An Expert Evaluation of Word-Sized Visualizations for Analyzing Eye Movement Data

Fabian Beck[1], Yasett Acurana[2], Tanja Blascheck[1], Rudolf Netzel[1], and Daniel Weiskopf[1]

[1]University of Stuttgart, Germany [2]University of Chile, Chile

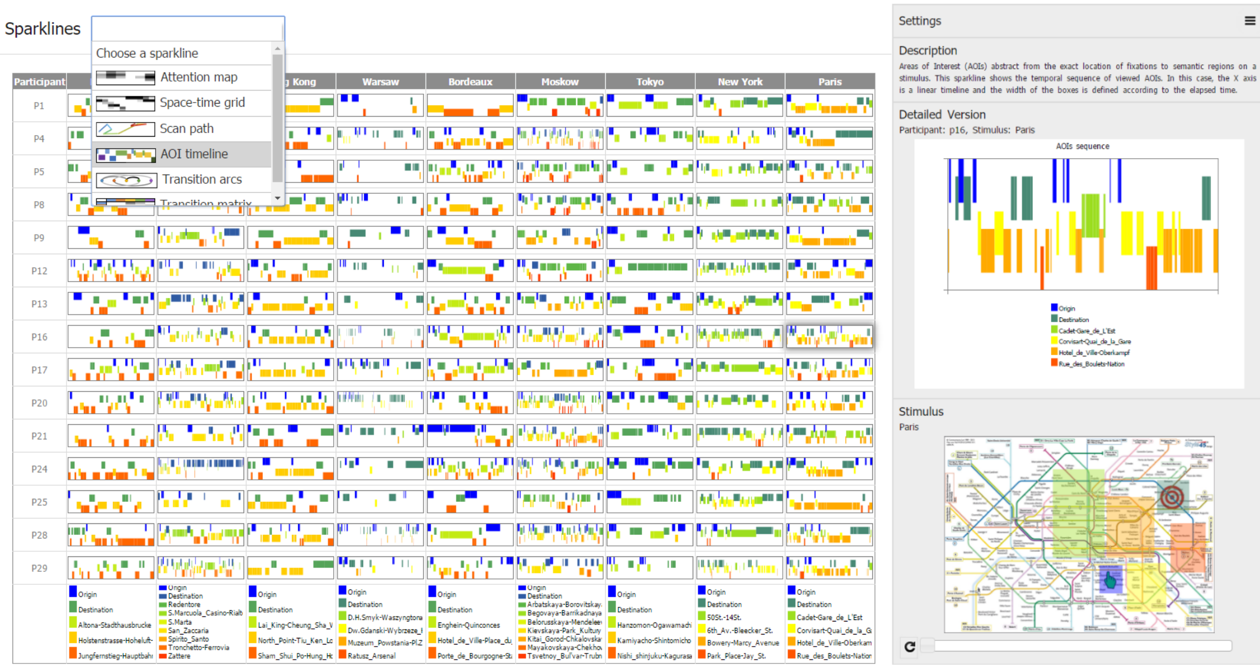


Figure 1: Analysis framework depicting eye movement data for different stimuli (columns) and participants (rows) in a tabular grid as word-sized visualizations. An analyst can choose between different variants of visualizations. On selection of a cell, the framework provides details such as a description, an enlarged version, and the original stimulus.

Abstract

Word-sized visualizations for eye movement data allow analysts to compare a variety of experiment conditions or participants at the same time. We implemented a set of such word-sized visualizations as part of an analysis framework. We want to find out which of the visualizations is most suitable for different analysis tasks. To this end, we applied the framework to data from an eye tracking study on the reading behavior of users studying metro maps. In an expert evaluation with five analysts, we identified distinguishing characteristics of the different word-sized visualizations.

