Int. J. Human-Computer Studies 96 (2016) 54–66

Contents lists available at ScienceDirect

Int. J. Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Using lag-sequential analysis for understanding interaction sequences in visualizations

Margit Pohl [a,*], Günter Wallner [b], Simone Kriglstein [a]

[a] Institute for Design and Assessment of Technology, Vienna University of Technology, Argentinierstrasse 8, 1040 Vienna, Austria
[b] Institute of Art and Technology, University of Applied Arts Vienna, Oskar Kokoschka Platz 2, 1010 Vienna, Austria

ARTICLE INFO

Article history:
Received 26 July 2015
Received in revised form
8 July 2016
Accepted 13 July 2016
Available online 30 July 2016

Keywords:
Interaction sequences
Lag-sequential analysis
Visualization
Log files
Thinking aloud

ABSTRACT

The investigation of how users make sense of the data provided by information systems is very important for human computer interaction. In this context, understanding the interaction processes of users plays an important role. The analysis of interaction sequences, for example, can provide a deeper understanding about how users solve problems. In this paper we present an analysis of sequences of interactions within a visualization system and compare the results to previous research. We used log file analysis and thinking aloud as methods. There is some indication based on log file analysis that there are interaction patterns which can be generalized. Thinking aloud indicates that some cognitive processes occur together with a higher probability than others.

1. Introduction

Understanding the sequential structure of interaction with software artifacts can contribute to a deeper understanding of how humans make use of technology and adapt to different situations (see, e.g., Sanderson and Fisher, 1994; Olson et al., 1994). In this regard sequential analysis methods provide a means to determine whether and how events unfolding in time are related to each other (cf. McComas et al., 2009). On the one hand, such information can help to detect usability issues. For example, if there is a recommended sequence of activities to achieve a goal then it would be interesting to identify deviations from this procedure. Such deviations might be indicators of usability problems. On the other hand, the analysis of interaction sequences can provide more information about how users solve problems with software. Nevertheless, the processes of insight generation are still not very well understood. Pike et al. (2009) therefore argue that sequences of interaction steps should be mapped to cognitive events. Such a mapping could clarify processes of insight generation which is relevant to know for designing user-centered and problem driven systems. However, we still do not know very much about what the users' activities mean and how they are related to their cognitive processes, although, for example, the theory of distributed cognition (Hollan et al., 2000) can help to clarify issues in this context. Many research methods such as interviews or questionnaires, or measurement of time and error are not appropriate for the investigation of interaction sequences because they cannot represent the chronology of the users' activities (Pohl, 2012). The methods which are most appropriate to capture interaction sequences are observation, log files, thinking aloud, and eye-tracking (cf. Pohl, 2012).

Because of the emphasis on analytical activities such aspects are, for instance, highly relevant for visualizations. A user-centered design can ensure the utility and usability of such systems to support users to understand the structure of as well as the interconnections between the data. Furthermore, it has been argued that the interaction with visualization systems can be conceptualized as problem solving activity (Green et al., 2008) and hence the analysis of interaction processes is considered an important topic in the information visualization community (cf. Pike et al., 2009). Therefore, theories of reasoning might be a valuable source for the development of visualizations.

The goal of this paper is to contribute to the clarification of some of the issues pointed out above. In particular our contributions are:

- We aim to advance the understanding of users' interaction patterns with visualizations systems. To this end, we present a study which draws upon theories of graph comprehension and

* Corresponding author.
E-mail addresses: margit@igw.tuwien.ac.at (M. Pohl), guenter.wallner@uni-ak.ac.at (G. Wallner), simone.kriglstein@tuwien.ac.at (S. Kriglstein).

http://dx.doi.org/10.1016/j.ijhcs.2016.07.006
1071-5819/

distributed cognition and which utilizes data gathered via log files and thinking aloud to identify common patterns of behavior within a visualization system for exploring contingency tables. Results of the study are also related to results from previous studies to find out whether it is possible to generalize the results of this research.

- We propose the use of transitional probabilities and lag-sequential analysis (LSA) (Bakeman and Gottman, 1997) – a statistical technique well suited for analyzing event-based sequential data and which is commonly used in behavioral sciences – for analyzing the interaction streams gathered through log files and thinking aloud.

Although we focus on visualizations in this paper we think that – due to the emphasis on reasoning processes – this research is also relevant for human computer interaction (HCI) in general. Designing systems so that users can fluidly interact with them is important for all complex information processing tasks, not only for the usage of visualizations. Interacting with, for instance, the web or with spreadsheet systems may also constitute a complex task which encompasses sequences of various interactions. These systems can only be useful when the right interaction processes are supported by the system and when these interactions are easy to use and well designed.

2. Related work

There are several theories of human reasoning which may be relevant for the explanation of the interaction of users with visualization systems (for an overview see, e.g., Pohl et al., 2012a). We found two of these approaches especially relevant for our empirical research: graph comprehension and distributed cognition. In general, graph comprehension theories are more concerned with cognitive processes going on inside people's minds, whereas distributed cognition emphasizes the interaction of the users with artifacts and the influence that the design of an artifact has on the solutions reached. As a consequence, different methodologies are used for investigations in the context of these two approaches.

Graph comprehension: Graph comprehension is a set of theories which try to describe and explain how people make sense of graphs (Friel et al., 2001). In this context, the term graph is defined as a graphical representation of data, as, for example, in bar charts or line charts. This definition differs from the definition of graphs in graph theory which describes graphs as consisting of nodes (vertices) and links (edges). The term as defined in graph comprehension is much wider in scope than in graph theory. Graph comprehension theory describes how inferences are drawn from simple diagrams. It is typically based on perceptual principles and focuses on less complex graphs like line plots and bar charts, but in recent years more complex visualizations have also been investigated. An important aspect of graph comprehension is that viewers develop a mental model based on several cycles of examining the graphs in which important variables and relations between these variables are identified. In the context of the theory of graph comprehension a considerable amount of research has been conducted to identify characteristics of this process. One major aspect of this research has been to clarify whether viewers only get a superficial impression of the meaning of a graph or if they get a deeper understanding which enables them to go beyond the data and predict developments based on the information from the graph (Tversky, 2005). Friel et al. (2001), for example, reviewed several models on graph comprehension and, based on that, distinguish between three different levels of graph comprehension: (1) reading the data (i.e., extracting data and locating data points), (2) reading between the data (i.e., finding connections between data), and (3) going beyond the data (i.e., making inferences).

Ratwani et al. (2008) also investigated the main issue of distinguishing between activities which aim at reading of data and, on the other hand, making inferences. They conducted a thinking aloud study and analyzed the protocols with the following coding scheme: extraction of quantitative and qualitative data; searching for specific objects; making inferences; making comparisons between components of the graph. Transition probability matrices of activities were utilized to analyze the data. These matrices indicate the probability with which one activity of a certain category follows another activity. Subjects had to solve two types of tasks: (a) tasks with the goal to extract single values and (b) integration tasks (basically inferences and comparisons). They assumed that solving the extraction tasks would be straightforward, while the integration tasks would contain repeated cycles of the same sequence of activities because going beyond the data is necessary in this context. They found evidence for this assumption. This study by Ratwani et al. (2008) influenced the thinking aloud study described in this paper.

Trafton et al. (2000) also conducted a thinking aloud study to analyze a graph comprehension task. They distinguished between the categories goal (talking about the goal of the task), extracting quantitative and qualitative values, inference, and brief writing. Based on this approach, Trickett and Trafton (2006) developed a more comprehensive model of graph comprehension. They argue that in many cases users are not able to directly extract information from a graphical representation. In such situations, users adopt spatial transformation as a strategy to get insights into the data. These spatial transformations are mental operations, in contrast to physical operations, for example, on a computer screen. Trickett and Trafton (2006) argue, based on their empirical research, that they found significantly more spatial transformations than physical transformations. This implies that even if an interactive visualization is available, people will still conduct spatial transformations in their minds in many cases.

Distributed cognition: One of the main reasons why distributed cognition might explain the interaction of users with software artifacts is the fact that distributed cognition takes the artifact itself into account (Hollan et al., 2000; Hutchins, 1995; Kirsh and Maglio, 1994). The interaction between the user and the artifact plays an important role in the context of this theory. Cognitive processes are shaped by the design of the artifacts. Liu et al. (2008) describe how distributed cognition could be applied as a theoretical foundation for the explanation of the interaction with visualizations. They argue that the form of the representation may evoke different solution strategies because of different affordances of the representations. Kirsh (2010) notes that the utility of external representations is not only due to the fact that they lighten the cognitive load of human memory. He describes seven additional advantages external representations might have. Andrews and North (2012) point out that distributed cognition emphasizes the importance of observing the interaction processes of the users with visualization systems. These interaction processes allow researchers to make inferences on the nature of the underlying cognitive processes as interaction and cognitive processes are tightly linked.

