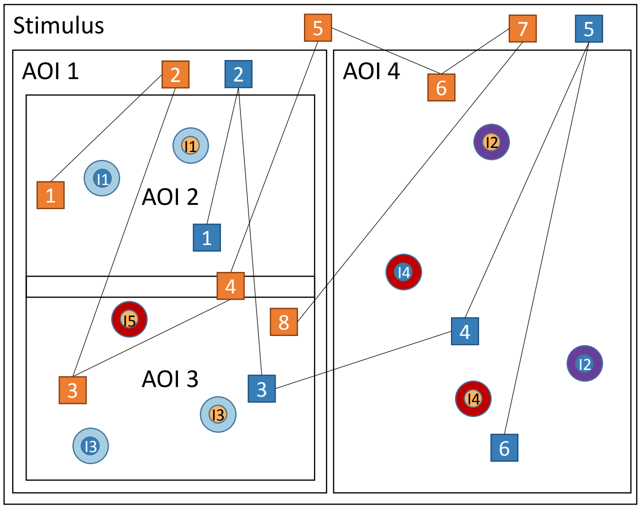


Figure 5: *Example stimulus with interleaved and overlapping AOIs. Rectangles connected by lines represent the fixations of two scanpaths (orange and blue). The circles depict interactions where the ordering is not necessarily temporally aligned to the fixations. The smaller inner circle represents the color of the participant and the outer circles is related to the interaction category.*

6.1 String-based Analysis

We propose to represent our data using a string. Each AOI hit (i.e., matching of a fixation or interaction to an AOI) represents a part of this string. This string includes AOIs for eye fixations as well as AOIs for interaction data, and the interaction category. This results in strings which are made up of 3-tuples (α, β, γ) where the first element in the tuple represents the AOI of a fixation, the second element the AOI of the interaction data, and the third element the interaction category. The symbols of the tuple have been chosen arbitrarily. We decided to use integers for AOIs and characters for interaction categories. One element is used to define fixations or interactions without an AOI or category. In our case, a 0 represents these AOIs and an X depicts these interaction categories. In Figure 5, the first string of the orange scanpath would be (2, 2, A) or short 22A, with fixation F1 in AOI 2, interaction I1 in AOI 2, and category A standing for an encode interaction. If no interaction took place the AOI of the current mouse position is used leading to strings like 22X. We added this exception in this paper, to also include the mouse position during the complete study.

Levenshtein Distance. A classical approach for comparing eye tracking scanpaths is calculating the string edit distance or **Levenshtein Distance** [Levenshtein 1966]. AOIs are used to represent scanpaths as strings. Based on these scanpath strings the Levenshtein distance is calculated. Privitera and Stark [2000] were the first to use Levenshtein distances to calculate similarities of scanpaths. The edit distance between two string is computed based on the minimum number of insert, delete, and substitute operations when transforming one string into another. If the number of transformations is small, scanpaths are more similar.

In the classical Levenshtein algorithm, transforming one string into another has the cost functions: replace = 1, delete and insert = 2. However, in our case different replace functions are used. If two elements in two compared tuples are the same a replace costs $1/3$, if one element is the same the cost is $2/3$, and if the complete string has to be replaced the cost is 1. The cost functions for delete and insert are still 2. We originally chose this approach, as we believe that when two participants are looking at the same AOI or using the same interaction in the same AOI this is more similar than if they are doing completely different things [Blascheck et al. 2016b]. To illustrate how we calculate the distances, Figure 5 shows an example stimulus with two scanpaths. The orange scanpath depicts the same data as the data in Figure 4. We assume the mouse stays where the interaction took place until the next interaction is executed. The string of the orange scanpath is 22B 14C 34C 34X 03B 43X 34A 33A. Fixation F7 is left out since it is not assigned to an AOI and has not corresponding interaction and fixation F8 has two interactions. The string of the blue scanpath, based on arbitrary times for the interaction data, is 22B 14C 34C 43B 03A 44X. Transforming the orange scanpath into the blue would require a Levenshtein distance of 5 $2/3$. The first three tuples are the same. The fourth requires a replace of 34X into 43B costing 1, the 03B is replaced with 03A costing $1/3$, 43X is replaces with 44X costing $1/3$, and 34A and 33A are deleted each costing 2. This results in a distance of $1 + 1/3 + 1/3 + 2 + 2 = 5\ 2/3$.

