Word-Sized Eye Tracking Visualizations

Fabian Beck, Tanja Blascheck, Thomas Ertl, and Daniel Weiskopf

Abstract In user studies, eye tracking is often used in combination with other recordings, such as think-aloud protocols. However, it is difficult to analyze the eye tracking data and transcribed recordings together because of missing data alignment and integration. We suggest the use of word-sized eye tracking visualizations to augment the transcript with important events that occurred concurrently to the transcribed activities. We explore the design space of such graphics by discussing how existing eye tracking visualizations can be scaled down to word size. The suggested visualizations can optionally be combined with other event-based data such as interaction logs. We demonstrate our concept by a prototypical analysis tool.

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 provides 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 user 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 (Fig. 1). Unlike other visualization approaches (Section 2), we handle the transcribed text as

Fabian Beck, Tanja Blascheck, Thomas Ertl, and Daniel Weiskopf
University of Stuttgart, Germany, e-mail: {firstname.lastname}@vis.uni-stuttgart.de

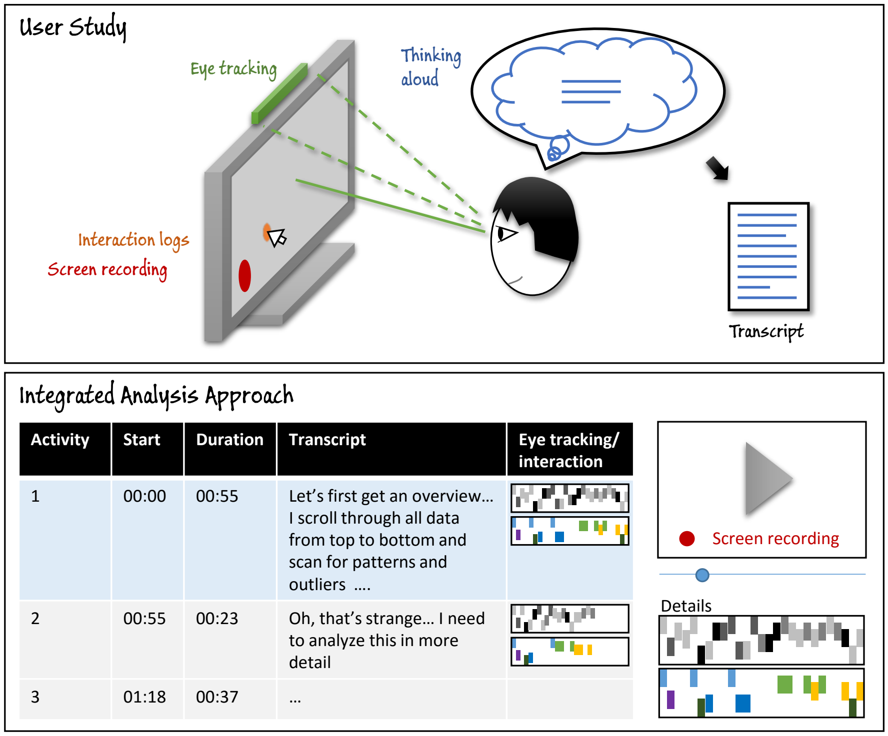


Fig. 1 Illustration of our approach to analyze transcribed recordings of user studies (e.g., based on think-aloud protocols): we integrate word-sized eye tracking and interaction visualizations into a tabular representation of the transcript and provide screen recordings and enlarged visualizations on demand.

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* [49], 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 (Fig. 1, bottom) as a details-on-demand view for a visual analytics framework for eye tracking data [6] (Section 5). To illustrate the applicability of our approach, we used the implementation to reanalyze and detail the results of an eye tracking study (Section 6). A discussion sheds light on strengths, shortcomings, and other areas of application of our approach (Section 7). 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 to visualize eye tracking data as Blascheck et al. [7] surveyed. Those focusing the analysis to an individual participant are 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 are available, 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 [50] 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* [49]. They can be integrated in all textual representations, such as natural-language text [20, 49], tables [49], 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 annotation and coding tools for transcribed experiment recordings. In the context of psycholinguistics, ELAN [11, 46] 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.

We assume that 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 [12].

Formally, we define a sequence of fixations as $F = (f_1, \ldots, f_n)$ where $f_i = (x_i, y_i, t_{1_i}, t_{2_i}, c_i)$ describes an individual fixation as a 5-tupel: x_i and y_i denote the coordinates of the fixation, t_{1_i} and t_{2_i} are the start and end times, and c_i refers to a categorial attribute, such as an AOI. A derived attribute is the duration $t_{d_i} := t_{2_i} - t_{1_i}$. Moreover, we define an aggregated duration Δ as a function of different arguments summing up all durations t_{d_i} for $f_i \in F$ under certain conditions:

- for $\Delta(x,y)$, in an area around (x,y) (with constants a_x, a_y)

$$\Delta(x,y) = \sum_{f_i \in F_{x,y}} t_{d_i}, \quad F_{x,y} = \{f_i \in F | x \leq x_i < x + a_x \wedge y \leq y_i < y + a_y\},$$

- for $\Delta(x)$, in an area around x (with constant a_x)

$$\Delta(x) = \sum_{f_i \in F_x} t_{d_i}, \quad F_x = \{f_i \in F | x \leq x_i < x + a_x\},$$

- for $\Delta(t,x)$, in a spatio-temporal area around (t,x) (with constants a_x, b; fixations first need to be split into separate fixations at t and $t+b$ as long as there exists a fixation that spans across one of these interval limits)

$$\Delta(t,x) = \sum_{f_i \in F_{t,x}} t_{d_i}, \quad F_{t,x} = \{f_i \in F | t \leq t_{1_i} \wedge t_{2_i} < t + b \wedge x \leq x_i < x + a_x\}, \text{ and}$$

- for $\Delta(c)$, with a specific category c

$$\Delta(c) = \sum_{f_i \in F_c} t_{d_i}, \quad F_c = \{f_i \in F | c = c_i\}.$$

Analogously, function σ counts the number of fixations according to the same arguments and conditions.

To define a temporal transition graph $G = (V,E)$, we use the categories as nodes ($V = \{c_1,\ldots,c_n\}$) and insert edges $e = (c_i, c_{i+1}, t_{2_i}, t_{1_{i+1}})$ to E for every pair of subsequent fixations f_i, f_{i+1} in F. Hence, the graph represents subsequent fixations in transition edges between categories, also encoding the time that the transtion takes (the two timestamps t_{2_i} and $t_{1_{i+1}}$ mark the end of the first fixation and the beginning of the following). We further derive the time-aggregated graph $G' = (V,E')$ by adding edges $e' = (c_j, c_k, w_{j,k})$ to E' where weight $w_{j,k}$ is the number of edges connecting the two categories c_j and c_k in E, if $w_{j,k} > 0$.

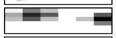
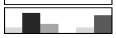
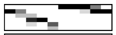
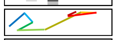
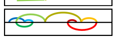
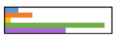
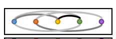
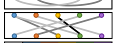
Our solution as outlined in Figure 1 (bottom) is based on representing the transcript in a table, showing one activity per line in chronological order. Besides a column containing the actual transcript text, additional columns provide context about timing, eye tracking, and interaction events that happened 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 graphics 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 The Design Space of 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, 45] (advanced circular layouts are hard to fit to the elongated format of a sparkline), or a specialization to particular kinds of stimuli [5, 48]. As

Table 1 Design space of *word-sized eye tracking visualizations*: data dimensions are mapped to visual attributes of the graphical representations; some visualizations are only defined for coordinate x, but there always exist equivalent ones for coordinate y; variables with index i refer to a specific fixation f_i, whereas variables without an index or with a different index are defined by the interval sizes and categories used for the visualization.

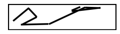
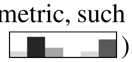
Visualization	*Data*	*Encoding*	*X-Axis*	*Y-Axis*	*Color*	*Ref.*
Point-Based Visualizations						
P1 	Space	Lines	x_i	y_i	–	[24, 38]
P2 	Space	Cells	x	y	Dur. $\Delta(x,y)$	[24, 35]
P3 	Space	Bars	x	Freq. $\sigma(x)$	Dur. $\Delta(x)$	
P4 	Space-time	Cells	Time t	x	Dur. $\Delta(t,x)$	[21, 51]
P5 	Space-time	Lines	x_i	y_i	Time t_{1_i}	
P6 	Space-time	Arcs	x_i	Direct. $x_{i+1} - x_i$	Time t_{1_i}	[13]
AOI-Based Visualizations						
A1 	AOI statistics	Bars	Freq. $\sigma(c)$ or dur. $\Delta(c)$	AOI c	AOI c	
A2 	AOI seq.	Columns	Events i	–	AOI c_i	[24]
A3 	AOI seq.	Boxes	Events i	AOI c_i	AOI c_i	
A4 	AOI seq.	Boxes	Events i	AOI c_i	Dur. t_{d_i}	[41]
A5 	AOI seq.	Boxes	Time t_{1_i}, t_{2_i}	AOI c_i	AOI c_i	[15, 28]
A6 	AOI trans.	Arcs	Trans. e'	Direct. of e'	Weight $w_{j,k}$	[37]
A7 	AOI trans.	Lines	Trans. e'	Direct. of e'	Weight $w_{j,k}$	[12]
A8 	AOI trans.	Cells	AOI c_j	AOI c_k	Weight $w_{j,k}$	[22]

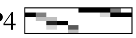
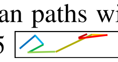
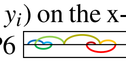
Legend: seq.–sequence; trans.–transition; freq.–frequency; dur.–duration; direct.–direction; ref.–references.