1 Introduction

Visually analyzing eye movement data, an analyst can pick from a diverse set of visualizations [6] to study fixation distribution, areas of interest (AOIs), temporal events, etc. When comparing multiple participants or repetitions based on the same or similar task, an analyst might juxtapose two or more visualizations. When scaling down the visualizations to the size of a word, the visual comparison can be scaled up to dozens of different conditions and can be integrated into natural language text or other user interfaces.

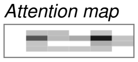
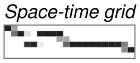
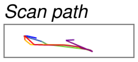
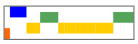
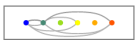
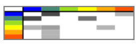
In this work, we target at evaluating word-sized visualizations to find suitable representations for different eye tracking data analysis tasks. To this end, we implemented previously proposed *word-sized eye tracking visualizations* [1] in a small-multiples approach. Like Figure 1 illustrates, we list participants as rows of a table and different stimuli as columns. Each cell shows the respective eye tracking data in a word-sized visualization, which is enlarged on demand in a side-bar. The main contributions of this paper are:

- to implement and make available previously suggested *word-sized eye tracking visualizations* [1] as part of a Web-based visual analysis framework (Section 3),[1]
- to apply these visualizations for partly revisiting an experiment on the readability of metro maps [20] (Section 4), and
- to conduct an expert evaluation comparing these visualizations regarding different usage scenarios (Section 5).

We start with briefly discussing related work (Section 2) and conclude the paper by summarizing results (Section 6).

[1]`https://github.com/Yasett/ETSparklines`

Table 1: Selected *word-sized eye tracking visualizations*.

Point-Based Visualizations	
Attention map 	A gridded attention map that aggregates fixations and encodes the duration in the darkness of the cells.
Space-time grid 	The X axis represents a timeline, the Y axis shows X or Y coordinates of fixations; darkness indicates fixation durations.
Scan path 	Path of eye movement with temporal information represented by the edge colors from blue to purple.
AOI-Based Visualizations	
AOI timeline 	Temporal sequence of viewed AOIs encoded in color and position, width of the boxes is defined according to the elapsed time.
Transition arcs 	Transitions between AOIs (nodes) and aggregated transitions as weighted links, where the darkness encodes frequency.
Transition matrix 	A matrix of transitions between AOIs, where the darkness of each box encodes the frequency of the transition.

2 Related Work

Visualizations of eye movement data are various, as Blascheck et al. [6] surveyed. Techniques comparing eye movement data of multiple participants typically depict participants in a sequential order [5, 9, 11, 15, 17]. These eye tracking visualizations usually represent either point-based (i.e., fixation data) [9, 19, 21] or AOI-based information [10, 11, 12, 13, 22]. We use a variety of point-based and AOI-based representations depicting them in a matrix layout to compare multiple participants and stimuli.

Word-sized visualizations or *sparklines* [23] are representations scaled to the height of a line of text. Thus, they can easily be integrated into a text or other visual representations to convey further information. Sparklines have been explored for natural language text [8, 23], visualizations [7, 18], user interfaces [2, 3], or source code [2, 4]. We apply them in a tabular representation.

3 Word-Sized Eye Tracking Visualizations

Beck et al. [1] discussed existing eye movement visualizations and how they can be represented as word-sized visualizations. We used these *word-sized eye tracking visualizations* in this paper and implemented all suggested fourteen variants. Since some of the original visualizations are similar to each other, we selected a subset of six for the analysis framework and evaluation described in the following. To make a representative selection that works well together with the studied stimuli, four of the authors rated each visualization—we took each the best three point-based and AOI-based visualizations, which are depicted and briefly explained in Table 1. Some of the visualizations are configurable if they only show either the X or Y coordinate of the data.

We developed a Web-based analysis framework integrating word-sized visualizations into a tabular representation (Figure 1). In contrast to previous work that integrates word-sized graphics with other data recorded during an experiment [1], we focus here on the effective representation of eye movement data. The cells of the table each show a word-sized visualization for a specific stimulus (column) and participant (row). We assume that all data refers to the same task per stimulus, which makes results comparable across participants. The type of word-sized visualization can be selected from a combo box at the top of the interface. When clicking on a specific instance of a visualization in the table, details are provided such as a description of the visualization, an enlarged and labeled version of the visualization, and a zoomable representation of the stimulus overlaid with AOI information.

