1 INTRODUCTION

Eye tracking data recorded in user studies is commonly analyzed using statistical methods. Visualizations depicting the data complements these methods by supporting more exploratory analysis and providing deeper insights into the data. Visualization research nowadays provide a body of techniques to visually represent the spatial and temporal dimensions of the recorded eye movements [7]. Eye tracking data, however, is only one of many data streams—such as video, audio, and interactions—that are usually recorded during an experiment. For instance, when applying a think-aloud protocol, a transcript of the oral statements is a particularly rich source that could explain the behavior of the participant on a higher level. To support an analyst to leverage the full potential of the recordings, it is important to integrate all streams of information within a single approach.

In this work, we focus on the integration of transcribed statements of individual participants and eye tracking data into a visually augmented user interface (Figure 1). Unlike other visualization approaches (Section 2), we handle the transcribed text as a first-class entity which we complement with word-sized eye tracking visualizations in a tabular chronological representation (Section 3). We systematically explore the design space of these word-sized visualizations, also known as sparklines [47], for eye tracking data by discussing how existing eye tracking visualizations can be scaled down to word size (Section 4). Similar visualizations can be used for representing interaction logs; the small size of visualizations allows us to combine multiple eye tracking and interaction visualizations within a user interface. We implemented a prototype of the suggested user interface (Figure 1, right) as a details-on-demand view for a visual analytics framework for eye tracking data by Blascheck et al. [6] (Section 5). We see our main contributions in designing novel word-sized variants of established eye tracking visualizations and demonstrating how these can be leveraged as part of an interactive transcript-focused analysis tool.

2 RELATED WORK

There are various approaches for visualizing eye tracking data as Blascheck et al. [7] surveyed. Those focusing the analysis to an individual participant are more closely related to our work, for instance, approaches that represent the spatial coordinates of fixations and saccades [13, 21, 35] or approaches that abstract this data to fixations on areas of interest (AOIs) and transitions between those [12, 15, 22, 25, 28, 41]. Also, a number of visualizations of interaction logs have been investigated, for instance, for interactions of software developers in IDEs [36], interactions with visualization systems [17, 34, 43], or provenance information in scientific workflows [23]. However, only few approaches integrate eye tracking or log visualizations with transcribed experiment recordings: Holsanova [26] connect transcribed picture descriptions with picture viewing data on a simple timeline showing both text and events. Franchak et al. [19] extend such a timeline with other events, in their case, interactions of infants with their environment. ChronoViz [48] includes a transcript view complementing a separate timeline view of eye tracking data and other event-based data. Blascheck et al. [6] combine eye tracking and interaction data in an extended timeline; the transcript is retrievable on demand only for individual time spans. Our approach, in contrast to these, puts a greater focus on text and handles eye tracking and interaction data only as context of the transcript.

A common method for integrating text and visualization—in particular when the text should not only be a supplement to the visualization—are word-sized graphics, also called sparklines [47].
They can be integrated in all textual representations, such as natural-language text [20, 47], tables [47], source code [2, 4], visualizations [9, 33], or user interfaces [3]. In this paper, we integrate them into columns of a tabular representation as additional information for transcribed experiment recordings. Being a kind of scaled-down information visualization, sparklines might represent any kind of abstract data, however, only under restricted space constraints. To the best of our knowledge, sparklines have not been used so far for representing eye tracking or interaction log data.

There are also annotation and coding tools for transcribed experiment recordings. In context of psycholinguistics, ELAN [11, 45] supports the analysis of orthographic and phonetic transcriptions. Another tool for linguistic analysis of spoken text is ANVIL [29]. It allows the integration of multimodal audiovisual material and was later extended to include spatiotemporal information of videos [30] and motion capturing [31]. None of these tools, however, supports the analysis of eye tracking and interaction data along with the text.

3 Setting

Our goal is to provide an analysis tool that enriches a transcribed experiment recording (e.g., from a think-aloud protocol) with eye tracking information. We focus on analyzing a single participant at a time, for instance, as part of a data exploration step or a systematic coding of performed activities. The integrated visualization, in addition to text, should enable the analyst to make informed data analysis and coding decisions without having to switch between multiple tools or visualizations.