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 based on the formalism introduced in Section 3. 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 prototypical analysis tool (Section 5).

4.1 *Point-Based Visualizations*

Each fixation f_i is assigned a pair of coordinates (x_i, y_i) 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.

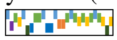
Space. Focusing on the spatial part of the data, the standard representations of eye tracking data are *scan paths* and *heat maps*. Scan path visualizations simply overlay the trajectory of the gaze $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$ onto the stimulus [38], often encoding fixations as circles scaled according to their duration t_{d_i} [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 (x_i, y_i), 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 $\Delta(x, y)$ in the darkness of the grid cells (P2 ). As an alternative, we could focus on only one spatial axis (either x or y), again encode duration (here, either $\Delta(x)$ or $\Delta(y)$) in the color, and use bar charts to encode another metric, such as the frequency of fixations $\sigma(x)$ or $\sigma(y)$ within the respective area (P3 ). Spatial information can also be restricted otherwise to make them representable at small scale, for instance, encoding angles of the trajectory in radial diagrams [21].

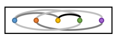
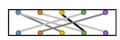
Space and Time. The temporal sequence of fixations is also important for some analysis scenarios. Mapping time t to a spatial dimension, however, requires the encoding of spatial information to be limited [21, 51]. For instance, using the longer x-axis as a timeline, the y-axis could encode one of the spatial coordinates of the fixations (either x or y) while darkness indicates the distribution of fixation durations, here, either $\Delta(t, x)$ or $\Delta(t, y)$ (P4 ). We can also extend scan paths with temporal information 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 (either x_i or y_i) on the x-axis and connecting points with arcs according to observed saccades (P6 )—like in the original approach, arcs are directed 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 visualizations 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.

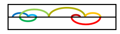
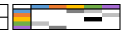
AOI Statistics. One of the simplest AOI-based visualizations is to depict the frequency $\sigma(c)$ or total duration $\Delta(c)$ each AOI c was fixated, for instance, in a line or bar chart. Such diagrams can be directly transferred into word-sized graphics. We decide to use bar charts because lines are harder to perceive if only little space is available. We use the y-axis to distinguish AOIs to have more spatial resolution for reading the value from the x-axis and redundantly color-code the AOI to improve the discernibility of the bars (A1 ).

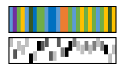
AOI Sequences. The temporal sequence of viewed AOIs reveals, on the one hand, what a participant saw and, on the other hand, in which order. This sequence of AOIs $(c_1, c_2, \ldots, c_n)$ might be visually encoded in any list representation showing, for instance, the logical temporal sequence of events from left to right. This has been done in various eye tracking visualizations, for instance, connecting subsequent AOI fixations by lines and encoding the AOI fixation durations in node sizes [41] or in the horizontal length of a line [28]. In a sparkline, the sequence is easily visualized as a sequence of blocks each representing an AOI event. The different AOIs might be discerned by color (A2 ) or redundantly as a combination of position and color (A3 ). When the duration of each AOI fixation t_{d_i} is important, it can be encoded in the darkness of the boxes if the position encoding is used for discerning AOIs (A4 ) or a linear timeline can be employed scaling the width of the boxes according to the time span from t_{1_i} to t_{2_i} (A5 ) [15, 28].

AOI Transitions. Transitions between AOIs might also be depicted as a graph G' with AOIs as nodes V, and aggregated transition frequencies as weighted links E' [28]. Considering the temporal dimension of the data as well, the aggregated static graph becomes dynamic and might be visualized by animation- or timeline-based dynamic graph visualization approaches [12]. Graphs are, however, difficult to represent as a sparkline because nodes and links require a certain amount of 2D space to be discernible. Arranging the nodes in only one dimension simplifies the problem: like in *ArcTrees* [37], we draw nodes V on a vertical axis connected by arcs according to transitions $e' = (c_j, c_k, w_{j,k})$, having the weight $w_{j,k}$ encoded in the link darkness (A6 ). A more scalable variant is the *Parallel Edge Splatting* approach [14], which was already applied to AOI transitions graphs [12]: the graph is interpreted as a bipartite graph duplicating the nodes V to two horizontal axes; all transitions E' are drawn as straight lines connecting a source AOI c_j at the top to a target AOI c_k at the bottom (A7 ). Furthermore, matrix representations of graphs are space-efficient and have already been employed to represent eye tracking data [22]. A transformation into a sparkline is straightforward, for instance, color-coding the AOIs V (first row and column) in addition to the transition weights $w_{j,k}$ within the matrix cells (A8 ). A limitation, however, is that they are inherently quadratic—although they can be stretched to fill an arbitrary rectangle, additional vertical space does not necessarily improve their readability.