Needleman-Wunsch. Other algorithms for sequence comparison were developed in Bioinformatics. For example, the Needleman-Wunsch algorithm [1970] is a related approach to Levenshtein distance. However, it uses a substitution matrix instead of cost functions and calculates a similarity score where a higher score indicates that two sequences are more similar. The Needleman-Wunsch algorithm has some benefits over the Levenshtein distance. First, it includes the relationship between AOIs. This is particularly helpful when analyzing interactive visualization systems, as views or windows are often used in combination. In Figure 5 fixations F1, F2, F3, F4, and F8 of the orange scanpath indirectly belong to AOI 1, as AOIs 2 and 3 are nested into AOI 1. We can include this information and define a similarity matrix where AOIs 1, 2 and 3 are more similar (distance is 1) than AOI 4 (distance is -1), as shown in Table 1. Second, the Needleman-Wunsch algorithm can include temporal aspects of data such as the duration of a fixation using temporal binning (cf. Section 5.3). If in the blue scanpath fixation F1 lasts for 100 ms, fixation F3 for 150 ms, interaction I3 for 100 ms, and the rest for 50 ms then the string would be: 22X 22B 14C 34C 34C 34X 43B 03A 44X.

	AOI1	**AOI2**	**AOI3**	**AOI4**
AOI1	2	1	1	-1
AOI2	1	2	1	-1
AOI3	1	1	2	-1
AOI4	-1	-1	-1	2

Table 1: *Needleman-Wunsch similarity matrix. We define that AOIs 1, 2, and 3 are more closely related and thus assign them values of 1. Between this group and AOI 4 we define the similarity to be -1. Each AOI is equal to itself, thus a score of 2 is assigned.*

6.2 Visualization-based Analysis

We propose two visualization approaches for analyzing eye tracking, interaction, and think aloud data. The first approach is based on an AOI sequence chart and the second is an interactive radial approach based on the parallel scanpath visualization [Raschke et al. 2012]. A visual approach has the benefit that an analyst can inspect and explore data to find unexpected patterns or behavior of participants and to develop hypotheses about the causes of patterns.

AOI Sequence Chart. An AOI sequence chart represents AOIs of a stimulus as separate timelines for each participant [Holmqvist et al.

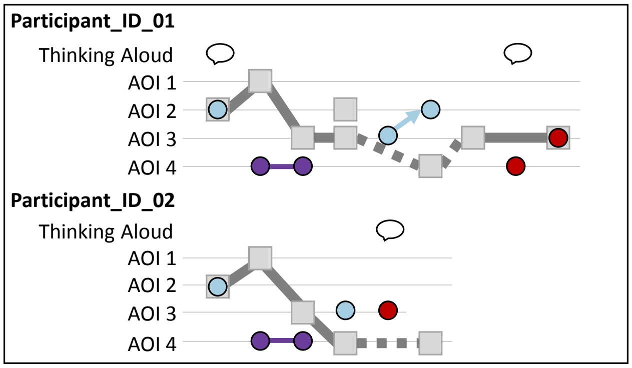


Figure 6: *An AOI sequence chart [Holmqvist et al. 2011, p. 197] showing discrete time steps. The interrupted timeline for AOI 3 of participant 2 indicates that this AOI was invisible for some time. The gray rectangles represent eye fixations and are connected through gray lines. If the gray line is dotted the eyes moved to a part of the stimulus not covered by any AOI. Two rectangles temporally aligned represent transparent AOIs. Colored circles represent interactions. If an interaction is started in one AOI and influences other AOIs an arrow is drawn. Speech bubbles represent think aloud data on a separate timeline above the AOI timelines.*