We assume that a transcript is divided into activities having a precise start and end time. The stimulus used in an experiment can either be static or dynamic. In the dynamic case, we want to be flexible enough to support video stimuli as well as interactively changeable stimuli such as user interfaces. A visual encoding of interaction logs is a secondary goal for our approach. Interaction events typically carry a timestamp when a participant triggered them, a spatial position that describes their location, and can be classified into different abstract categories such as selection, encoding, navigation, etc.

The eye tracking data consists of a sequence of fixations with spatial coordinates as well as start and end times; saccades describe quick eye movements between individual fixations. Some of the visualizations discussed in the following require that a stimulus has been annotated with areas of interest (AOIs), summarizing sets of fixations into spatial groups. Individual transitions between AOIs can be considered as a graph, either aggregated over time as a static graph or reflecting the temporal order of transitions as a dynamic graph [12].

Our solution as outlined in Figure 1 (right) is based on representing the transcript in a table, showing one activity per line in chronological order. Besides a column containing the actual transcript text, additional columns provide context about timing, eye tracking, and interaction events that happened during the respective activity. Since the tabular representation does not allow us to integrate large visualizations, we use word-sized eye tracking visualizations. Due to the division of time into short activities, each sparkline only needs to show a small amount of data. As an additional help to make the visualizations more readable, a larger version of each word-sized graphic is retrievable on demand as part of a sidebar. The sidebar also allows us to show the recorded video stream of a specific activity, with eye tracking and interaction data potentially overlaid.

4 Word-Sized Eye Tracking Visualizations

A central element of our approach is the representation of eye tracking data as word-sized visualizations. Since many approaches already exist for visualizing this data in normal-sized graphics [7], we take these as a starting point for developing word-sized variants showing similar data. This transformation usually requires one to simplify the visualization approach: in particular, one cannot, or at least should not, label visual objects with text, use thin lines or border lines for objects, waste space by separating objects using white space, or show 3D graphics. Moreover, a sparkline—like a word—usually has a panorama format, being limited to the line height of the text but having some space on the horizontal axis.

To explore the design space of those visualizations in a systematic way, we analyze all eye tracking visualization techniques Blascheck et al. [7, Table 1] surveyed and try to transfer each approach to a word-sized visualization. Since we only target at visualizing the data recorded for a single participant, we exclude all visualizations focusing on comparing or aggregating multiple participants. Further, we are not able to suggest meaningful word-sized variants of some techniques, in particular, because of the use of 3D views [1, 18, 32, 39], the original stimulus [16, 27, 42] (the stimulus usually is too complex to be represented within a sparkline), circular layouts [8, 27, 40, 44] (advanced circular layouts are hard to fit to the elongated format of a sparkline), or a specialization to particular kinds of stimuli [5, 46]. As a result, we come up with a list of visualization techniques that can be adequately transferred to miniaturized graphics. Below, we discuss all these miniaturized visualization techniques by showing an example embedded in the text and defining the specific visual encoding in Table 1. We furthermore describe the modifications needed when using the visualizations as word-sized graphics. All visualizations shown in this section are manually created drafts encoding artificial data. Some of them are implemented as examples in our prototype analysis tool (Section 5).

4.1 Point-Based Visualizations

Each fixation is assigned a coordinate on the stimulus that represents the estimated location a participant looked at. This information is a rich data source for interpreting eye movement data, together with durations and saccades between fixations.

Space. Focusing on the spatial part of the data, the standard representations of eye tracking data are scan paths and heat maps. Scan path visualizations simply overlay the trajectory of the gaze onto the stimulus [38], often encoding fixations as circles scaled according to their duration [24]. For the word-sized variant, we do not show the stimulus or fixations, but just plot the trajectory as a line (P1

In contrast, heat maps, also called attention maps, aggregate fixation durations for spatial coordinates, which are color-coded and overlaid onto the stimulus [24, 35]. For a word-sized attention map, we suggest to plot a coarsely gridded map [24] into the sparkline representation and encode the duration in the darkness of the grid cells (P2