4.3 Combination and Extension

The suggested visualizations provide a flexible framework for encoding eye tracking data. To decide between the different encodings is not an either–or decision because visualizations can be combined with each other to build an even more expressive analysis tool. Moreover, the framework of visualizations might be extended with only little adaption to also depict interaction data.

Juxtaposing Visualizations. Since word-sized visualizations are space-efficient, they can easily be juxtaposed within one line, each graphic providing a different perspective onto the data. For instance, it could be useful to combine a point-based and an AOI-based visualization:  . 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:



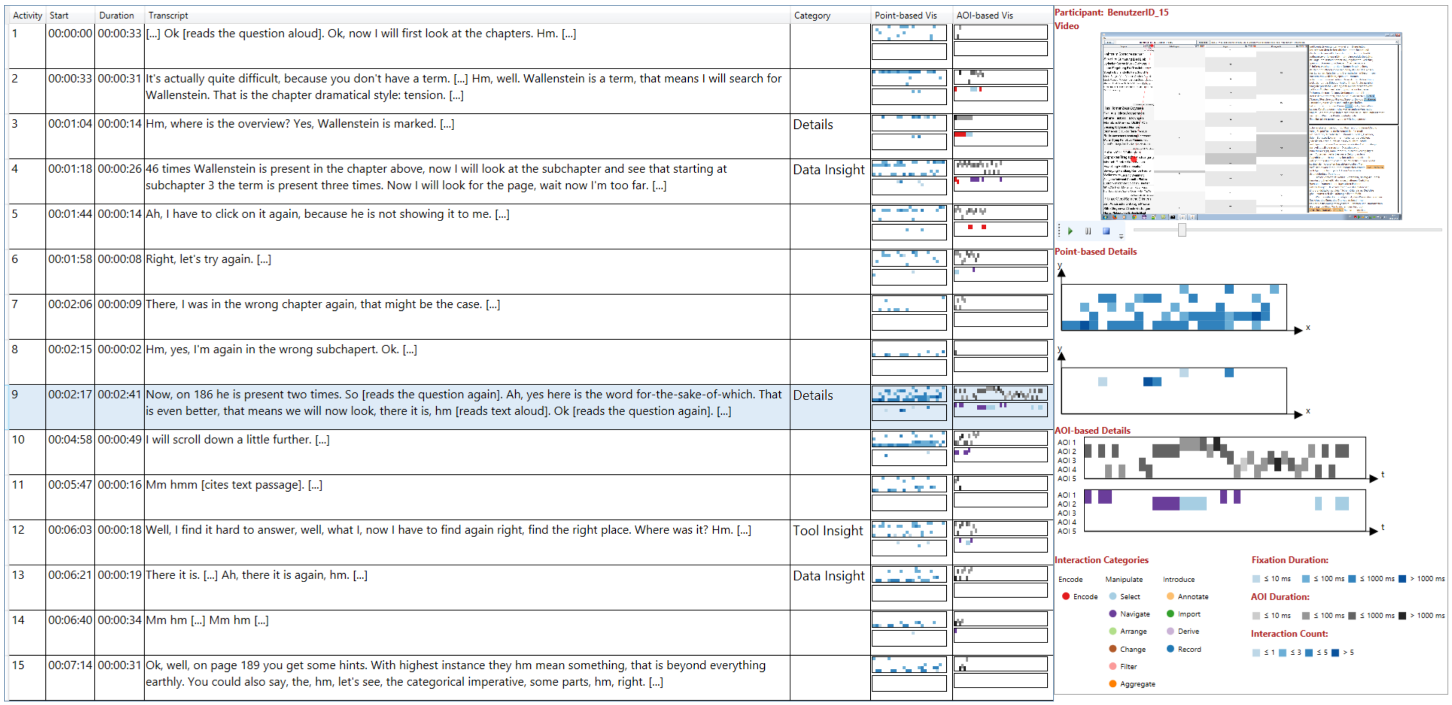
Interaction Data. Interaction data shares characteristics to eye tracking data: Much like fixations, interactions are temporal events on the same experiment time dimension. They can be classified according to their type into categories or assigned to AOIs based on their location. Also, transitions between interactions might be derived from the sequence of logged events. One difference, however, is that typical interaction events do not have a duration; they only get a duration if they are abstracted to longer sequences of semantically linked interactions. Hence, an interaction can be represented as a 5-tupel like a fixation $f_i = (x_i, y_i, t_{1_i}, t_{2_i}, c_i)$, often with $t_{1_i} = t_{2_i}$. 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; even those representing durations are applicable if we just assume constant durations $t_{d_i} > 0$. 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 analysis framework for eye tracking studies [6]. The visual analysis framework is intended to support the joint analysis of eye tracking and interaction data. 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 visualizations 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.