2011, p. 197]. In our previous publication, we described how we adapted this visualization to show also interaction and think aloud data [Blascheck et al. 2016b]. Figure 6 shows an AOI sequence chart representing the data from the two scanpaths in Figure 5. We use linearly ordered time sequences to depict the data. A square represents each AOI hit of eye fixations. Gray lines representing the transitions between AOIs connect eye fixation rectangles. If AOIs do not cover the complete stimulus, gray dotted lines are used to represent time spans where a participant was looking at a part of a stimulus no AOI covers. On top of eye fixations we draw interaction data as a circle. The color-coding of the circles represents the interaction category assigned to an interaction. Lines connect interactions if they represent continuous interactions, for example, a scroll or drag interaction lasting for a certain time period. Think aloud data is represented as speech bubbles on a separate timeline on top of the AOIs, as they are independent of specific AOIs.We have further adapted this visualization technique to represent temporarily hidden and transparent AOIs and interactions influencing multiple AOIs. An arrow represents this case and depicts interactions triggered in one AOI and affecting other AOIs, as shown in Figure 6 (participant 1, interaction I3). If an AOI timeline is interrupted, as shown in Figure 6 (participant 2, AOI 3), an AOI was invisible (i.e., collapsed) for a certain timespan. Furthermore, we now support transparent AOIs. A transparent AOI may contain one or multiple other AOIs leading to time sequences where two AOI rectangles are drawn at the same time as depicted in Figure 6 (participant 1, fixation F4).

Some disadvantages of the AOI sequence chart are, that data may be represented multiple times as discussed in Section 5.4. Additionally, if the recorded data is long or the number of AOIs becomes large, the AOI sequence charts get large and would require a lot of space or scrolling. Furthermore, representing an AOI sequence chart for each participant individually, makes a comparison of participants challenging. In addition, this approach does not include mouse movements at the moment.

Radial AOI Timeline. A different representation using a radial layout of AOIs is shown in Figure 7. We implemented an interactive prototype to show how to use this approach for analyzing eye tracking and interaction data. At the moment this prototype does not include think aloud data, however, it includes mouse movements.

Evaluation View. The main evaluation view contains eye fixations, mouse fixations, and interactions for one study task and all selected participants. We chose a space saving visualization technique using a radial layout for a good comparison and overview of participants. Each circle segment, delineated with a black line, contains data of one participant. Each small segment within such a participant segment represents AOIs of a study (gray line). These separate timelines of participants and AOIs reduce visual clutter. Time starts in the middle and ends at the outer part of the circle. Green lines with circles represent a mouse path. Circles represent mouse fixations where the radius of a circle corresponds to a fixations' duration. Brown circles serve as eye fixations where the radius also represents the duration of a fixation. AOIs contain and depict corresponding mouse and eye fixations. This data is depicted on top of each other, to show correlations between them. In contrast to the AOI sequence chart in this technique data is not duplicated, for example, when a fixation contains two interactions as the duration of the fixation is represented explicitly. The evaluation view represents interactions as lines across all AOIs, since they influence a stimulus as a whole. The color of each interaction corresponds to the category this interaction belongs to. For example, red corresponds to the encode, purple to the navigate, and blue to the select category [Blascheck et al. 2016b]. The color coding is chosen from Color Brewer[1] for representing qualitative data. Thus, an analyst can see which interactions participants used the most or the least or at what times different interaction types were frequently used.

Visualization Settings and Interactions. The visualization offers different interaction mechanisms. In Figure 7 (A), an analyst can change visualization settings, for example, resize the view, zoom in and out, and shift the data part shown. Zooming in shows a specific time range in more detail. For example, an analyst might be interested especially in the end of a task. After zooming into the view, time intervals can be selected using a shift slider to look at specific time intervals in more detail (e.g., the beginning of a task). The size of the view can be changed on demand, depending on the screen resolution. Additionally, the view may represent data in relation to all participants or absolute for each participant (in Figure 7 icon). This feature is especially important as different participants may spent a different amount of time on the stimulus leading to a cluttered representation close to the center if all participants are shown in relation to each other. To counteract this issue, our approach offers to switch between a relative and absolute representation of the completion times of participants. Figure 7 shows data represented in relation to all participants. Thus, an analyst is able to investigate how long individual participants took. If absolute times of each individual participant are depicted the whole space of each circle segment is used, enabling to see more details about individual participants.