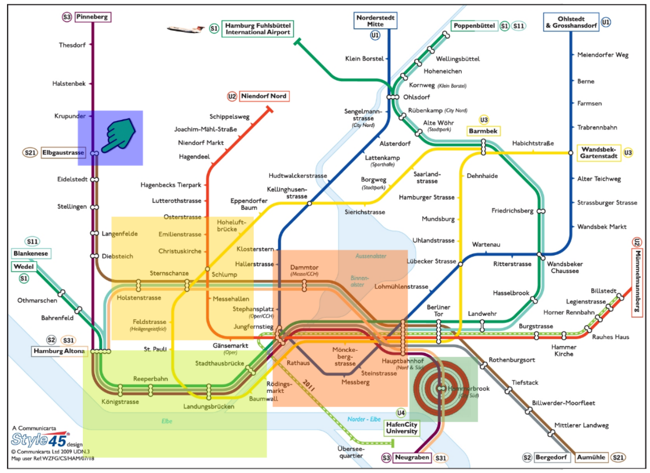
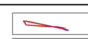
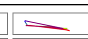
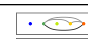
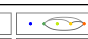
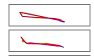
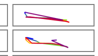
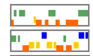
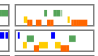
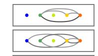
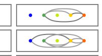


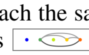
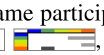
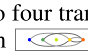
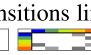
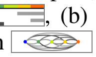
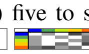
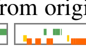
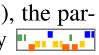
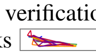
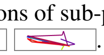
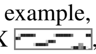
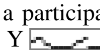
Figure 2: Metro map of *Hamburg*. Blue and green AOI rectangles indicate origin and destination of a task, while the other three AOIs are possible paths a participants could have traced.

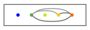
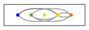
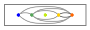
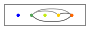
Table 2: Identified groups of participants (a–c) applying different search strategies for stimulus *Hamburg*; three different visualizations show each the same two users per group (left and right sparkline).

Group	*Scan path*		*AOI timeline*		*Transition arcs*	
(a)						
(b)						
(c)						

4 Application Example

We demonstrate our approach with eye tracking data obtained in a controlled experiment with 40 participants about reading behavior of metro maps [20]. Participants were asked to find a route from an origin to a destination and state the number of transitions they needed. All maps had the same design but differed in color encoding (color vs. gray-scale) and complexity measured by the number of nodes. During the experiment, 24 metro maps from all over the world were used. The statistical evaluation showed that participants performed better while using colored maps, therefore, we only investigate data collected of colored maps using word-sized visualizations. We define AOIs for origin, destination, and for each possible correct route by defining areas that need to be crossed. Figure 2 shows a sample stimulus with marked origin (green hand), destination (red target), and AOIs (colored rectangles).

With the help of the visualizations, we separate participants into different behavioral groups. For the metro map of *Hamburg* (Figure 2), we identify three categories of transition complexities in the transition graph (here showing *transition arcs* and *transition matrix* for each the same participant): (a) three to four transitions links of AOIs  , (b) five to seven  , and (c) more than seven  . Table 2 shows samples for these three groups with different types of visualizations. Participants in Group (a) found a path from origin to destination quickly and did not verify it fully  . In Group (b), the participants performed a final verification more thoroughly  . Finally, in Group (c) there are more complex patterns also involving verifications of sub-paths as well as tracing and checking side tracks  . We also observe jumping back and forth to double-check solutions or trying to find alternative routes in *space-time grid* visualizations, for example, a participant who jumped back to the origin two times (X , Y ).