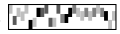
The point-based visualizations are gridded attention maps (Table 1, P2 ) or, respectively, maps showing the spatial distribution of interactions. We divided the stimulus into 25 columns and five rows. For each cell, we counted the fixation durations and the count of interactions and color-coded the cells accordingly. The color



Activity	Start	Duration	Transcript	Category	Point-based Vis	AOI-based Vis
1	00:00:00	00:00:33	[...] Ok [reads the question aloud]. Ok, now I will first look at the chapters. Hm. [...]			
2	00:00:33	00:00:31	It's actually quite difficult, because you don't have a term. [...] Hm, well. Wallenstein is a term, that means I will search for Wallenstein. That is the chapter dramatical style: tension. [...]			
3	00:01:04	00:00:14	Hm, where is the overview? Yes, Wallenstein is marked. [...]	Details		
4	00:01:18	00:00:26	46 times Wallenstein is present in the chapter above, now I will look at the subchapter and see that starting at subchapter 3 the term is present three times. Now I will look for the page, wait now I'm too far. [...]	Data Insight		
5	00:01:44	00:00:14	Ah, I have to click on it again, because he is not showing it to me. [...]			
6	00:01:58	00:00:08	Right, let's try again. [...]			
7	00:02:06	00:00:09	There, I was in the wrong chapter again, that might be the case. [...]			
8	00:02:15	00:00:02	Hm, yes, I'm again in the wrong subchapert. Ok. [...]			
9	00:02:17	00:02:41	Now, on 186 he is present two times. So [reads the question again]. Ah, yes here is the word for-the-sake-of-which. That is even better, that means we will now look, there it is, hm [reads text aloud]. Ok [reads the question again]. [...]	Details		
10	00:04:58	00:00:49	I will scroll down a little further. [...]			
11	00:05:47	00:00:16	Mm hmm [cites text passage]. [...]			
12	00:06:03	00:00:18	Well, I find it hard to answer, well, what I, now I have to find again right, find the right place. Where was it? Hm. [...]	Tool Insight		
13	00:06:21	00:00:19	There it is. [...] Ah, there it is again, hm. [...]	Data Insight		
14	00:06:40	00:00:34	Mm hm [...] Mm hm [...]			
15	00:07:14	00:00:31	Ok, well, on page 189 you get some hints. With highest instance they hm mean something, that is beyond everything earthly. You could also say, the, hm, let's see, the categorical imperative, some parts, hm, right. [...]			

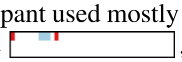
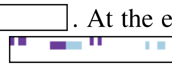
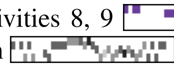
Fig. 2 Screenshot of the prototype implementation of our approach that shows the think-aloud protocol in a tabular fashion, containing an activity ID, start time and duration, a transcript of the audio recording, a category, and point-based as well as AOI-based word-sized visualizations for eye movements (top) and interactions (bottom). The sidebar provides a video replay of an enriched stimulus and enlarged word-sized visualizations of a selected activity.

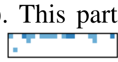
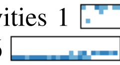
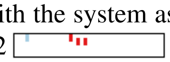
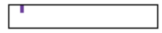
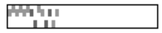
coding was obtained from ColorBrewer [10], using a sequential, single-hue blue color and a gradation of four (fixation duration $\leq$ 10 ms, $\leq$ 100 ms, $\leq$ 1000 ms, and $>$ 1000 ms; interaction count $\leq$ 1, $\leq$ 3, $\leq$ 5, and $>$ 5).

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 gray scale and a logarithmic gradation of four (AOI fixation duration $\leq$ 10 ms, $\leq$ 100 ms, $\leq$ 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 scan path overlay obtained from Tobii Studio. Below the video, the visualizations of a selected row are shown enlarged and annotated with labels.

6 Application Example

For a small use case example, we re-analyzed data from a user study testing a visual text analysis tool [6] (Participant 13 and 15, transcript partly translated from German to English). For Participant 15 (Figure 2), 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 , 3, 4, 5, 6) while focusing mostly on AOI 1 and 2 . At the end, the participant used more *navigate* interactions (Activities 8, 9 , 10, 12) and was looking at AOIs 4 and 5 more often .

We can observe a similar behavior when looking at another participant (Participant 13). This participant also focused on the upper part at the beginning (Activities 1 , 2, 3, 4) and at the lower part at the end (Activitites 12–15, 16 , 17). However, this participant has not interacted with the system as intensely as Participant 15 did, only at the beginning (Activities 1, 2 , 4, 5) and once more at the end (Activity 16 ). In the meantime, this participant was switching focus from top to bottom and looking at AOIs 1, 2, and 5 mainly (Activities 5 , 6, 7, 9, 10).