Interaction provenance: The analysis of users' interactions with a system has also become prominent in the field of analytical provenance (cf. Dou et al., 2009; Groth and Streefkerk, 2006; Gotz and Zhou, 2009; Sun et al., 2013; Walker et al., 2013; Xu et al., 2015) which focuses *on understanding a user's reasoning process through the study of their interactions with a visualization* (North et al., 2011). In this context several evaluation studies (e.g., Dou et al., 2009; Gotz and Zhou, 2009; Lipford et al., 2010) were conducted to understand a user's visual analytic behavior. For example, Dou et al. (2009) investigated how much of the analyst's reasoning process can be recovered by only using the captured interactions between an analyst and the system. Brown et al. (2014) investigated whether it is possible to automate the process of analyzing the users' interactions. Gotz and Zhou (2009) conducted an empirical study with 30 users who performed typical visual analytic tasks. They analyzed the records of user activities like videos, user notes, and log files. The findings of their study – based upon Activity Theory (Nardi, 1996) – allowed them to characterize the users' analytic behavior. Lipford et al. (2010), on the other hand, showed that participants had problems to recall the steps they performed during the analysis process from memory alone. However, it proved useful to provide the interaction logs for the analysts themselves as an effective memory aid and as a basis for the discussion of their decisions and strategies.

The importance of the analysis of users' interactions to support the reasoning process is also reinforced by a number of different approaches which have been developed over the last years. For example, Cowley et al. (2005) presented a system called Glass Box which captures user interactions and system activities to investigate the analytic process. The approach demonstrated by Groth and Streefkerk (2006) allows users to interact with the history of their interaction and to annotate the data to give them additional opportunities to communicate what and how they found it. Shrinivasan and van Wijk (2008) presented an approach that uses a history tracking mechanism to automatically capture the states of a visualization. Their system allows users to go back to the different visited states in order to review and to validate their findings and also to find alternative solutions. Several investigations show that visualizations of users' activities can be helpful for the analysis of the reasoning process. To name but a few: Kadivar et al. (2009) developed the visual analytics tool CzSaw that records and visualizes the history of the users' interactions. Jankun-Kelly (2008) presents a method to support exploration of the different visualization sessions by using visualization process graphs. The process view of SocialNetSense developed by Gou et al. (2012) – a framework supporting social network sensemaking activities – captures and visualizes users interactions. Scheidegger et al. (2007) and Bavoil et al. (2005) developed approaches using node-link diagrams to give an overview of the sensemaking process. The approach demonstrated by Dunne et al. (2012) supports visual feedback that helps users to track their interactions. Isenberg and Fisher (2011) presented Pairgrams which were developed to visualize log files in combination with observation data to analyze the patterns of reading and searching in a collaboration environment. Furthermore, recording and analysis of users' activities are also important for systems to provide suggestions to the users interacting with the system (see, e.g., Jankun-Kelly et al., 2007; Koop et al., 2008).

Sequence analysis: Analysis of sequences of user interactions is quite common in HCI research – especially for the investigation of the behavior of web users (cf. Benson and Karger, 2014; Radinsky et al., 2013). Mirel (2004) has pointed out the importance of this kind of research especially in the context of complex tasks. Visualizations support such complex tasks. Guo et al. (2016), for example, analyze the interactions of intelligence analysts with visualizations. Similar to the approach described in Pohl et al. (2012b), they used log file analysis. They also adopted the categorization scheme of Yi et al. (2007) and could identify similar interaction sequences. Another area where such complex interaction processes can be observed is exploratory search. Such search processes go beyond simple lookup tasks and comprise cycles of searching and browsing (Marchionini, 2006). Detailed investigations of such processes can yield insights about how to design systems to better support such tasks.

To date, several approaches for sequence analysis have been proposed. For example, Cuomo (1994) as well as Hilbert and Redmiles (2000) examined the applicability of various exploratory sequential data analysis methods, among them lag-sequential analysis (LSA), for HCI related research. Generally speaking, LSA is a statistical procedure that can be employed to find chains of events that repeatedly occur in sequence data. LSA has been first proposed by Sackett (1979) and has since then become an important method to identify significant behavior sequences. LSA has shown to be useful for explorative studies where no specific a priori hypothesis about expected patterns exist (e.g., Bakeman and Gottman, 1997; Mazzi et al., 2003) – as it is the case in the present study. To the best of our knowledge, LSA in particular has not been used widely in HCI yet but existing work in this area shows that LSA provides added value for understanding human computer interaction behaviors. For example, Cuomo (1994) found LSA useful for understanding what activities have preceded an error and Wu et al. (2015) successfully utilized LSA to understand behavioral patterns during online discussions. Consolvo et al. (2002) report that LSA of coded video recordings helped them to verify that an ubiquitous computing application for biologists did smoothly integrate into existing laboratory workflows. Chung and Baker (2003), in turn, applied LSA to logged user-interface events in an interactive learning environment and concluded that these events can be a measure of problem-solving processes if their sequential structure is taken into account. In contrast to these studies, our aim is to understand the reasoning processes of users while interacting with software artifacts – specifically visualizations – and to determine if these processes can be generalized across different products.

Task taxonomies: Several different frameworks have been proposed to describe the activities of users when they interact with information visualization systems (see, e.g., Wehrend and Lewis, 1990; Zhou and Feiner, 1998). The goal of such systems is, among others, to assess the quality of information visualization systems and to describe the activities of the users of such systems. Gotz and Zhou (2009) developed a multi-tier system to characterize users' analytic activities. The authors distinguish between system-oriented taxonomies and user-oriented taxonomies. They argue that user-oriented taxonomies should reflect the users' actions and intentions. Consequently, they developed their taxonomy based on user observation. The authors also conducted a practical validation of their taxonomy with users. They especially identified an intermediate level of actions, consisting of activities such as delete, edit, filter, pan/zoom, remove, and sort. These activities are domain independent but still represent meaningful units of user interactions. The authors also point out that these user actions were performed in various visualizations and are independent of the specific tool which is used. Yi et al. (2007) developed a similar taxonomy reflecting the user's side of view and uses intermediate level categories which can be employed to categorize users' activities. Yi et al.'s taxonomy is based on a considerable amount of previous work on taxonomies. Pohl et al. (2012b) used Yi et al.'s taxonomy successfully to analyze interaction patterns. They could identify certain interaction patterns which were valid for several different visualizations. Kang et al. (2011) demonstrate in their study that, based on a categorization of users' activities, different usage patterns can be identified which, in their case, reflect the properties of the various tools which were used by the subjects and indicate that these tools afford different usage strategies. The idea of an intermediate level of activities reflecting the users' sensemaking strategies and the notion of meaningful interaction patterns influenced the approach we adopted in our research. A more novel typology of visualization tasks was developed by Brehmer and Munzner (2013). Their goal was to develop general categories to compare visualizations from different application domains. Their work is based on the results of previous researchers, for example, Gotz and Zhou (2009) and Yi et al. (2007). Their approach is novel insofar as they distinguish between why a task is performed, how it is performed, and what the task pertains to. They argue that their typology can help to compare the tasks of different users in evaluation studies. In general, the taxonomies described above have many similarities. This is not surprising as the tasks in which users of information visualizations engage in are similar in nature. It is also not surprising that the intermediate level of activities attracts so much attention in research because this is the most interesting level. Gotz and Zhou (2009) argue that high-level activities are not task independent. It is therefore difficult to generalize results in this context. Low-level activities, on the other hand (as, e.g., mouse-over) are difficult to interpret and therefore not useful for evaluation studies.

3. Background

In recent years, the increased complexity of IT systems has emphasized the interactive nature of this technology. Mirel (2004) points out that complex problem solving has to be supported by systems enabling users to engage in open ended inquiry. These activities have to be analyzed closely to adapt the interface to the exploration processes of the users. The goal is to identify patterns of inquiry, that is *recurring sets of actions and strategies that have a successful record in resolving particular types of problems* (Mirel, 2004). Mirel also points out that complex problem solving is typically supported by interactive visualization systems.

In our previous work, we conducted three studies which analyzed interaction sequences of users of visual analytics systems (Pohl et al., 2012b; Rind et al., 2011). All three systems were concerned with time-oriented data from the medical domain (see Fig. 1 for an overview of the systems). Nevertheless, the visualizations differed to a considerable degree. One visualization compared many patients and was based on a *spring metaphor* (Fig. 1, left). Patients had to fill in questionnaires five times during a psychotherapy. This data was shown as an animation, and users were able to identify clusters of patients and whether their therapy was successful or not. The second visualization only presented one patient at a time and, at least partly, used well-known visualizations, for instance, line graphs (*static charts*, Fig. 1, middle). The data shown concerned diabetes patients and, for example, their blood sugar, blood pressure, and medication. Physicians were able to analyze the development of one patient over time. The third visualization used *animated scatterplots* (Fig. 1, right) and the same dataset as the second visualization (*static charts*). In the animated scatterplot visualization, the development of groups of patients over time was shown as animation. The subjects of the first study were 32 students, the subjects of the second study were 9 physicians, and in the third case, there were 10 physicians who participated. In all three cases, the subjects had to solve several tasks developed by experts. The activities of the participants were logged and then categorized according to the categorization scheme developed by Yi et al. (2007). The tasks the participants had to solve with the *spring metaphor* visualization concerned the positive/negative therapy outcome of patients and its relationship to various variables (e.g., depression, number of friends, and relationship to parents). Tasks concerning the *static chart* visualization were, for example, whether users were able to detect relations between the various variables concerning diabetes in one patient and whether the therapy was successful or not over the years. Tasks for the *animated scatterplot* visualization were, for example, to detect the development of groups of patients over time in relation to specific relevant variables. In both visualizations for diabetes data, the question also was whether medication was successful or not. The tasks were always adapted to the needs of practitioners in the field to make them more valid and to be able to achieve realistic tests for the visualizations. Therefore, the tasks for the three visualizations differed to a certain extent. We think that in the context of our study, this is an advantage because if common interaction patterns across all visualizations are detected this is all the more significant as visualizations and tasks and target groups were all different. It is obvious, however, that more rigorous testing has to be conducted to be able to get a more comprehensive idea of the circumstances when certain interaction patterns appear.