The evaluation view represents eye movements (), interactions (), or mouse fixations () individually or in combination. If an analyst is interested only in eye movements or mouse movements, the other data sources can be switched off. Hovering over any visual element in the view shows additional information. Each data source depicts additional information, such as timestamps within a tooltip. For eye tracking data and mouse fixations, the AOI and fixation duration are shown. For interactions, it depicts the category and specific information of this category. Details shown on demand reduce clutter in the visualization. Clicking on any visual element in the view shows a snapshot of the stimulus at this particular timestamp. Here, an analyst can see the states of an analyzed system for an individual participant.

[1]http://colorbrewer2.org/

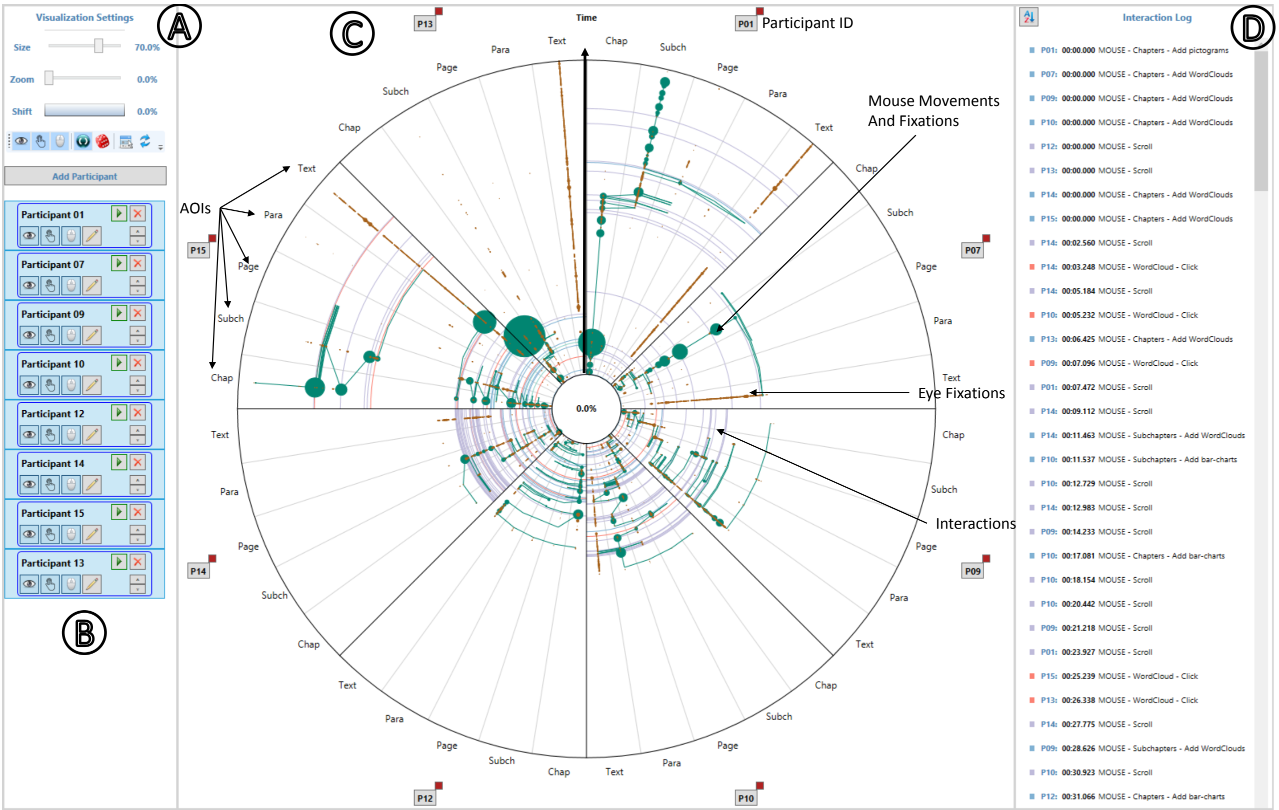


Figure 7: *The radial AOI timeline represents eye tracking and interaction data in a combined way. On the upper left (A), an analyst can change settings for the visualization. The lower left (B) shows a list of participants. The middle part (C) depicts the radial AOI timeline, and the right side (D) contains an interaction log of all participants.*