Furthermore, we can infer that some paths were not considered or other paths were used by most of the participants. It is easy to see which AOIs participants visited, for instance, in the *transition arcs* visualization. In the example of *Hamburg*, 65% of the participants did not look at the bottom left AOI (no arcs to the green AOI, e.g.,   ). These participants did not find that there is a line through this AOI directly connecting origin and destination. Also, the *attention map* visualizations allow for discerning the users who noted the direct connection  from those who did not . This is in accordance with the results of Huang et al. [14] that people have a geodesic-path tendency while solving such a task, hence, tend to choose a path that is closer to the direct connection between two locations. Another surprising outcome is that about 40% of the participants did not look at the origin AOI (e.g., ). Reasons for this surprising result could be that participants noted the big hand symbol only through their peripheral vision, that we defined AOIs too small, or that the recorded eye movement data is not accurate enough.

5 Expert Evaluation

We conducted an expert evaluation to compare the usefulness and effectiveness of the different types of word-sized visualizations. In particular, we wanted to learn which visualizations are most suitable for which tasks and why. We invited five experts to participate in a semi-structured user study. The experts worked with our framework to re-analyze the data from the eye tracking study described above. We designed the evaluation to last about 60 minutes per participant. All questionnaires and results are available as supplemental material.[2]

5.1 Participants and Method

All participants were researchers affiliated with the University of Stuttgart, including four PhD students and one postdoc (not involving any co-authors of this paper). Based on self-assessment, all researchers considered their expertise as *knowledgeable* or *expert* in at least two of the three areas: *visualization*, *eye tracking*, *human–computer interaction* (available options: *no knowledge*, *passing knowledge*, *knowledgeable*, *expert*). We refer to them as Expert 1–5 (E1–E5) in the following.

At the beginning of the evaluation and after the self-assessment of expertise, the participants were provided with a short tutorial explaining the visualized data, the analysis framework, and the visualizations. The tutorial included an oral introduction and a brief tool demo using a small sample of three stimuli (disjoint from those used in the later main phase).[3] Then, the experts were allowed to familiarize themselves with the tool for a few minutes. In this introduction and during the whole evaluation, experts were allowed to ask questions at any time. Throughout the experiment, we used a setup with two monitors, one showing the visual analysis framework, the second one showing the required questionnaires. While filling in the questionnaires, the experts could go back to the tool.

For the main part of the study, the experts re-analyzed data from the described eye tracking experiment based on nine stimuli.[4] The stimuli were selected to cover all levels of complexity and difficulty as defined in the original study [20]. Their task was to investigate the data and, for instance, look for "common patterns, outliers, surprising results, search strategies, clusters, etc." For each visualization type, three open-ended questions asked them to describe (i) their findings, (ii) what they liked about the visualization, and (iii) what they disliked and would suggest as improvements. We asked them to proceed after answering all questions or if they spent more than ten minutes with one visualization. We randomized the order of visualizations for every expert.

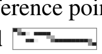
[2]`http://etsparklines.fbeck.com/etvis16-suppl.zip`
[3]`http://etsparklines.fbeck.com/pre_study.html`
[4]`http://etsparklines.fbeck.com/main_study.html`

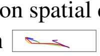
At the end, we asked the experts to rate the six visualization types according to different tasks. Kurzhals et al. [16] presented a taxonomy of analysis tasks applied to eye movement data. We use their high-level classification of tasks for our evaluation, in particular, the three main types of tasks: *where* (space-related tasks, e.g., "Where did a participant look?", "What information did a participant see?", "How long did a participant look at a specific position?"), *when* (time-related tasks, e.g., "When was something investigated?", "When did a fixation start/end?", "When did viewing behavior change?"), and *who* (participant-related tasks, e.g., "Who showed a certain viewing behavior?", "Who had a different/similar viewing behavior?", "How many participants had a similar viewing behavior?"). Moreover, the experts had the opportunity to give other comments in a text field.

5.2 Results

We first summarize the experts' comments regarding the individual visualization types and then evaluate the comparative ratings provided in the final questionnaire. The summaries focus on describing central findings and statements, but we could not include all comments of the experts.