We found the categorization scheme developed by Yi et al. (2007) especially valuable because it is designed to capture the users' intentions. Our experience also indicates that it is fairly comprehensive and the categorization of activities according to these categories is quite straightforward in general. Therefore, and in order to allow for comparison with our previous studies, we used this categorization scheme again for the log file analysis in this study. More information about the categorization is provided in Section 4.2.1.

4. Empirical evaluation

In this study we expand upon our previous work in two ways. First, we could observe that it was often difficult to interpret the users' activities solely based on log files. Since thinking aloud can help to clarify such issues, we complement log file analysis with thinking aloud. Furthermore, thinking aloud analysis gives us a possibility to find out what kind of utterances the study participants make while working with a system. Second, we extend our analysis of interaction sequences with LSA. Compared to transitional probabilities, as used in the previous studies, LSA offers two major advantages: (a) transitional probabilities do not reveal if a two-event sequence occurs at a rate greater than chance (cf. McComas et al., 2009) and (b) LSA is well suited for understanding

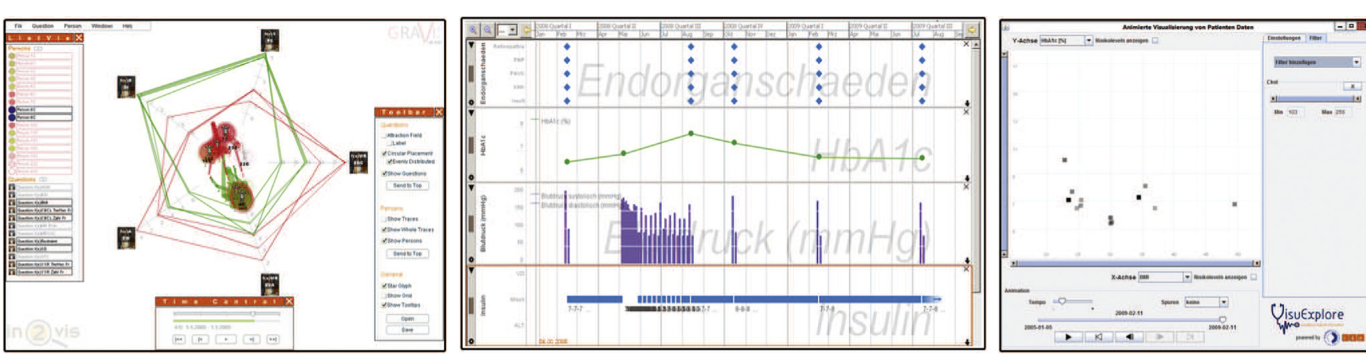

Fig. 1. Screenshots of the three systems evaluated in previous studies: *spring metaphor* (left), *static charts* (middle), and *animated scatterplot* (right).

sequences consisting of more than two events (cf. Olson et al., 1994). In addition, we compare the results from the log file analysis to the findings of our previous studies to determine if similarities in interaction behavior exist.

One basic assumption which has been frequently studied in graph comprehension is that inexperienced users tend to stick to single data values while experts go beyond the data and make inferences about the data (Tversky, 2005). Ratwani et al. (2008) argue that more complex tasks would contain repeated cycles of the same sequence of activities. In our study we wanted to find out how these sequences look like. These cognitive activities can be identified by categorizing verbal utterances.

In particular, this study aims to address the following research questions (R1–R4) which are partly based on our previous work, especially the research questions concerning log file analysis (methods of analysis are listed in parenthesis):

- *R1 – log file analysis* (*frequencies*): Which interaction processes do the users of visualizations engage in? Are there any preferred interactions?
- *R2 – log file analysis* (*transitional probabilities*, *LSA*): Are there any patterns in the interactions? Do users engage in certain interaction sequences repeatedly?
- *R3 – thinking aloud* (*frequencies*): Which kind of utterances dominate in thinking aloud protocols? Is there a larger amount of utterances related to single data values or to more comprehensive reasoning processes?
- *R4 – thinking aloud* (*transitional probabilities*, *LSA*): Which kind of sequences occur in the thinking aloud protocols. It seems sensible to assume that subjects would first talk about single values and then engage in more comprehensive reasoning and inference processes. Are such sequences visible in the data?

4.1. Tested system

For our study we decided to use a system which visualizes contingency tables. Such tables summarize the relation between two categorical variables and arise in both scientific and business domains. Traditional visualizations of contingency tables can only show a limited number of variables. The visualization we investigate in this paper is more complex and enables the users to analyze contingency tables with a low number of columns but a large number of rows. The system allows users to study positive associations in contingency tables and adopts a wheel metaphor for their visualization (see Fig. 2). Data are mapped to sectors and dots. The sectors represent the table columns and form a ring chart. Dots represent a subset of the table cells. A dot is created for a cell in the sector of its column if the rows and columns are positively associated. The system also computes associations between the sectors. Lines represent the existence of shared data between sectors. A slider enables the user to filter the data. When the slider is further away from the center of the visualization fewer dots are represented on the screen. The visualization also shows the original contingency table in an additional panel and a list view in a third panel.

There are systematic differences between the compared systems (the one described in this paper and the three systems tested in previous studies, see Table 1) as one of our goals is to find out whether common interaction patterns across different visualizations with different features and different interaction possibilities exist. The systems especially differ along four dimensions: familiarity, time, animation, and multiple views. One visualization used familiar visualization techniques (e.g., line charts and bar charts), whereas the others required the users to first familiarize themselves with the employed techniques. In contrast to the previous systems, the *contingency table* visualization does not represent time-dependent data. One of the visualizations uses animation

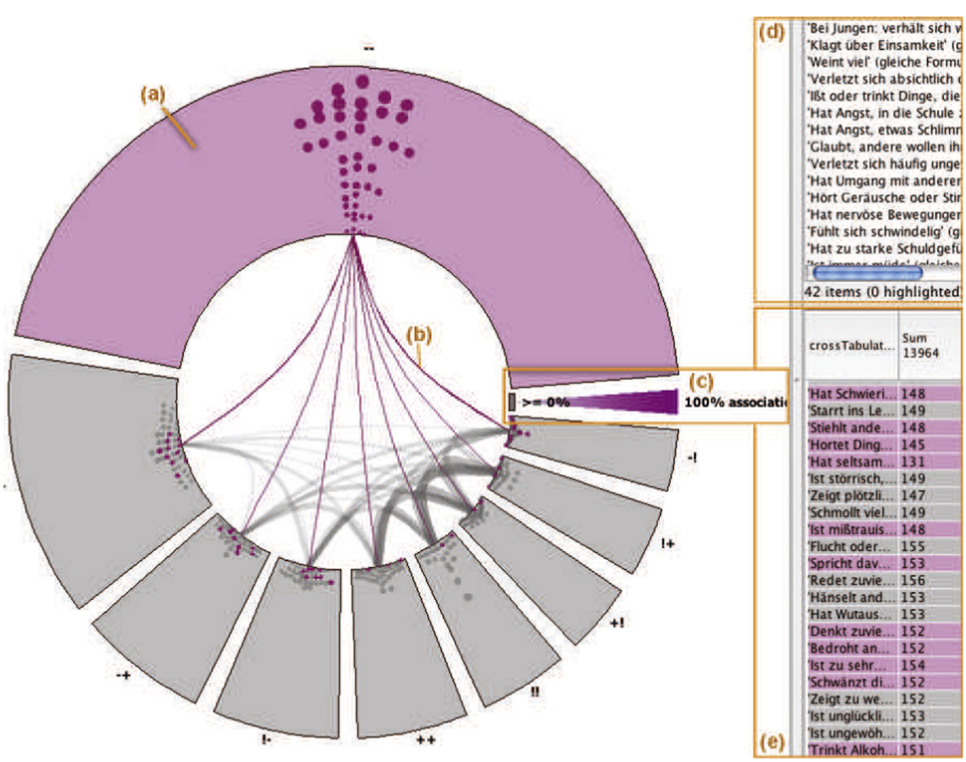

Fig. 2. The *contingency table* visualization: (a) data is shown in sectors, (b) line width of the connections between sectors correlates with the number of connections, (c) slider to filter the data, (d) list of selected entities, and (e) original contingency table. The currently selected sector and all its related dots and selections are marked in pink. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

Table 1
Differentiation between the systems from the previous studies in comparison with the tested system in this study.

Dimension	*Spring metaphor*	*Static charts*	*Animated scatterplot*	*Contingency table*
Familiarity		✓		
Time	✓	✓	✓	
Animation			✓	
Multiple views		✓		✓

and two visualizations use multiple views. In terms of interaction possibilities, the *contingency table* visualization does not offer extensive possibilities to scroll or pan. One of the more important interaction possibilities is the filtering mechanism. In addition, and opposed to the other visualizations, the *contingency table* visualization allows us to establish connections between the data explicitly.