Participant List. In the participant list (cf. Figure 7 (B)) an analyst can select or deselect participants to be shown in the radial AOI timeline. Selecting only some participants might be useful if an analyst wants to compare participant groups. The actual participant video may be replayed ( ). This permits to look at the actual screen cast of a participant performing a task. Additionally, data for each participant can be edited ( ) and the analyst may change the order of participants( ). For example, if participants display similar behavior changing the order in the participant list would lead to a reordering of participants in the visualization. This would facilitate to better compare those participant groups.

Interaction Log. On the right side of Figure 7 (D), an interaction log is depicted additionally, ordered by the timestamp. However, the list can be reordered ( ) to show interactions for each participant separately. Enabling to analyze which interaction a single participant used at what points in time. Each entry contains a rectangle in the color of the interaction category. Next to a rectangle, the log shows a participant ID, and detailed information about an interaction. The interaction log allows to see which interactions all participants used at what point in time. Additionally, individual interactions of participants can be investigated in more detail. For example, how often a participant used which interaction categories.

The radial approach has some drawbacks. For example, if the number of participants or AOIs is large, it is difficult to visualize everything in the same radial AOI timeline. Although, the radial layout is space saving it is restricted to representing only few AOIs and participants. Currently, think aloud data is not included into this approach. On the other hand, the advantage of this space-saving visualization technique is, that it facilitates an analyst to get a fast overview of multiple participants. Furthermore, an analyst can see which AOIs have been looked at and which interactions have been used the most frequently.

7 Conclusion and Future Work

We presented an approach for evaluating interactive visualization systems by collecting, synchronizing, and analyzing eye tracking, interaction, as well as think aloud data. We discussed how the different data sources are collected and synchronized afterwards to detect common strategies of participants performing a task. Our approach works best, if these data sources are temporally aligned using eye fixations and AOIs. We showed how the three data sources are analyzed using a string-based and visualization-based analysis. The radial visualization was implemented as a prototype to show how parts of the data is analyzed visually.

In the future, we plan to integrate think aloud data into the analysis more seamlessly, to improve the correlation of modal, causal, and intentional behavioral aspects between participants. Furthermore, we want to investigate the behavior of dynamic AOIs and how they can be integrated into our approach, especially if the AOI is changing as well as the content. Currently, we have evaluated two visualization systems with our approach. We plan to analyze more visualization systems in the future. This may require to develop different visualization techniques to overcome the drawbacks of the current approaches and integrate further requirements.

Acknowledgments

This work has been partially funded by the German Federal Ministry of Education and Research (BMBF) as part of the 'ePoetics' project.

References

AIGNER, W., MIKSCH, S., SCHUMANN, H., AND TOMINSKI, C. 2011. *Visualization of Time-Oriented Data*. Human-Computer Interaction Series. Springer.

BECK, F., BLASCHECK, T., ERTL, T., AND WEISKOPF, D. 2015. Exploring word-sized graphics for visualizing eye tracking data within transcribed experiment recordings.

BLASCHECK, T., AND ERTL, T. 2014. Towards analyzing eye tracking data for evaluating interactive visualization systems. In *Proceedings of the 2014 BELIV Workshop*, 70–77.

BLASCHECK, T., KURZHALS, K., RASCHKE, M., BURCH, M., WEISKOPF, D., AND ERTL, T. 2014. State-of-the-Art of Visualization for Eye Tracking Data. In *EuroVis - STARs*, 63–82.

BLASCHECK, T., KURZHALS, K., RASCHKE, M., STROHMAIER, S., WEISKOPF, D., AND ERTL, T. 2016. AOI hierarchies for visual exploration of fixation sequences. In *Proceedings of the Symposium on Eye Tracking Research & Applications*, (in press).