Attention map  – *Findings:* E2, E3, E4, E5 described locations that received high attention. E3, E4, E5 noted special properties in the spatial distribution of attention. *Positive:* E2, E3, E4, E5 highlighted that the visualization showed well how long participants looked at a certain location. E1, E2, E5 considered the visualization to be easy to understand or use. *Negative:* E2, E5 missed temporal information. E1, E3, E5 liked to see the mapping to the stimulus. *Improvements:* E1, E3, E5 suggested showing the stimulus in the background of the visualization or (E5) at least origin and destination as reference points.

Space-time grid  – *Findings:* E2, E4 identified search strategies of participants and E2 analyzed their task completion time. E3, E5 noted relative spatial locations of origin and destination. *Positive:* E2, E3 highlighted the effectiveness to find spatio-temporal patterns simplified to one spatial dimension. E4 remarked that the word-sized visualization was already sufficient and the enlarged version was not needed. *Negative:* E1, E3, E5 found the visualization misleading because of the mapping of dimensions and E2, E5 criticized spatial patterns were obfuscated because either the X- or Y-dimension was shown. *Improvements:* E1, E2 suggested using the Y-axis for time if the visualization is intended to show the X-dimension spatial data.

Scan path  – *Findings:* E1, E2, E4 identified points of interest and E5 spatial patterns. E2, E4, E5 noted specific search strategies. E1 found stimuli with fewer distractors. *Positive:* E1, E5 found it easy to retrieve spatial positions. E2 highlighted that one easily saw search strategies and inefficient participants. E4 appreciated to see both space and time in one visualization. *Negative:* E1 noticed cluttered lines. E2, E4, E5 criticized the color coding, that it did not work for complex lines (E2, E4) and that the direction was hard to interpret (E5). *Improvements:* E2 suggested combining close fixations. E5 wanted to use a different color map and mark origin and destination in the visualization. E5 also recommended an interactive restriction of the displayed time span.

AOI timeline  – *Findings:* E2, E3, E4 observed specific search strategies and patterns. E2, E5 identified AOIs that attracted considerable attention and E1, E4, E5 AOIs that received no attention. *Positive:* E1 found the visualization to be "informative", E3 "easy to interpret", E4 "very useful", and E5 to provide "good overview". E2 appreciated that the AOIs abstracted from actual spatial coordinates. *Negative:* E1, E2, E3 did not appreciate the redundancy in encoding AOIs in position and color. E5 missed spatial information that would allow reconstructing the scan path. *Improvements:* E3 suggested making the AOIs fill the full height of

Table 3: Expert ratings of *word-sized eye tracking visualizations* with respect to different groups of tasks.

Task						
Where?	++	+	++	∘	−−	−−
When?	−−	+	+	++	−−	−−
Who?	++	+	++	++	∘	∘
Overall	+	∘	+	++	−	∘

the visualization and E5 suggested making the AOIs selectable and highlighting the selected one.

Transition arcs – *Findings:* E2, E3, E5 noted different densities of graphs. E4 further observed that participants transitioned much between intermediate AOIs. E5 spotted an outlier without any transitions. *Positive:* E1, E5 positively mentioned the arcs to represent transitions and E4 considered the encoding of frequencies as an advantage. *Negative:* E4, E5 noted that the exact temporal order of transitions was lost. E2 wished an indication of edge direction, but maybe missed that the direction of links was encoded in their placement (top: left to right; bottom: right to left). E1, E5 perceived some examples as cluttered. *Improvements:* E1, E2, E3, E5 suggested integrating or linking the visualization with spatial information. E5 recommended scaling the AOI nodes according to dwell times and using color instead of a gray scale for encoding frequencies.