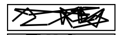
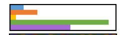
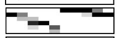
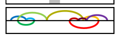
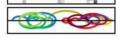
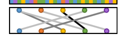
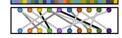
These kinds of analyses allows us to classify the participants' activities and manually assign categories in the respective table column (Fig. 2). To categorize some of the data, we use categories Saraiya et al. [44] define for visualization systems, namely *overview*, *patterns*, *groups*, and *details*. Smuc et al. [47] added the categories *data insights* and *tool insights*. For example, if we want to categorize Activity 9, highlighted in Figure 2, we can look at the transcript and read that the participant found a specific text passage with the word she was searching for and some related terms after a while. Looking at the point-based word-sized graphics of the fixations  shows that the participant was looking at large parts and most AOIs  of the stimulus while using many navigate interactions and at the end she selected some items . This behavior could be classified as *details*, because the participant was inspecting the system in order to find details about the analysis task she was solving. Thus, we can add this category to Activity 9 in the transcript.

7 Discussion

We explored the design space of *word-sized eye tracking visualizations* systematically in the sense that we miniaturized existing eye tracking visualizations following a taxonomy [7]. This strategy provides a variety of meaningful solutions. However, we cannot guarantee that we covered the full available design space: First, we only focused on single-participant visualizations and might have excluded other visualizations that would be well suitable for word-sized representation. Second, we did not analyze the design space for word-sized representations independently of previous normal-sized visualizations—there might be visualizations that make sense as word-sized graphics but would not be used as normal-sized visualizations.

Word-sized visualizations, by design, have only limited space available and hence scale worse to larger data sets. By splitting up the data into short activities and present multiple word-sized visualizations, we circumvent this problem to some extent. Our application example in Section 6 already shows that at least the implemented visualizations scale well to a meaningful amount of data. However, if the activities become longer or if we have to handle more AOIs, visualizations that do not aggregate this information would become cluttered; word-sized graphics only leave few opportunities for improving visual scalability. According to our experience with the current implementation, we estimate that the suggested visualizations scale sufficiently up to 30 events (i.e., fixations and interactions) or 30 time intervals respectively and 10 AOIs per activity. To substantiate this estimation, Table 2 provides some mockup examples showing increasing numbers events for visualizations that do not aggregate time and examples for increasing numbers of AOIs. For the number of events, the number of crossing lines and the width of the visualization

Table 2 Examples of *word-sized eye tracking visualizations* to investigate scalability by number of events/intervals for point-based visualizations and number of AOIs for AOI-based visualizations.

	10 events/intervals	*30 events/intervals*		*5 AOIs*	*10 AOIs*
P1			A1		
P4			A3		
P6			A7		

restrict their scalability. Limiting factors for the number of AOIs is the amount of discernible colors and the height or width of the visualizations.

We studied in detail the embedding of *word-sized eye tracking visualizations* within transcribed experiment recordings of single participants. But beyond that, we see broader opportunities for application of our approach: First, the visualizations can be easily used to compare similar activities of multiple participants by juxtaposing the visualizations as demonstrated in the application example (Section 6). Those visualizations that aggregate time to duration or transitions (Table 1, P2, P3, A1, A6, A7, A8) can even be directly used to summarize data of several participants within a single representation. A second application is the use of our visualizations within scientific texts, for instance, to report the results of a study like demonstrated in Section 6. Since publications are a static medium, however, it is a restriction in this scenario that interactions are not available to view a larger version of the visualization. In general, an open question is still how well the suggested word-sized visualizations would be received by experimenters analyzing their eye tracking data and readers seeing the visualizations within publications.

8 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 a detail view of a larger 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 interaction techniques, 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