4.2. Methodology

Carpendale (2008) provides a very comprehensive overview of methods from HCI which might be applied in information visualization. The problem we encountered is still more specific. For the analysis of interaction sequences, not every method of analysis applied in HCI is appropriate. Many methods only yield very general values (e.g., measures of time and error). Any method used in the context of interaction processes should preserve the sequence of events, like log file analysis and thinking aloud (for a discussion of this issue see Pohl, 2012).

Categorization is an essential element of the investigation of interaction processes of users with visualization tools. The choice of categories decides which variables of the large amount of observed behavior are considered as relevant. Categorization also allows to aggregate the data and to look at the users' behavior on a more general level and to generalize the results.

In the following, we will briefly describe our employed data collection methods (log files and thinking aloud), along with the applied categorization scheme and the methods we used for analyzing the collected data.

4.2.1. Data collection and categorization

Log files are a common research methodology in HCI and is usually applied to detect usability problems (cf. Ivory and Hearst, 2001). However, data collected via log files can also be used for the analysis of cognitive processes of the users (see, e.g., Dou et al., 2009). Log files enable researchers to study interaction sequences in great detail. In this context, it is essential to decide which data to capture and how to aggregate these data (Lazar et al., 2010). This is related to the problem of categorization and probably the biggest challenge in this context. In our case we chose the categorization scheme developed by Yi et al. (2007) for reasons discussed in Section 3.

Table 2 describes these categories and shows how we categorized user activities logged from our tested systems in regard to this categorization scheme. Our experience indicates that one of the limitations of log files is that they are dependent on the nature of the tool which is used. For instance if the system does not allow filtering then this category will not appear in the analysis. Seen from the point of view of distributed cognition this is, however, also a strength because it emphasizes the relationship to the artifact. The artifact affords some cognitive activities and limits others and, in this way, shapes how the users' cognitive activities will develop.

Thinking aloud is a methodology developed by Ericsson and Simon (1993). The original goal was the investigation of cognitive processes, especially of problem solving activities. It is, therefore, especially appropriate for the investigation of interactions with visualization tools because this process can also be seen as a problem solving activity (cf. Green et al., 2008). Thinking aloud has often been successfully used by researchers in HCI.

In order to categorize the thinking aloud data we derived a categorization scheme or coding frame (Schreier, 2012) for the analysis of the users' utterances based on the theory of graph comprehension (Friel et al., 2001; Ratwani et al., 2008; Trafton et al., 2000; Tversky, 2005) (see also Section 2). The categories we adopted are the main categories that are used in the research of graph comprehension to clarify the interaction of users with graphical representations. An important aspect in this context is the distinction between a superficial understanding of the data which does not go beyond what is visible on the screen and often concentrates on details rather than on the whole picture and, on the other hand, a deep understanding of the data which enables the user to go beyond the data and make inferences and predictions. The categories Quan and Qual (see Table 3) reflect a superficial understanding related to the perception of single data points. The categories Find and Reas indicate inferential activities and the process of going beyond the data. The category Goal denotes actions when users discuss the goals they have when they try to solve a task. Most of the categories in our categorization scheme (see Table 3) have been already used successfully in the research mentioned above. We conducted a trial coding (Schreier, 2012) to adapt our categorization scheme to the data and added another category which we did not find in the literature: Pause. This is a break of more than four seconds. Eventually, we came up with a set of six categories (see Table 3) which reflect the users' activities when interacting with visualizations very well. In this context, segmentation is a challenging problem. In general, an appropriate segmentation should reflect the goal of the research questions (Schreier, 2012). In the research described in this study, the goal is to identify interaction sequences. Therefore, one unit of analysis is a self-contained activity of the users which represents one category of the categorization scheme. Segments, therefore, tend to be fairly small.

In summary we would argue that log file analysis is a methodology which is more apt to clarify issues in the context of

Table 2
Categorization of user activities (log files).

Category	*Description*	*Interactions*
Select	Mark something as interesting	Selection of a sector, of an element in the diagram, or of a row in a table
Explore	Show me something else	Scrolling in the table or list view
Reconfigure	Show me a different arrangement	Merging of sectors
Encode	Show me a different representation	[*Not supported by the tested system*]
Abstract/Elaborate	Show me more/less detail	Moving from one panel to another
Filter	Show me something conditionally	Movement of slider
Connect	Show me related items	Selection of edges

Table 3
Categorization of user activities (thinking aloud).

Category	*Description*
QUAN	Extracting quantitative information
QUAL	Extracting qualitative information
FIND	Finding relationships
GOAL	Discussing the goal of the tasks
PAUSE	Break lasting longer than four seconds
REAS	Reasoning, interpretation of the data, developing a comprehensive mental model of the data, going beyond the data, predicting future trends

distributed cognition research. Thinking aloud, on the other hand, will rather be a research method to clarify issues related to the theory of graph comprehension. Graph comprehension is rather influenced by cognitive psychology and therefore is concerned about the ways how users process information in their heads. Nevertheless, we would like to point out that these relationships cannot be seen as absolute. It is certainly possible to get information relevant for distributed cognition from thinking aloud research and vice versa.

4.2.2. Data analysis

We used content analysis to categorize the data from the log files and thinking aloud protocols. Content analysis is a well-established method in social sciences. It aims at analyzing all kinds of recorded communications (transcripts of interviews, transcripts of verbal communication, documents, observations, characteristics of document layout, etc., Mayring, 2000). Its application area goes beyond written text in a narrow sense and also encompasses other aspects of communication (Schreier, 2012). Important characteristics of content analysis are that it is systematic and has specific steps, and that it helps to reduce data and identify main themes in a communication process (Mayring, 2000; Schreier, 2012). One of the main steps of content analysis is the development of a coding frame (Schreier, 2012) which consists of the categories relevant for the investigation. These categories have to be developed according to the aims of the research question (Mayring, 2000). Such categories can be developed in a deductive (a priori) or an inductive manner based on the material which is analyzed (Harding, 2013). Very often they are a combination of both approaches. Deductive categories, for example, sometimes do not fit the material exactly and have to be revised during the coding process. The aim of content analysis is to identify commonalities and patterns in the material, but also to see occurrences which only happen once, (Harding, 2013). Content analysis is based on coding processes. Saldana (2012) describes several different well-defined procedures of coding.

Ericsson and Simon (1993) who developed the thinking aloud methodology pointed out that coding is an important part of this approach, therefore content analysis is appropriate for the investigation. Content analysis for log files is not an obvious choice, but we think that interaction between human user and computer can also be seen as a kind of communication, therefore we also use content analysis in this case. In both cases, we used deductive categories for the content analysis, but in the case of the thinking aloud protocols we adapted the deductive categories based on the characteristics of the protocols we analyzed.

Once the activities tracked via software logging and the utterances in thinking aloud protocols were categorized using the respective categorization scheme, the elements in each category were counted. These frequencies can show us which interactive possibilities were preferred by the participants and which cognitive activities participants engaged in (as reported in thinking aloud). Subsequently, transitions between categories were tallied in a table using overlapped sampling (i.e., the target of one transition is the antecedent of the next transition) and transitional probabilities were calculated. Transitional probabilities express the likelihood that an activity follows another (or the same) activity (see, e.g., Bakeman and Gottman, 1997; McComas et al., 2009). They are also an indication of certain sequences of activities. Lastly, a LSA was conducted to identify event chains that occur at frequencies greater than chance. The term *lag* in LSA refers to the position of a target code relative to a given criterion code. For example, *lag* 1 denotes that the target code immediately follows the criterion code, *lag* 2 refers to the code two positions ahead of the criterion code and so on. At this point it should also be noted that we are dealing with untimed events, that is, we are interested in the stream of events but not in their particular duration.

In the following we will briefly review the basics of LSA (for an in-depth discussion see, e.g., Bakeman and Gottman, 1997) as used in the context of this work. First, once the interactions performed by the user – as recorded via log files or screen capture – have been coded according to a categorization scheme, the transitions from one code to the subsequent code (*lag* 1) are tallied in a table using overlapped sampling and transitional probabilities are calculated. Following common convention, the rows of the table refer to the preceding event and columns to the following event. Next, *z*-scores are calculated for each transition to determine if the transitional probabilities deviate significantly from their expected value. For that purpose we employed the adjusted residual equation suggested in Bakeman and Gottman (1997) and given by

$$z_{i\to j} = \frac{o_{ij} - e_{ij}}{\sqrt{e_{ij}\left(1 - x_{i+}/N\right)\left(1 - x_{+j}/N\right)}}$$

where o_{ij} is the observed number and $e_{ij} = x_{i+}x_{+j}/N$ is the expected number of transitions from event i to event j, with x_{i+} being the total observed counts of the i-th row, x_{+j} being the total observed counts for the j-th column, and N being the total number of records in the table.

However, as noted by Wampold (1992) – as cited in McComas et al. (2009) – a *z*-score *does not indicate the degree to which a pattern is present*. Following McComas et al. (2009) we are therefore using the *z*-score in conjunction with a strength of association measurement, specifically Yule's *Q*. Yule's *Q* is a transformation of the odds ratio to a $[-1\cdots+1]$ range. For our purposes, a transition from one code to another was then only considered significant if the *z*-score was above the 1.96 level (the critical value assuming a normal distribution and a significance level of 0.05) and the *Q*-value was at least 0.30 (a moderate association according to the convention of Davis, 1971). At this point significant two-event chains are identified. For a three-event chain to be significant, the *lag* 2 transition as well as the two *lag* 1 transitions in the sequence have to be significant (see Olson et al., 1994).