Transition matrix – *Findings:* E3 observed different densities of graphs. E3, E4, E5 noted specific transition patterns and frequencies. *Positive:* E2 found it easy to identify similar search strategies, E3 to read the individual transitions, E4 to get an overview of transition distributions, and E5 to compare transition frequencies and number of transitions. *Negative:* E1 had problems understanding the visualization, E2 said it was "hard to interpret". E2, E4 missed information on timing. E5 found the colors on the axes distracting and the visualization not scalable for a larger number of AOIs. *Improvements:* E3 suggested interactively linking the visualizations. E5 would have preferred to use other identifiers at the axes and color to encode transition frequencies.

Ratings – The final comparative ratings provide a relative judgment of the visualization types. The results summarized in Table 3 are (mathematically) rounded average ratings on a 5-point scale from (−−) *not useful at all* to (++) *very useful*. We observe clear differences in the experts' ratings between visualization types: Overall, the *AOI timeline* performed best (++), followed by the *attention map* and the *scan path* (+). No visualization received the highest rating (++) in all tasks. We note, for instance, that the *attention map* and the *scan path* seem to be suitable for space- and participant-related tasks (*when* and *who*). An *AOI timeline*, by contrast, could much better reveal time-related aspects (*when*), but also supports participant-related tasks (*who*). The *space-time grid* received average ratings with respect to all tasks (+ for each task, ∘ overall). The two transition graph representations might only be suitable to some extent (∘) for *who* questions, but not for *where* and *when* questions (−−).

As further comments, E3 and E5 suggested a stronger link between the visualizations and stimuli, for instance, by interactive selection. E5 added that the special aspect ratio we apply to the visualizations to make them word-sized was distracting, particularly if the stimulus had a totally different aspect ratio. Visualizations that show multiple participants in one representation and ordering of participants could further improve the analysis framework (E5).

5.3 Threats to Validity

The results are based on subjective ratings of the experts and subjective interpretation of the authors. By choosing experts instead of arbitrary users, we tried to counterbalance to some extent: we believe that these experts could make quite informed, objective judgments on the provided visualizations. In general, as a qualitative analysis with few users, the evaluation is constrained to qualitative findings and cannot make any judgment on quantitative measures such as accuracy or efficiency of usage of the different visualizations. Moreover, we only compared the different visualization with one data set as part of a specific analysis framework. Other data sets, a different embedding, or a comparison to other visualizations could influence the results.

5.4 Discussion and Guidelines

Based on the results from the application example and expert evaluation, we discuss our findings and condense them into guidelines. The tested visualizations helped identify diverse information from eye movement data, for instance, spatial distribution of attention, search strategies (spatio-temporal patterns), groups of participants, transition patterns, or outliers. These could not be found with a single visualization technique, but only with a set of complementing ones, covering spatial point-based data, temporal information, AOIs, and AOI transitions. All visualizations showed unique advantages and disadvantages.

> **Guideline 1** – If possible, use a diverse set of visualizations.

Sometimes, however, there is only little space, like in the cells of our tabular representation. Then, the user has to switch between visualizations to get the full picture. Hence, we formulate the following guidelines based on the comparative ratings:

> **Guideline 2** – To cover all main tasks (*where*, *when*, and *who*), use an *AOI timeline* plus an *attention map* or a *scan path* visualization.

> **Guideline 3** – If you need to focus on one visualization, use a *scan path* or an *AOI timeline* visualization.

Although the ratings of the transition graph representations are quite low, the application example showed that there are scenarios where they provide relevant insights. The low ratings, also for the *space-time grid* visualization, might partly be influenced by misinterpretations of some users. Hence, it would be important to first train users how to read these visualizations.

6 Conclusion

Our application example and expert evaluation contrast previously suggested word-sized visualizations of eye movement data. The results suggest guidelines which visualizations can be used for different analysis tasks. The introduced Web-based analysis framework brings these visualizations into practical application.

Acknowledgements

We thank all participants of the expert evaluation for their time and valuable feedback. 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. We thank the German Research Foundation (DFG) for financial support within project B01 of SFB/Transregio 161. Metro maps are designed by Robin Woods (www.robinworldwide.com) and licensed from Communicarta Ltd (www.communicarta.com).