1. Baldauf, M., Fröhlich, P., Hutter, S.: KIBITZER: A wearable system for eye-gaze-based mobile urban exploration. In: Proceedings of the 1st Augmented Human International Conference, AH, pp. 9:1–9:5 (2010)
2. Beck, F., Dit, B., Velasco-Madden, J., Weiskopf, D., Poshyvanyk, D.: Rethinking user interfaces for feature location. In: Proceedings of the 23rd IEEE International Conference on Program Comprehension, ICPC, pp. 151–162. IEEE (2015)
3. Beck, F., Koch, S., Weiskopf, D.: Visual analysis and dissemination of scientific literature collections with SurVis. IEEE Transactions on Visualization and Computer Graphics **22**(1), 180–189 (2016)
4. Beck, F., Moseler, O., Diehl, S., Rey, G.D.: In situ understanding of performance bottlenecks through visually augmented code. In: Proceedings of the 21st IEEE International Conference on Program Comprehension, ICPC, pp. 63–72 (2013)
5. Beymer, D., Russell, D.M.: WebGazeAnalyzer: A system for capturing and analyzing web reading behavior using eye gaze. In: CHI '05 Extended Abstracts on Human Factors in Computing Systems, CHI EA, pp. 1913–1916 (2005)
6. Blascheck, T., John, M., Kurzhals, K., Koch, S., Ertl, T.: VA^2: A visual analytics approach for evaluating visual analytics applications. IEEE Transactions on Visualization and Computer Graphics **22**(1), 61–70 (2016)
7. Blascheck, T., Kurzhals, K., Raschke, M., Burch, M., Weiskopf, D., Ertl, T.: State-of-the-art of visualization for eye tracking data. In: EuroVis - STARs, pp. 63–82 (2014)
8. Blascheck, T., Raschke, M., Ertl, T.: Circular heat map transition diagram. In: Proceedings of the 2013 Conference on Eye Tracking South Africa, ETSA, pp. 58–61 (2013)
9. Brandes, U., Nick, B.: Asymmetric relations in longitudinal social networks. IEEE Transactions on Visualization and Computer Graphics **17**(12), 2283–2290 (2011)
10. Brewer, C.A., Harrower, M.: ColorBrewer 2.0. http://www.colorbrewer.org
11. Brugman, H., Russel, A.: Annotating multimedia/multi-modal resources with ELAN. In: Proceedings of the Fourth International Conference on Language Resources and Evaluation, LREC, pp. 2065–2068 (2004)
12. Burch, M., Beck, F., Raschke, M., Blascheck, T., Weiskopf, D.: A dynamic graph visualization perspective on eye movement data. In: Proceedings of the 2014 Symposium on Eye Tracking Research & Applications, ETRA, pp. 151–158 (2014)
13. Burch, M., Schmauder, H., Raschke, M., Weiskopf, D.: Saccade Plots. In: Proceedings of the 2014 Symposium on Eye Tracking Research & Applications, ETRA, pp. 307–310 (2014)
14. Burch, M., Vehlow, C., Beck, F., Diehl, S., Weiskopf, D.: Parallel Edge Splatting for scalable dynamic graph visualization. IEEE Transactions on Visualization and Computer Graphics **17**(12), 2344–2353 (2011)

15. Crowe, E.C., Narayanan, N.H.: Comparing interfaces based on what users watch and do. In: Proceedings of the 2000 Symposium on Eye Tracking Research & Applications, ETRA, pp. 29–36 (2000)
16. Dorr, M., Jarodzka, H., Barth, E.: Space-variant spatio-temporal filtering of video for gaze visualization and perceptual learning. In: Proceedings of the 2010 Symposium on Eye Tracking Research & Applications, ETRA, pp. 307–314 (2010)
17. Dou, W., Jeong, D.H., Stukes, F., Ribarsky, W., Lipford, H., Chang, R.: Recovering reasoning processes from user interactions. IEEE Computer Graphics and Applications **29**(3), 52–61 (2009)
18. Duchowski, A., Medlin, E., Courina, N., Gramopadhye, A., Melloy, B., Nair, S.: 3D eye movement analysis for VR visual inspection training. In: Proceedings of the 2002 Symposium on Eye Tracking Research & Applications, ETRA, pp. 103–155 (2002)
19. Franchak, J.M., Kretch, K.S., Soska, K.C., Babcock, J.S., Adolph, K.E.: Head-mounted eye-tracking of infants' natural interactions: a new method. In: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications, ETRA, pp. 21–27 (2010)
20. Goffin, P., Willett, W., Fekete, J.D., Isenberg, P.: Exploring the placement and design of word-scale visualizations. IEEE Transactions on Visualization and Computer Graphics **20**(12), 2291–2300 (2014)
21. Goldberg, J.H., Helfman, J.I.: Visual scanpath representation. In: Proceedings of the 2010 Symposium on Eye Tracking Research & Applications, ETRA, pp. 203–210 (2010)
22. Goldberg, J.H., Kotval, X.P.: Computer interface evaluation using eye movements: Methods and constructs. International Journal of Industrial Ergonomics **24**, 631–645 (1999)
23. Hlawatsch, M., Burch, M., Beck, F., Freire, J., Weiskopf, D., Silva, C.: Visualizing the evolution of module workflows. In: Proceedings of the International Conference on Information Visualisation, IV, pp. 40–49 (2015)
24. Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., Van de Weijer, J.: Eye Tracking: A Comprehensive Guide to Methods and Measures, 1st edn. Oxford University Press (2011)
25. Holsanova, J.: Picture viewing and picture description: Two windows on the mind. Ph.D. thesis, Lund University (2001)
26. Holsanova, J.: Dynamics of picture viewing and picture description. Advances in Consciousness Research **67**, 235–256 (2006)
27. Hurter, C., Ersoy, O., Fabrikant, S., Klein, T., Telea, A.: Bundled visualization of dynamic graph and trail data. IEEE Transactions on Visualization and Computer Graphics **20**(8), 1141–1157 (2014)
28. Itoh, K., Tanaka, H., Seki, M.: Eye-movement analysis of track monitoring patterns of night train operators: Effects of geographic knowledge and fatigue. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 44, pp. 360–363 (2000)
29. Kipp, M.: ANVIL – A generic annotation tool for multimodal dialogue. In: Proceedings of the 7th European Conference on Speech Communication and Technology, Eurospeech, pp. 1367–1370 (2001)
30. Kipp, M.: Spatiotemporal coding in ANVIL. In: Proceedings of the 6th International Conference on Language Resources and Evaluation, LREC, pp. 2042–2045 (2008)
31. Kipp, M.: ANVIL: The video annotation research tool. In: U.G. Jacques Durand, G. Kristoffersen (eds.) The Oxford Handbook of Corpus Phonology, chap. 21, pp. 420–436. Oxford (2014)
32. Lankford, C.: Gazetracker: Software designed to facilitate eye movement analysis. In: Proceedings of the 2000 Symposium on Eye Tracking Research & Applications, ETRA, pp. 51–55 (2000)
33. Lee, B., Henry Riche, N., Karlson, A.K., Carpendale, S.: SparkClouds: Visualizing trends in tag clouds. IEEE Transactions on Visualization and Computer Graphics **16**(6), 1182–1189 (2010)
34. Lipford, H.R., Stukes, F., Dou, W., Hawkins, M.E., Chang, R.: Helping users recall their reasoning process. In: Proceedings of the 2010 IEEE Symposium on Visual Analytics Science and Technology, VAST, pp. 187–194 (2010)