If LSA is used in an exploratory manner – as in the present case – it is prone to Type I errors due to the large number of involved significance tests (cf. Bakeman and Gottman, 1997; Olson et al., 1994). We therefore followed the advice given in Bakeman and Gottman (1997) and performed a Pearson chi-square test on the frequency table to first determine if a significant dependence between rows and columns exists and calculated individual *z*-scores only if the chi-square test was actually significant.

4.3. Sample and procedure

To address our research questions, it was important that participants could focus on the interaction with the visualization itself. Hence, we tested the *contingency table* visualization with ten computer science students who were familiar with visualization and statistical methods. For the evaluation we had a 94×9

contingency table based on a dataset resulting from a questionnaire from 300 children who showed symptoms of ADHD (Attention Deficit Hyperactivity Disorder), their parents, and their teachers.

Each test session was conducted in a quiet office room and the participants were accompanied by a moderator and a note-taker. The test sessions lasted for about 90 minutes and started with an introduction about the structure of the dataset and the basic functions of the system. Afterward, the participants had to solve seven tasks which were developed in conjunction with domain experts. The tasks concentrated on (a) identifying specific data points (e.g., to find the point with the highest frequency in a specific sector), (b) recognizing dependencies between sectors, and (c) the interaction between the multiple views. During the work with the system the participants were encouraged to think aloud. Each test session was recorded by screen capture with audio. Video recordings were coded by two independent raters using VCode (Hailpern and Hagedorn, 2015). For this purpose each coder segmented the video based on the verbal utterances of the participant and assigned a code to the segment based on the categorization scheme given in Table 3. Afterward, disagreements in the coding (e.g., omissions of a segment and assignment of different codes to a segment) were resolved through discussion. Note, that an exact agreement on the onset and offset times of the segments was not required for our investigation as we are interested in the sequence of codes rather than their duration. The resulting coding was then exported to a text file for further analysis. In addition, software logging was used to record the interactions with the system. The recorded interactions were then categorized using the scheme given in Table 2.

5. Results

In the following we report our results organized by research question.

5.1. Log files

5.1.1. R1 – frequencies of activities

In total we counted 1159 activities, with the most common activity being Filter (44% of all activities) followed by Select (32%). This is, to a certain extent, influenced by the specific design of the system which supports Filter activities very well, but Explore (8%) activities only to a limited extent. On the other hand, we would like to point out that the *contingency table* visualization also supports Connect activities, but users did only rarely adopt this option (1%). Further research should probably try to clarify why some of the options of systems are adopted, and others not.

5.1.2. R2 – transitional probabilities, LSA

Table 4 shows the transition probability matrix between the different categories (since the system did not allow to switch between different representations, the category Encode is not listed in the table). A chi-square test confirmed a significant relation between the rows and columns of the tallied frequencies ($\chi^2 = 1088.04$, $df=25$, $p < .001$, Monte Carlo 2-sided). It can be seen that the most probable cases are those where one activity follows the same activity (e.g., Filter → Filter, Select → Select, Explore → Explore). This indicates that users have a tendency to do the same things again and again.

Table 4
Transitional probabilities (log files).

Activity	Filter	Select	Explore	A/E	Reconf.	Connect
Filter	85.33	10.42	0.00	3.28	0.19	0.77
Select	12.81	58.58	3.00	18.26	6.27	1.09
Explore	0.00	20.69	40.23	39.08	0.00	0.00
A/E	16.00	47.20	33.60	0.00	3.20	0.00
Reconf.	21.43	40.48	0.00	7.14	30.95	0.00
Connect	0.00	30.00	0.00	40.00	10.00	20.00
0%						100%

We conducted a LSA according to Bakeman and Gottman (1997). Depending on the computation of z-values and Yule's Q the following sequences of two elements were significant (that is $z > 1.96$ and $Q > 0.30$):

Filter → Filter	($z = 24.84$, $Q = 0.95$)
Select → Select	($z = 13.32$, $Q = 0.71$)
Explore → Explore	($z = 11.88$, $Q = 0.86$)
A/E → Explore	($z = 11.55$, $Q = 0.83$)
Reconfigure → Reconfigure	($z = 9.60$, $Q = 0.89$)
Explore → A/E	($z = 8.79$, $Q = 0.75$)
Connect → Connect	($z = 6.54$, $Q = 0.94$)
Connect → A/E	($z = 2.97$, $Q = 0.70$)

These transitions are depicted graphically in Fig. 3. It is interesting that even for the *contingency table* visualization, which differs from the other investigated systems, two of the most common interaction sequences are Explore → A/E and A/E → Explore. These seem to be interaction sequences which are common across fairly different systems. We also conducted a LSA for significant sequences of three elements. The chains

Explore → A/E → Explore	($z_{lag2} = 10.56$, $Q_{lag2} = 0.82$)
A/E → Explore → A/E	($z_{lag2} = 10.97$, $Q_{lag2} = 0.79$)

were significant in this case. This result is especially surprising because the *contingency table* visualization does not support these activities very well (especially Explore).

5.2. Thinking aloud

5.2.1. R3 – frequencies of activities

As mentioned earlier (cf. Section 4.2.1) the results of the thinking aloud analysis were categorized by a categorization scheme based on research in graph comprehension. From a total of 907 counted utterances, the two most frequent categories were Qual – extracting qualitative information (38%) and reas – interpreting the insights gained from the data (23%). These frequencies are also quite consistent across individual participants of the study. The least frequent categories were Find – find relationship (3%) and Quan – extract quantitative information (8%). The fact that reasoning occurs fairly often is a positive result as it is a common finding in many studies on graph comprehension that users stick to detailed information and refrain from *going beyond the data* (Tversky, 2005).

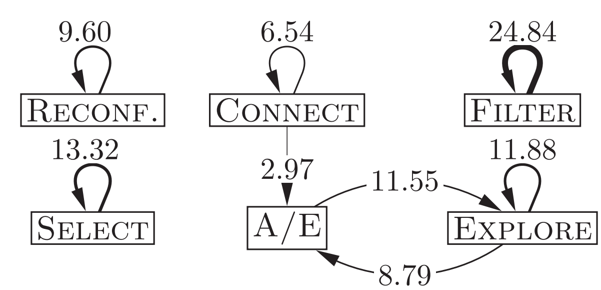

Fig. 3. State transition diagram of significant *lag* 1 transitions ($z > 1.96$, $Q > 0.30$) occurring in the log file data. Edges are labeled with their z-score.

Table 5
Transitional probabilities (thinking aloud).

Activity	Goal	Qual	Quan	Reas	Pause	Find
Goal	12.41	55.47	9.49	8.03	13.14	1.46
Qual	11.95	35.28	4.08	30.32	14.29	4.08
Quan	11.94	22.39	22.39	25.37	16.42	1.49
Reas	15.17	36.97	8.06	24.17	11.85	3.79
Pause	25.66	37.17	13.27	15.93	7.08	0.88
Find	7.69	38.46	0.00	46.15	7.69	0.00
0%						100%

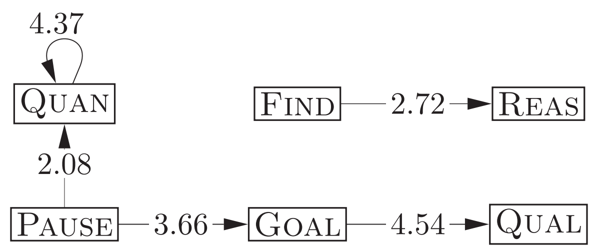

Fig. 4. State transition diagram of significant *lag* 1 transitions ($z > 1.96$, $Q > 0.30$) occurring in the thinking aloud data. Edges are labeled with their z-score.

5.2.2. R4 – transitional probabilities, LSA

Table 5 lists the transitional probabilities between the different categories. Again, a chi-square test (conducted on the observed frequencies) showed a dependence between rows and columns ($\chi^2 = 98.18$, $df=25$, $p < 0.001$).

The LSA revealed the following significant sequences of two elements (see also Fig. 4):

Goal → Qual	($z = 4.54$, $Q = 0.40$)
Quan → Quan	($z = 4.37$, $Q = 0.58$)
Pause → Goal	($z = 3.66$, $Q = 0.40$)
Pause → Quan	($z = 2.08$, $Q = 0.31$)
Find → Reas	($z = 2.72$, $Q = 0.40$)

It is interesting to note that in the thinking aloud analysis there are fewer significant sequences which consist of the same sort of activity. The only exception is Quan → Quan. It might be assumed that this is due to the fact that the thinking aloud analysis reflects considerably less detail, but this is not the case. In the case of thinking aloud a total sum of 907 occurrences were categorized and in the case of the log files 1159 occurrences. This variation does not seem big enough to explain this difference. Talking about the goal of the tasks often occurred at the beginning of the task. Apparently, the first thing the participants did after that was trying to identify qualitative information. When users start to extract quantitative information they usually stick to this activity and repeat it. A pause can perhaps be interpreted as silent reasoning activity. Participants, for example, needed a pause before they talked about quantitative data or their goal. It seems plausible to assume that they first tried to clarify issues before they talked about them. We also observed that finding relations and reasoning were activities which occur together fairly often.