As an alternative, we could focus on only

<table>
<thead>
<tr>
<th>Table 1. Design space of word-sized eye tracking visualizations.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Visualization</strong></td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td><strong>Point-Based Visualizations</strong></td>
</tr>
<tr>
<td>P1</td>
</tr>
<tr>
<td>P2</td>
</tr>
<tr>
<td>P3</td>
</tr>
<tr>
<td>P4</td>
</tr>
<tr>
<td>P5</td>
</tr>
<tr>
<td>P6</td>
</tr>
<tr>
<td><strong>AOI-Based Visualizations</strong></td>
</tr>
<tr>
<td>A2</td>
</tr>
<tr>
<td>A3</td>
</tr>
<tr>
<td>A4</td>
</tr>
<tr>
<td>A5</td>
</tr>
<tr>
<td>A6</td>
</tr>
<tr>
<td>A7</td>
</tr>
<tr>
<td>A8</td>
</tr>
</tbody>
</table>

Legend: seq.–sequences; trans.–transitions; freq.–frequency; dur.–duration; direct.–direction; ref.–references.
one spatial axis, again encode duration in the color, and use bar charts
to encode another metric, such as the frequency of fixations within
the respective area (P3). Spatial information can also be
restricted otherwise to make them representable at small scale, for in-
tance, encoding angles of the trajectory in radial diagrams [21].

Space and Time. The temporal sequence of fixations is also im-
portant for some analysis scenarios. Mapping time to a spatial di-
imension, however, requires the encoding of spatial information to be
limited [21, 49]. For instance, using the longer x-axis as a time-
line, the y-axis could encode one of the spatial coordinates of the
fixations where darkness indicates the distribution of fixation dura-
tions (P4). We can also extend scan paths with temporal infor-
mation by using the edge color for encoding time (P5).

This is similar to Saccade Plots [13] that show saccades (i.e., the jumps
between fixations) at the side of a stimulus. Leaving out the stimulus,
we could use a similar approach within a sparkline plotting a spatial
coordinate on the x-axis and connecting points with arcs according to
observed saccades (P6)—like in the original approach, arcs are
redirected from left to right on top of the axis, whereas arcs in the
opposite direction are below.

4.2 AOI-Based Visualizations

AOIs abstract from the exact location of fixations to semantic regions
on a stimulus, which an analyst usually defines manually. AOIs also
allow us to build a transition graph connecting the AOIs according
to the sequence they were looked at. We assume for the following visual-
izations that we have to handle five to ten different AOIs. Due to
the limited size of our visualizations, most of the suggested approaches
do not scale to more AOIs, but according to our experience, ten AOIs
suffice for the majority of application scenarios.

AOI Statistics. One of the most simple AOI-based visualizations is
to depict the frequency or total duration each AOI was fixated, for
instance, in a line or bar chart. Such diagrams can be directly trans-
ferred into word-sized graphics. We decide to use bar charts because
ingraphic providing a different perspective onto the data. For instance, it
could be useful to combine a point-based and an AOI-based visualiza-
tion. If the application scenario allows the use of several lines en bloc, a vertical stacking of the sparklines (i.e., placing
them on top of each other) is possible. To align both visualizations,
the x-axes should have the same encoding, for example, a color-coded
sequence of AOIs combined with a duration encoding:

Interaction Data. Interaction data shares characteristics to eye
tracking data: Much like fixations, interactions are temporal events
on the same experiment time dimension. They can be classified ac-
cording to their type into categories or assigned to AOIs based on their
location. Also, transitions between interactions might be derived from
the sequence of logged events. One difference, however, is that usu-
ally interaction events do not have a duration; they only get a tempo-
ral dimension if they are abstracted to longer sequences of semanti-
cally linked interactions. The general similarity between the two data
streams now allows us to reuse most of the suggested word-sized eye
tracking visualizations for interaction data. Furthermore, the discussed
horizontal and vertical juxtaposition of these sparklines provides an
easy way of integrating both data sources within one user interface.

5 Prototype Implementation

We implemented the approach as a detail view of a larger visual analy-
sis framework for eye tracking studies [6]. The visual analysis frame-
work is intended to analyze eye tracking and interaction data together.
In the original implementation, think-aloud data was added to enrich
the other two data sources. In the new detail view, in contrast, we
intend to present the think-aloud protocol in detail and enrich it with
eye tracking and interaction data. This prototype is a proof of concept
implementing two AOI-based and two point-based versions of word-
sized visualizations.