35. Mackworth, J.F., Mackworth, N.H.: Eye fixations recorded on changing visual scenes by the television eye-marker. Journal of the Optical Society of America **48**(7), 439–444 (1958)
36. Minelli, R., Mocci, A., Lanza, M., Baracchi, L.: Visualizing developer interactions. In: Proceedings of the Second IEEE Working Conference on Software Visualization, VISSOFT, pp. 147–156 (2014)
37. Neumann, P., Schlechtweg, S., Carpendale, S.: ArcTrees: Visualizing relations in hierarchical data. In: Proceedings of the 7th Joint Eurographics / IEEE VGTC Conference on Visualization, EuroVis, pp. 53–60 (2005)
38. Noton, D., Stark, L.: Scanpaths in saccadic eye movements while viewing and recognizing patterns. Vision Research **11**, 929942 (1971)
39. Paletta, L., Santner, K., Fritz, G., Mayer, H., Schrammel, J.: 3D attention: Measurement of visual saliency using eye tracking glasses. In: CHI '13 Extended Abstracts on Human Factors in Computing Systems, CHI EA, pp. 199–204 (2013)
40. Pellacini, F., Lorigo, L., Gay, G.: Visualizing Paths in Context. Tech. Rep. #TR2006-580, Department of Computer Science, Dartmouth College (2006)
41. Räihä, K.J., Aula, A., Majaranta, P., Rantala, H., Koivunen, K.: Static visualization of temporal eye-tracking data. In: M.F. Costabile, F. Paternò (eds.) Human-Computer Interaction – INTERACT 2005, *LNCS*, vol. 3585, pp. 946–949. Springer (2005)
42. Ramloll, R., Trepagnier, C., Sebrechts, M., Beedasy, J.: Gaze data visualization tools: Opportunities and challenges. In: Proceedings of the 8th International Conference on Information Visualization, IV, pp. 173–180 (2004)
43. Reda, K., Johnson, A., Leigh, J., Papke, M.: Evaluating user behavior and strategy during visual exploration. In: Proceedings of the 2014 BELIV Workshop, pp. 70–77 (2014)
44. Saraiya, P., North, C., Duca, K.: An insight-based methodology for evaluating bioinformatics visualizations. IEEE Transactions on Visualization and Computer Graphics **11**(4), 443–456 (2005)
45. Schulz, C.M., Schneider, E., Fritz, L., Vockeroth, J., Hapfelmeier, A., Brandt, T., Kochs, E.F., Schneider, G.: Visual attention of anaesthetists during simulated critical incidents. British Journal of Anaesthesia **106**(6), 807–813 (2011)
46. Sloetjes, H., Wittenburg, P.: Annotation by category – ELAN and ISO DCR. In: Proceedings of the 6th International Conference on Language Resources and Evaluation, LREC, pp. 816–820 (2008)
47. Smuc, M., Mayr, E., Lammarsch, T., Aigner, W., Miksch, S., Gartner, J.: To score or not to score? tripling insights for participatory design. IEEE Computer Graphics & Applications **29**, 29–38 (2009)
48. Špakov, O., Räihä, K.J.: KiEV: A tool for visualization of reading and writing processes in translation of text. In: Proceedings of the 2008 Symposium on Eye Tracking Research & Applications, ETRA, pp. 107–110 (2008)
49. Tufte, E.R.: Beautiful Evidence, 1st edn. Graphics Press (2006)
50. Weibel, N., Fouse, A., Emmenegger, C., Kimmich, S., Hutchins, E.: Let's look at the cockpit: Exploring mobile eye-tracking for observational research on the flight deck. In: Proceedings of the 2012 Symposium on Eye Tracking Research & Applications, ETRA, pp. 107–114 (2012)
51. Yarbus, A.L.: Eye Movements and Vision. Plenum Press (1967)