We also conducted an analysis of three-element sequences. In this case, only the combination

Quan → Quan → Quan ($z_{lag2} = 6.41$, $Q_{lag2} = 0.71$)

was significant. This is especially noticeable as Quan (that is identifying a quantitative value) is an activity that does not occur very often.

6. Discussion

In this investigation we studied one visualization in depth – the *contingency table* visualization. We used methods which are appropriate for the analysis of sequences of interaction processes, i.e., log files and thinking aloud. There are other methods which might be used such as eye-tracking and observation. However, eye-tracking poses problems because it is only appropriate for short and well-defined tasks, not for explorative tasks. Observation, on the other hand, can be very time consuming. The combination of log files and thinking aloud is especially appropriate because log files cannot provide information about the users' intentions, thought processes and ideas. This information can be provided by thinking aloud. In addition, log files and thinking aloud do not interfere with each other. In contrast to that, it is difficult to combine eye-tracking and thinking aloud.

6.1. Log file analysis – comparison to previous studies

In the following we compare our results to those of three other visualizations. As discussed earlier these visualizations differ in several important aspects (cf. Table 1). They were developed for specific goals and user groups. The tasks were developed together with domain experts and are fairly realistic. The tasks were different for the various visualizations.

Table 6 shows the percentage of the frequencies of activities of the three systems (*static charts*, *animated scatterplot*, and *spring metaphor*) taken from previous studies (Pohl et al., 2012b) and the results concerning the visualization (*contingency table*) tested in this study. The three visualizations with time-oriented data (*static charts*, *animated scatterplot*, and *spring metaphor*) apparently yield fairly similar results. The majority of activities has an explorative character. It is obvious that users adopt explorative activities (e.g., scrolling, panning, etc.) to a much larger extent than activities which change the appearance of the visualization fundamentally (Encode and Reconfigure respectively). Users also seem to be very interested in the exact data values (A/E). This means, they either look up the exact values of the data by mouse over or by looking up the values in a table or by zooming into the data. The other possibilities were not used very often. Some of the activities were not supported by some visualizations (Connect was not supported by the *spring metaphor* and *static charts* visualizations; Filter was not supported by the *static charts*). From these data it might be concluded that users prefer explorative activities and looking at exact data values, but they avoid activities which might change the appearance of the visualization on the screen. The *contingency table* visualization shows a different picture. Exploration is not supported very well by the system and therefore is seldom used. The majority of activities are related to filtering and selecting objects on the screen. Looking at precise values of the data was

Table 6
Percentage of the frequencies of activities of the systems from the previous studies in comparison with the tested system in this study. Activities not supported by a system are marked with an ×.

Activity	*Spring metaphor*	*Static charts*	*Animated scatterplot*	*Contingency table*
Explore	34.72	52.67	22.43	8.00
A/E	44.78	30.36	41.61	11.00
Encode	0.99	1.57	9.57	×
Select	6.09	10.72	0.24	32.00
Reconf.	8.92	4.67	11.89	4.00
Filter	4.50	×	9.57	44.00
Connect	×	×	4.68	1.00
0%				100%

supported by the system but presumably underestimated by the log file data because a considerable part of the original contingency table was visible on the screen. Therefore, few activities related to A/E were recorded by the log files. Encode was not supported. One conclusion which might be drawn from these data is that if similar features are offered by visualizations then users adopt similar forms of interaction.

In our previous study for the spring metaphor, static charts, and animated scatterplot visualizations (Pohl et al., 2012b) the results indicate that there are a few two-element interaction sequences occurring frequently in all three visualizations, for example, A/E→A/E. In general, participants tended to repeat the same activities again and again. This is similar to what we found for the contingency table visualization.

We only found two three-element interaction sequences in our current study which occurred with a probability greater than chance: Explore→A/E→Explore and A/E→Explore→A/E. These sequences were also found in the three previous studies. They are obviously fundamental activities. Users first explore the visualization (e.g., by scrolling and panning) and then look at specific data (A/E) and explore further (or vice versa). In the three previous studies (Pohl et al., 2012b) visual inspection of the coded sequences and the results from transitional probabilities also suggested additional sequences: Explore→A/E→Reconfigure and Filter→A/E→Filter. These sequences were not found in the current study. The methodology we used to identify the sequences in the three previous studies was transitional probabilities and visual inspection. The lag sequential analysis used in the current study allows for a much more reliable identification of statistically significant sequences. The differences in the findings may be due to this change in methodologies or to the difference in the visualization. This has to be clarified in further research.

6.2. Thinking aloud

The thinking aloud data yield some interesting results compared to previous research in the area of graph comprehension. The fact that a considerable proportion of the utterances expressed reasoning processes (23%) of the users indicates that an appropriate design of a visualization might be able to support such processes and counteract a tendency to avoid more comprehensive interpretation of the data. It is noticeable that there are only very few utterances concerning Find (3%), although the *contingencies tables* system does support such interactions. The data which was extracted from the visualization was mostly qualitative. This conforms to research by Trafton et al. (2000). They argue that users usually develop a qualitative mental model depending on the qualitative data in the visualization. Based on this qualitative model experts can also derive quantitative information. From an informal inspection of the data we know that there were some regularities concerning the utterances in thinking aloud. First, the users discuss the goal of their task, then they study the data (quantitative and qualitative extraction) and in a final stage they interpret the data, find relationships and reason about the data. This structure of solving a task also seems quite plausible and it is, to a certain extent, supported by the results of the LSA. There is a high probability, that utterances concerning the goal will be followed by a discussion of qualitative data (Goal→Qual). If quantitative data are discussed, users will stick to this activity and do this again and again. When users talk about finding relationships, this is very often followed by reasoning activities (Find→Reas).

Ratwani et al. (2008) found that for simple lookup tasks there is usually just one sequence of utterances. Complex tasks are more exploratory and contain repeated cycles of sequences. Ratwani et al. (2008) found that reasoning processes and processes of qualitative data extraction followed each other frequently. We could not observe this specific cycle of utterances. Nevertheless, we found other cycles, for example, that extraction of quantitative data followed extraction of quantitative data (Quan→Quan) and reasoning followed utterances about finding relationships (Find→Reas).

7. Limitations and future work

There are some limitations of the approach described in this paper. Probably the most important limitation relates to the fact that the participants of the studies described here only interacted with the systems for a limited amount of time. Some of the results we got could be due to this fact. It would, for example, make sense to assume that persons who interact with a system only for a short time only use basic functionalities (e.g., Explore and A/E). It would be interesting to clarify whether persons who use the systems for a longer period of time also use functionalities which change the appearance of the visualization on the screen to a larger extent. For future work it might also be desirable to investigate longer interaction sequences. However, it should be noted that LSA requires a sufficient data size – which increases with the length of the sequences under investigation and the number of codes – in order to draw meaningful conclusions (see, e.g., Bakeman and Gottman, 1997 for a more detailed discussion). These data demands might not be an issue if interaction logs can be automatically gathered from a large user base but might be prohibitive if analysis relies on thinking aloud protocols or on interaction data from a limited number of experts.

Another limitation concerns the fact that the four different visualizations were not compared systematically. Not only the visualizations were different, but also the tasks and the samples (experts or students). In an exploratory study this can be an advantage because the identification of common patterns of interactions across these four different cases seems to be quite valid. Despite the differences, we could identify one common significant three-element interaction pattern and several common significant two-element interaction patterns. Nevertheless, it is necessary to compare visualizations more systematically in the future to find out whether there are patterns which occur only in specific situations.

Another open issue concerns the different possible levels of interactions. Gotz and Zhou (2009), for example, developed a multi-tier model of interaction with visualizations. We think that one should at least distinguish between a low level (keystroke level), an intermediate level (interactions like, delete, edit, filter, pan/zoom, etc.), and a high level which is not domain independent (e.g., identifying patterns of development of groups of patients concerning therapy outcome). It is an open question which level is the most interesting for the analysis of the interactions of users of information visualizations. We chose the intermediate level because it is still domain independent, and different visualizations can be compared, and has, on the other hand a fairly coarse granularity to ensure that interactions become meaningful. Nevertheless, an analysis of other levels of granularity is certainly also meaningful.

There are several issues we would like to address in future research. We want to conduct long-term studies to clarify whether our results can also be observed in such situations. Furthermore, we also want to investigate whether there are interrelationships between data from log files and from thinking aloud. The question is whether activities in the thinking aloud protocol often occur together with activities from log files. This would enable researchers to find a relationship between activities which can be observed easily and cognitive processes going on in a person's mind. So far, we also did not validate whether the activities the users engage in are successful or not. We think, it would be interesting to investigate how the users act and what results they get.

8. Conclusions

In this paper we described research concerning interaction processes with visualizations systems. We described a study concerning a visualization tool representing data from contingency tables. Then we compared the results of the log file analysis with three different studies from previous research. Our investigation indicates that some patterns can be observed across fairly different visualizations:

- We analyzed the frequencies of activities in log files and found that some of them occur significantly more often than others (especially exploration and retrieval of specific data as opposed to changing the appearance of a visualization which does not occur very often).
- We found a significant sequence of three activities in the log file data: Explore → A/E → Explore. This is surprising because in the contingency wheel investigation these activities were quite rare. Nevertheless, this sequence was significant.
- The above-mentioned example shows that one of the strengths of LSA is the fact that it can identify sequences of interactions which occur statistically more often than others while the single interactions themselves occur only rarely.
- We also found significant results for sequences of two interactions. In this context, it is noticeable that most of the significant sequences consist of an activity followed by the same activity (e.g. Explore → Explore). This is specific for the log file data and could not be observed in the thinking aloud data.