BLASCHECK, T., JOHN, M., KURZHALS, K., KOCH, S., AND ERTL, T. 2016. VA^2: A visual analytics approach for evaluating visual analytics applications. *IEEE Transactions on Visualization and Computer Graphics 22*, 01, 61–70.

BREHMER, M., AND MUNZNER, T. 2013. A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics 19*, 12, 2376–2385.

CARPENDALE, S. 2008. Evaluating information visualizations. In *Information Visualization*, A. Kerren, J. T. Stasko, J.-D. Fekete, and C. North, Eds., vol. 4950 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 19–45.

CRISTINO, F., MATHÔT, S., THEEUWES, J., AND GILCHRIST, I. 2010. ScanMatch: A novel method for comparing fixation sequences. *Behavior Research Methods 42*, 3, 692–700.

DOU, W., JEONG, D. H., STUKES, F., RIBARSKY, W., LIPFORD, H., AND CHANG, R. 2009. Recovering reasoning processes from user interactions. *IEEE Transactions on Computer Graphics and Applications 29*, 3, 52–61.

HOLMQVIST, K., NYSTRÖM, M., ANDERSSON, R., DEWHURST, R., JARODZKA, H., AND VAN DE WEIJER, J. 2011. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press.

KURZHALS, K., HEIMERL, F., AND WEISKOPF, D. 2014. ISeeCube: Visual analysis of gaze data for video. In *Proceedings of the Symposium on Eye Tracking Research & Applications*, 43–50.

LEVENSHTEIN, V. 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics-Doklady 10*, 8, 707–710.

LIPFORD, H., STUKES, F., DOU, W., HAWKINS, M., AND CHANG, R. 2010. Helping users recall their reasoning process. In *IEEE Symposium on Visual Analytics Science and Technology*, 187–194.

MUNZNER, T. 2014. *Visualization Analysis and Design*, 1 ed. AK Peters Visualization Series. CRC Press Taylor and Fracis Group.

NEEDLEMAN, S. B., AND WUNSCH, C. D. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology 48*, 3, 443 – 453.

NORTH, C. 2006. Toward measuring visualization insight. *IEEE Transactions on Computer Graphics and Applications 26*, 3, 6–9.

PIKE, W. A., STASKO, J., CHANG, R., AND O'CONNELL, T. A. 2009. The science of interaction. *Information Visualization 8*, 4, 263–274.

POHL, M. 2012. Methodologies for the analysis of usage patterns in information visualization. In *Proceedings of the 2012 BELIV Workshop*, 17:1–17:3.

PRIVITERA, C., AND STARK, L. 2000. Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *IEEE Transactions on Pattern Analysis and Machine Intelligence 22*, 9, 970–982.

RASCHKE, M., CHEN, X., AND ERTL, T. 2012. Parallel scan-path visualization. In *Proceedings of the Symposium on Eye Tracking Research & Applications*, 165–168.

REDA, K., JOHNSON, A., LEIGH, J., AND PAPKE, M. 2014. Evaluating user behavior and strategy during visual exploration. In *Proceedings of the 2014 BELIV Workshop*, 70–77.

SARAIYA, P., NORTH, C., AND DUCA, K. 2005. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE Transactions on Visualization and Computer Graphics 11*, 4, 443–456.

SMUC, M., MAYR, E., LAMMARSCH, T., AIGNER, W., MIKSCH, S., AND GARTNER, J. 2009. To score or not to score? tripling insights for participatory design. *IEEE Transactions on Computer Graphics and Applications 29*, 29–38.

TENBRINK, T. 2014. Cognitive discourse analysis: Accessing cognitive representations and processes through language data. *Language and Cognition 7*, 1, 98 – 137.

YI, J. S., KANG, Y., STASKO, J. T., AND JACKO, J. A. 2008. Understanding and characterizing insights: How do people gain insights using information visualization? In *Proceedings of the 2008 BELIV Workshop*, 4:1–4:6.