Figure 2 shows a screenshot of our prototype, depicting data of one
participant in a temporal order. A tabular view represents the main
part of the prototype. For each verbal statement, word-sized visualiza-
tions are shown, in one column the two point-based visualizations, in
another the two AOI-based ones. In both columns, the visualizations
for eye movements and interactions are juxtaposed vertically, showing
the eye tracking visualization above the interaction visualization.

The point-based visualizations are gridded attention maps (Table 1, P2)
or, respectively, maps showing the spatial distribution of interactions. We divided the stimulus into 25 columns and five rows.
For each cell, we counted the fixation durations and the count of inter-
actions and color-coded the cells accordingly. The color coding was
obtained from ColorBrewer [10], using a sequential, single-hue blue
color and a gradation of four (fixation duration ≤ 10 ms, ≤ 100 ms,
≤ 1000 ms, and > 1000 ms; interaction count ≤ 1, ≤ 3, ≤ 5, and > 5).

Our AOI-based visualizations (Table 1, A4) represent each AOI as a row of rectangles. Since only one AOI is active at a
time, we assign a height to each rectangle greater than the row height to increase the size of the rectangles (which improves color perception). In the eye tracking visualization at the top, for each individual AOI fixation, the duration is calculated and the AOI rectangle is colored based on the duration. We chose a sequential, single-hue gray color scale and a logarithmic gradation of four (AOI fixation duration $\leq 10$ ms, $10 < $ AOI fixation duration $\leq 100$ ms, $100 < $ AOI fixation duration $\leq 1000$ ms, and $>1000$ ms). For the visualization of interaction data below, interactions are assigned to AOIs and the color is determined by the categorical interaction category. For example, an interaction from the category *encode* is shown in red, a *select* interaction in light blue, and a *navigate* interaction in purple. The interactions are temporally aligned with the AOI fixations, thus, representing interactions at the point in time of its corresponding AOI fixation.

Based on the eye movement and interaction data depicted in the word-sized visualizations, an analyst adds categories to the activities. Additionally, rows and columns might be reordered. On the right side of the prototype, a video playback is shown for further reference. The playback might be combined with an animated representation of the eye tracking or interaction data, in our case, a dynamic scanpath overlay retrieved from Tobii Studio. Below the video, the visualizations of a selected row are shown enlarged and annotated with labels.

For a small use case example, we re-analyzed data from a user study testing a visual analysis tool [6] (participant 15, transcript translated from German to English). We first explore the data trying to get an overview. We find that in the point-based visualization at the beginning most of the fixations are in the upper part of the stimulus (Activities 1, 2, 3, 5), whereas later, most of the fixations are in the lower part of the stimulus (Activities 13, 14, 15). In the AOI-based sparkline, it becomes apparent that, at the beginning, the participant used mostly *encode* and *select* interactions in the first two AOIs (Activities 1, 2) while focusing mostly in AOI 1 and 2. At the end, the participant used more *navigate* interactions (Activities 8, 9, 10) and was looking at AOIs 4 and 5 more often. These kinds of analysis allows us to classify the participant’s activities and manually assign categories in the respective table column (Figure 2).

6 CONCLUSION AND FUTURE WORK

With a focus on analyzing the transcribed experiment recording of a single participant, we suggested a novel approach to visually enrich the textual representation of a transcript with eye tracking and interaction data. This data is represented in word-sized visualizations that provide different perspectives onto the data. We systematically explored the design space of word-sized eye tracking visualizations and prototypically implemented the approach as the visual analysis framework for eye tracking studies.

Since our implementation is work in progress, it only partly covers the suggested visualizations yet, still lacks important interactive features, and only provides rudimentary support for coding. We will extend the implementation toward a full-fledged visual analysis and coding system. Moreover, we want to explore which of the suggested visualizations is most effective and efficient for analyzing the data and at the same time easy to understand for potential users. Beyond that, we are interested in exploring other application scenarios for the suggested visualizations, for instance, their use to communicate results of eye tracking studies in scientific publications.

ACKNOWLEDGMENTS

Fabian Beck is indebted to the Baden-Württemberg Stiftung for the financial support of this research project within the Postdoctoral Fellowship for Leading Early Career Researchers.

REFERENCES