The results from the log file analysis indicate that users tend to engage in activities resembling natural behavior in everyday life. Users explore the visualizations extensively. They repeat such processes again and again as if they had various hypotheses in their mind which they want to test. This is similar to moving around to look at objects from different angles. They also try to get more extensive information (Elaborate) which resembles going closer to see the details of an object. On the other hand, they avoid activities which change the appearance of objects on a screen (especially Encode – showing a different representation of the data on the screen). This might be due to the fact that in everyday activities it is usually not possible to change the appearance of objects fundamentally. We cannot, for example, change the appearance of houses when we try to find our way in a city to make it easier to remember a route or to build a coherent mental model of the layout of the streets. In addition, changing the representation of data on the screen is fairly disruptive. In general, users strive to develop relatively stable mental models of their environment. Changing the representation of data fundamentally is probably not helpful to retain such stable mental models. To a certain extent, these results contradict common assumptions in information visualization that presenting data in different views or as different kinds of visualizations is beneficial for the users. It should be mentioned, however, that there are still many open questions. It is, for example, not clear whether the strategies based on exploration and drilling down into the data are more successful than adopting more disruptive strategies as, for example, changing the visualization. In addition, it is necessary to clarify whether users will adopt more sophisticated activities when they have used a visualization for a longer period of time than in our investigations. If these usage patterns also lead to valid insights and are observed in long-term studies, one consequence for the design of visualizations might be to support exploration more extensively than changing the appearance of the visualization.

The thinking aloud study indicates that some sequences of cognitive processes occur with a higher probability than others (e.g., finding relationships is often followed by reasoning activities). The results from the thinking aloud protocols support the model that users first describe their goal, then extract data, and then reason about the data. This is not the only possible model. Ratwani et al. (2008) research suggests that there are repeated loops of a sequence of extracting data and reasoning about the data. We did, however, not observe such loops. This discrepancy might be due to differences in the experimental setting. The strategies adopted in this context probably depend on the types of tasks, the time the users spend on the task, or the tools that were used. These issues have to be clarified by future research.

For this study we used systems which were developed for specific target groups and realistic tasks. This increases the ecological validity of our results. We think that the research described in this paper provides some first insights into interaction processes of users with visualization systems. We hypothesize that these results allow us to make inferences about the users' reasoning processes. There are still many open questions, but if these results are refined and confirmed, this would allow us to develop visualizations that are adapted to specific reasoning processes of the users.

Acknowledgments

We acknowledge the contribution of Désirée Lavaulx-Vrécourt and Ruth Wittmann in collecting and analyzing the data. We also thank Bilal Alsallakh for the development of the tested system. This work is conducted in the context of the CVAST (Centre of Visual Analytics Science and Technology) project. It is funded by the Austrian Federal Ministry of Science, Research, and Economy in the exceptional Laura Bassi Centres of Excellence initiative (project number: 822746).

References

Andrews, C., North, C., 2012. Analyst's workspace: an embodied sensemaking environment for large, high-resolution displays. In: Proceedings of the VAST'12, pp. 123–131.

Bakeman, R., Gottman, J., 1997. Observing Interaction: An Introduction to Sequential Analysis. Cambridge University Press, Cambridge.

Bavoil, L., Callahan, S., Crossno, P., Freire, J., Scheidegger, C., Silva, C., Vo, H., 2005. VisTrails: enabling interactive multiple-view visualizations. In: Proceedings of the IEEE Visualization, pp. 135–142.

Benson, E., Karger, D.R., 2014. End-users publishing structured information on the web: an observational study of what, why, and how. In: Proceeding of the CHI'14. ACM, New York, pp. 1265–1274.

Brehmer, M., Munzner, T., 2013. A multi-level typology of abstract visualization tasks. IEEE Trans. Vis. Comput. Graph. 19 (December (12)), 2376–2385.

Brown, E.T., Ottley, A., Zhao, H., Lin, Q., Souvenir, R., Endert, A., Chang, R., 2014. Finding waldo: learning about users from their interactions. IEEE Trans. Vis. Comput. Graph. 20 (12), 1663–1672.

Carpendale, S., 2008. Evaluating information visualizations. In: Kerren, A., Stasko, J., Fekete, J.-D., North, C. (Eds.), Information Visualization. Lecture Notes in Computer Science vol. 4950. Springer, Berlin, Heidelberg, pp. 19–45.

Chung, G.K.W.K., Baker, E.L., 2003. An exploratory study to examine the feasibility of measuring problem-solving processes using a click-through interface. J. Technol. Learn. Assess. 2.

Consolvo, S., Arnstein, L., Franza, B., 2002. User study techniques in the design and evaluation of a UbiComp environment. In: Borriello, G., Holmquist, L. (Eds.), Proceedings of the UbiComp'02. Lecture Notes in Computer Science vol. 2498. Springer, Berlin, Heidelberg, pp. 73–90.

Cowley, P., Nowell, L., Scholtz, J., 2005. Glass Box: an instrumented infrastructure for supporting human interaction with information. In: Proceeding of the HICSS'05. IEEE, Washington, DC, p. 296c, http://dx.doi.org/10.1109/HICSS.2005.286.

Cuomo, D.L., 1994. Understanding the applicability of sequential data analysis techniques for analysing usability data. Behav. Inf. Technol. 13 (1–2), 171–182.

Davis, J., 1971. Elementary Survey Analysis. Methods of Social Science Series. Prentice-Hall, Englewood Cliffs.