References

[1] F. Beck, T. Blascheck, T. Ertl, and D. Weiskopf. Word-sized eye tracking visualizations. In *Extended Papers of ETVIS 2015 (to appear)*. 2016.

[2] F. Beck, B. Dit, J. Velasco-Madden, D. Weiskopf, and D. Poshyvanyk. Rethinking user interfaces for feature location. In *Proceedings of the 23rd IEEE International Conference on Program Comprehension*, ICPC, pages 151–162. IEEE, 2015.

[3] F. Beck, S. Koch, and D. Weiskopf. Visual analysis and dissemination of scientific literature collections with SurVis. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):180–189, 2016.

[4] F. Beck, O. Moseler, S. Diehl, and G. D. Rey. In situ understanding of performance bottlenecks through visually augmented code. In *Proceedings of the 21st IEEE International Conference on Program Comprehension*, ICPC, pages 63–72, 2013.

[5] T. Blascheck, M. John, K. Kurzhals, S. Koch, and T. Ertl. VA^2: A visual analytics approach for evaluating visual analytics applications. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):61–70, 2016.

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

[7] U. Brandes and B. Nick. Asymmetric relations in longitudinal social networks. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2283–2290, 2011.

[8] P. Goffin, W. Willett, J.-D. Fekete, and P. Isenberg. Exploring the placement and design of word-scale visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2291–2300, 2014.

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

[10] J. H. Goldberg and X. P. Kotval. Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics*, 24:631–645, 1999.

[11] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press, 1st edition, 2011.

[12] J. Holsanova. *Picture viewing and picture description: Two windows on the mind.* PhD thesis, Lund University, 2001.

[13] J. Holsanova. Dynamics of picture viewing and picture description. *Advances in Consciousness Research*, 67:235–256, 2006.

[14] W. Huang, P. Eades, and S.-H. Hong. A graph reading behavior: Geodesic-path tendency. In *Proceedings of the IEEE Pacific Visualization Symposium*, pages 137–144, 2009.

[15] K. Kurzhals, C. F. Bopp, J. Bässler, F. Ebinger, and D. Weiskopf. Benchmark data for evaluating visualization and analysis techniques for eye tracking for video stimuli. In *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, BELIV, pages 54–60, 2014.

[16] K. Kurzhals, M. Burch, T. Blascheck, G. Andrienko, N. Andrienko, and D. Weiskopf. A task-based view on the visual analysis of eye tracking data. In *Extended Papers of ETVIS 2015 (to appear)*. 2016.

[17] K. Kurzhals, M. Hlawatsch, F. Heimerl, M. Burch, T. Ertl, and D. Weiskopf. Gaze Stripes: Image-based visualization of eye tracking data. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):1005–1014, 2016.

[18] B. Lee, N. Henry Riche, A. K. Karlson, and S. Carpendale. SparkClouds: Visualizing trends in tag clouds. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1182–1189, 2010.

[19] J. F. Mackworth and N. H. Mackworth. Eye fixations recorded on changing visual scenes by the television eye-marker. *Journal of the Optical Society of America*, 48(7):439–444, 1958.

[20] R. Netzel, B. Ohlhausen, K. Kurzhals, R. Woods, M. Burch, and D. Weiskopf. User performance and reading strategies for metro maps: An eye tracking study. *Spatial Cognition and Computation: An Interdisciplinary Journal*, To appear (cf. supplemental material), 2016.

[21] D. Noton and L. Stark. Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research*, 11:929942, 1971.

[22] K.-J. Räihä, A. Aula, P. Majaranta, H. Rantala, and K. Koivunen. Static visualization of temporal eye-tracking data. In M. F. Costabile and F. Paternò, editors, *Human-Computer Interaction – INTERACT 2005*, volume 3585 of *LNCS*, pages 946–949. Springer, 2005.

[23] E. R. Tufte. *Beautiful Evidence*. Graphics Press, 1st edition, 2006.