Dou, W., Jeong, D., Stukes, F., Ribarsky, W., Lipford, H., Chang, R., 2009. Recovering reasoning processes from user interactions. IEEE Comput. Graph. Appl. 29 (3), 52–61.
Dunne, C., Henry Riche, N., Lee, B., Metoyer, R., Robertson, G., 2012. GraphTrail: analyzing large multivariate, heterogeneous networks while supporting exploration history. In: Proceeding of the CHI'12. ACM, New York, pp. 1663–1672.
Ericsson, K., Simon, H., 1993. Protocol Analysis – Verbal Reports as Data. MIT Press, Cambridge.
Friel, S., Curcio, F., Bright, G., 2001. Making sense of graphs: critical factors influencing comprehension and instructional implications. J. Res. Math. Educ. 32 (2), 124–158.
Gotz, D., Zhou, M.X., 2009. Characterizing users' visual analytic activity for insight provenance. Inf. Vis. 8 (1), 42–55.
Gou, L., Zhang, X., Luo, A., Anderson, P., 2012. SocialNetSense: supporting sensemaking of social and structural features in networks with interactive visualization. In: Proceeding of the VAST'12, pp. 133–142.
Green, T., Ribarsky, W., Fisher, B., 2008. Visual analytics for complex concepts using a human cognition model. In: Proceeding of the VAST'08, pp. 91–98.
Groth, D.P., Streefkerk, K., 2006. Provenance and annotation for visual exploration systems. IEEE Trans. Vis. Comput. Graph. 12 (6), 1500–1510.
Guo, H., Gomez, S.R., Ziemkiewicz, C., Laidlaw, D.H., 2016. A case study using visualization interaction logs and insight metrics to understand how analysts arrive at insights. IEEE Trans. Vis. Comput. Graph. 22 (1), 51–60.
Hailpern, J., Hagedorn, J., 2015. VCode ⟨http://social.cs.uiuc.edu/projects/vcode.html⟩ (accessed March 2015).
Harding, J., 2013. Qualitative Data Analysis from Start to Finish. Sage Publications Ltd., London.
Hilbert, D.M., Redmiles, D.F., 2000. Extracting usability information from user interface events. ACM Comput. Surv. 32 (4), 384–421.
Hollan, J., Hutchins, E., Kirsh, D., 2000. Distributed cognition: toward a new foundation for human–computer interaction research. ACM Trans. Comput.–Hum. Interact. 7 (2), 174–196.
Hutchins, E., 1995. Cognition in the Wild. MIT Press, Cambridge.
Isenberg, P., Fisher, D., 2011. Pairgrams: understanding collaborative analysis behavior with visualization. In: Proceeding of the CHI Workshop on Analytic Provenance: Process + Interaction+ Insight.
Ivory, M., Hearst, M., 2001. The state of the art in automating usability evaluation of user interfaces. ACM Comput. Surv. 33 (4), 470–516.
Jankun-Kelly, T.J., 2008. Using visualization process graphs to improve visualization exploration. In: Freire, J., Koop, D., Moreau, L. (Eds.), Provenance and Annotation of Data and Processes. Springer, Berlin, Heidelberg, pp. 78–91.
Jankun-Kelly, T.J., Ma, K.-L., Gertz, M., 2007. A model and framework for visualization exploration. IEEE Trans. Vis. Comput. Graph. 13 (2), 357–369.
Kadivar, N., Chen, V., Dunsmuir, D., Lee, E., Qian, C., Dill, J., Shaw, C., Woodbury, R., 2009. Capturing and supporting the analysis process. In: Proceeding of the VAST'09. IEEE, Atlantic City, Washington, DC, pp. 131–138.
Kang, Y.-A., Gorg, C., Stasko, J., 2011. How can visual analytics assist investigative analysis? Design implications from an evaluation. IEEE Trans. Vis. Comput. Graph. 17 (May (5)), 570–583.
Kirsh, D., 2010. Thinking with external representations. AI SOC. 25 (4), 441–454.
Kirsh, D., Maglio, P., 1994. Distinguishing epistemic from pragmatic action. Cogn. Sci. 18 (4), 513–549.
Koop, D., Scheidegger, C.E., Callahan, S.P., Freire, J., Silva, C.T., 2008. VisComplete: automating suggestions for visualization pipelines. IEEE Trans. Vis. Comput. Graph. 14 (6), 1691–1698.
Lazar, J., Feng, J., Hocheiser, H., 2010. Research Methods in Human–Computer Interaction. Wiley, Chichester.
Lipford, H., Stukes, F., Dou, W., Hawkins, M., Chang, R., 2010. Helping users recall their reasoning process. In: Proceeding of the VAST'10, pp. 187–194.
Liu, Z., Nersessian, N., Stasko, J., 2008. Distributed cognition as a theoretical framework for information visualization. IEEE Trans. Vis. Comput. Graph. 14 (6), 1173–1180.
Marchionini, G., 2006. Exploratory search: from finding to understanding. Commun. ACM 49 (4), 41–46.
Mayring, P., 2000. Qualitative content analysis. Forum Qualitative Sozialforschung/Forum: Qualitative Social Research 1 (June (2)).
Mazzi, M.A., Del Piccolo, L., Zimmermann, C., 2003. Event-based categorical sequential analyses of the medical interview: a review. Epidemiol. Psichiatr. Soc. 12 (2), 81–85.
McComas, J., Moore, T., Dahl, N., Hartman, E., Hoch, J., Symons, F., 2009. Calculating contingencies in natural environments: issues in the application of sequential analysis. J. Appl. Behav. Anal. 42 (2), 413–423.
Mirel, B., 2004. Interaction Design for Complex Problem Solving: Developing Useful and Usable Software. Interactive Technologies Series. Elsevier, San Francisco.
Nardi, B.A. (Ed.), 1996. Context and Consciousness: Activity Theory and Human–Computer Interaction. The MIT Press, Cambridge.
North, C., Chang, R., Endert, A., Dou, W., May, R., Pike, B., Fink, G., 2011. Analytic provenance: process+interaction+insight. In: Proceeding of the CHI EA'11. ACM, New York, pp. 33–36.
Olson, G., Herbsleb, J., Rueter, H., 1994. Characterizing the sequential structure of interactive behaviors through statistical and grammatical techniques. Hum.-Comput. Interact. 9 (4), 427–472.
Pike, W.A., Stasko, J., Chang, R., O'Connell, T.A., 2009. The science of interaction. Inf. Vis. 8 (4), 263–274.
Pohl, M., 2012. Methodologies for the analysis of usage patterns in information visualization. In: Proceeding of the BELIV'12. ACM, New York, pp. 17:1–17:3.
Pohl, M., Smuc, M., Mayr, E., 2012a. The user puzzle – explaining interaction with visual analytics systems. IEEE Trans. Vis. Comput. Graph. 18 (12), 2908–2916.
Pohl, M., Wiltner, S., Miksch, S., Aigner, W., Rind, A., 2012b. Analysing interactivity in information visualisation. KI – Kuenstliche Intell. 26 (2), 151–159.
Radinsky, K., Svore, K.M., Dumais, S.T., Shokouhi, M., Teevan, J., Bocharov, A., Horvitz, E., 2013. Behavioral dynamics on the web: learning, modeling, and prediction. ACM Trans. Inf. Syst. 31 (3), 16:1–16:37.
Ratwani, R., Trafton, J., Boehm-Davis, D., 2008. Thinking graphically: connecting vision and cognition during graph comprehension. J. Exp. Psychol.-Appl. 14 (1), 36–49.
Rind, A., Aigner, W., Miksch, S., Wiltner, S., Pohl, M., Drexler, F., Neubauer, B., Suchy, N., 2011. Visually exploring multivariate trends in patient cohorts using animated scatter plots. In: Proceeding of the EHAWC'11. Springer-Verlag, Berlin, Heidelberg, pp. 139–148.
Sackett, G., 1979. The lag sequential analysis of contingency and cyclicity in behavioral interaction research. In: Osofsky, J.D. (Ed.), Handbook of Infant Development. Wiley, New York, pp. 623–649.
Saldana, J., 2012. The Coding Manual for Qualitative Researchers, 2nd ed. Sage Publications Ltd., London.
Sanderson, P.M., Fisher, C., 1994. Exploratory sequential data analysis: foundations. Hum.-Comput. Interact. 9 (4), 251–317.
Scheidegger, C., Vo, H., Koop, D., Freire, J., Silva, C., 2007. Querying and creating visualizations by analogy. IEEE Trans. Vis. Comput. Graph. 13 (November (6)), 1560–1567.
Schreier, M., 2012. Qualitative Content Analysis in Practice. Sage Publications Ltd., London.
Shrinivasan, Y B., van Wijk, J.J., 2008. Supporting the analytical reasoning process in information visualization. In: Proceeding of the CHI'08. ACM, New York, pp. 1237–1246.
Sun, G.-D., Wu, Y.-C., Liang, R.-H., Liu, S.-X., 2013. A survey of visual analytics techniques and applications: state-of-the-art research and future challenges. J. Comput. Sci. Technol. 28 (5), 852–867.
Trafton, J., Kirschenbaum, S., Tsui, T., Miyamoto, R., Ballas, J.A., Raymond, P.D., 2000. Turning pictures into numbers: extracting and generating information from complex visualizations. Int. J. Hum.-Comput. Stud. 53 (5), 827–850.
Trickett, S.B., Trafton, J.G., 2006. Toward a comprehensive model of graph comprehension: making the case for spatial cognition. In: Barker-Plummer, D., Cox, R., Swoboda, N. (Eds.), Diagrammatic Representation and Inference. Lecture Notes in Computer Science vol. 4045. Springer, Berlin, Heidelberg, pp. 286–300.
Tversky, B., Visuospatial reasoning. In: Keith, J.H., Robert, G.M. (Eds.), The Cambridge Handbook of Thinking and Reasoning. Cambridge University Press, Cambridge, 2005, pp. 209–240.
Walker, R., Slingsby, A., Dykes, J., Xu, K., Wood, J., Nguyen, P.H., Stephens, D., Wong, B.L.W., Zheng, Y., 2013. An extensible framework for provenance in human terrain visual analytics. IEEE Trans. Vis. Comput. Graph. 19 (12).
Wampold, B., 1992. The intensive examination of social interactions. In: Kratochwill, T.R., Levin, J.R. (Eds.), Single-Case Research Design and Analysis: New Directions for Psychology and Education. L. Erlbaum Associates, Hillsdale, pp. 93–131.
Wehrend, S., Lewis, C., 1990. A problem-oriented classification of visualization techniques. In: Proceedings of the IEEE Visualization. IEEE, Los Alamitos, pp. 139–143.
Wu, S.-Y., Chen, S.Y., Hou, H.-T., 2015. A study of users' reactions to a mixed online discussion model: a lag sequential analysis approach. Int. J. Hum.-Comput. Interact. 31 (3), 180–192.
Xu, K., Attfield, S., Jankun-Kelly, T., Wheat, A., Nguyen, P., Selvaraj, N., 2015. Analytic provenance for sensemaking: a research agenda. IEEE Comput. Graph. Appl. 35 (May (3)), 56–64.
Yi, J., Kang, Y., Stasko, J., Jacko, J., 2007. Toward a deeper understanding of the role of interaction in information visualization. IEEE Trans. Vis. Comput. Graph. 13 (6), 1224–1231.
Zhou, M.X., Feiner, S.K., 1998. Visual task characterization for automated visual discourse synthesis. In: Proceeding of the CHI'98. ACM Press/Addison-Wesley Publishing Co., New York, pp. 392–399.

Margit Pohl studied computer science and psychology in Vienna. She is an associate professor at the Institute for Design and Technology Assessment at Vienna University of Technology. Her main research interest is in human–computer interaction and psychological aspects of interaction processes with computer systems. She participated in various projects, also EU projects. Main topics of these projects were information visualization and visual analytics in the medical domain, gender aspects of the work in control rooms, visualization design for traffic control systems, and e-learning/vocational training. She has published more than 100 peer reviewed scientific articles.

Günter Wallner is senior scientist at the University of Applied Arts Vienna. His research interests lie generally in the fields of HCI, information visualization, data analytics, and games. His current research focuses on analysis and visualization of game telemetry data. He holds a doctorate degree in natural sciences and a diploma degree in computer science. His work has been published in international journals and conference proceedings, such as Computers & Graphics, Entertainment Computing, SIGCHI, FDG, and ACE.

Simone Kriglstein studied computer science at the Vienna University of Technology and graduated in 2005. In 2011 she received her Doctoral degree from the University of Vienna. Since then she has worked on several projects at the University of Vienna (e.g., OCIS, Viz4PAIS, and playthenet) and the Vienna University of Technology (e.g., CVAST). She also worked as usability engineer consultant and user interface designer for several companies. Her research interests are interface and interaction design, usability, information visualization, and games.