Visual Perception of Parallel Coordinate Visualizations

Harri Siirtola, Tuuli Laivo, Tomi Heimonen, and Kari-Jouko Räihä
Tampere Unit for Computer-Human Interaction (TAUCHI)
Department of Computer Sciences, FIN-33014 University of Tampere, Finland
{harri.siirtola, tuuli.laivo, tomi.heimonen, kari-jouko.raiha}@cs.uta.fi

Abstract

Parallel coordinates is a visualization technique that provides an unbiased representation of high-dimensional data. The parallel configuration of axes treats data dimensions uniformly and is well suited for exploratory visualization. However, first-time users of parallel coordinate visualizations can find the representation confusing and difficult to understand. We used eye tracking to study how parallel coordinate visualizations are perceived, and compared the results to the optimal visual scan path required to complete the tasks. The results indicate that even first-time users quickly learn how to use parallel coordinate visualizations, pay attention to the correct task-specific areas in the visualization, and become rapidly proficient with it.

Keywords— **Parallel coordinate visualizations, eye tracking, perception of visualization**

1 Introduction

Exploratory data visualization as an approach requires more flexibility from a visualization technique than solutions that are tailored to specific data and tasks. Perhaps the most important requirement for exploratory visualization is the uniform treatment of data dimensions, which is offered only by few visualization techniques. The parallel coordinate plots technique [2] treats all data dimensions equally, displays them simultaneously, and the number of dimensions is limited only by the available screen real estate. In addition, parallel coordinate plots can be turned into an efficient interactive visualization by constructing a software tool, or parallel coordinate browser, to explore them. These software tools usually implement a number of interaction techniques that facilitate the acquisition of insight into data, such as brushing, highlighting, drill-down, and visual queries with logical connectives AND, OR, and XOR.

The fundamental idea in parallel coordinate plots is to use parallel coordinate axes instead of the standard orthogonal Cartesian layout (Figure 1). When the standard axes layout is exhausted already by three-dimensional data, the parallel layout can in theory handle an unlimited number of data dimensions (in practise limited only by the display width). A parallel coordinate plot effectively "flattens" a multidimensional data set into a two-dimensional image.

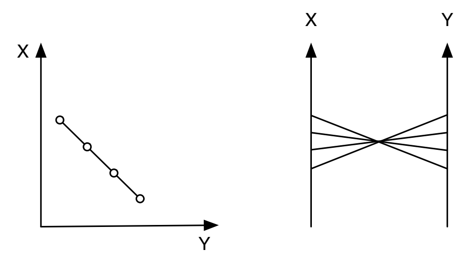

Figure 1: The idea of parallel coordinate plots: use of parallel axes layout instead of orthogonal one. The four points on the left are represented by the four lines on the right.

Encountering parallel coordinate visualizations for the first time can be an alienating experience, especially if the person has a negative attitude towards mathematical or statistical graphics. We have repeatedly observed this phenomenon in our undergraduate information visualization courses, both in the class and via lecture diaries. However, interactive exploration of parallel coordinate plots often overcomes this initial attitude, and the chance for graphical discovery provided by interaction is soon appreciated.

In our earlier experiment [11], we observed that participants learned to use parallel coordinate visualizations surprisingly rapidly. This suggests that although the visualizations based on the parallel coordinate plots are seen abstract and alienating by first-time users, their immediate usability is still quite good. This study focusses on the visual perception of parallel coordinate visualizations by using eye tracking equipment to capture the gaze behavior of participants. By inspecting how the parallel coordinate visualizations are perceived, we hope to uncover whether users focus their attention to the relevant areas while performing a task and how these scan paths correspond to the optimal, task-specific order of area visitation.

978-0-7695-3733-7/09 $25.00 © 2009 IEEE
DOI 10.1109/IV.2009.25

2 Related work

The idea of parallel coordinates was developed by Inselberg [1] (please see [11, 10] for a thorough review) and further revised by him and his colleagues. The use of parallel coordinate plots in interactive visualizations has been studied widely, including our own work on direct manipulation of parallel coordinates [9] and interaction with parallel coordinates [11]. In the latter article, we described an experiment where sixteen database professionals performed a set of query tasks both with the SQL query language and a parallel coordinate browser. Those results show that although many of the participants had doubts about the usefulness of parallel coordinate visualizations, they could perform tasks more efficiently with the visualization interface than with their familiar database query language.

The use of gaze analysis to study information visualizations has recently become more popular as eye tracking equipment allows unobtrusive gaze tracking. Spence et al. [12] did a comparison of different RSVP visualization modes by utilizing gaze analysis. They found that visualization techniques significantly affected users' gaze behaviour. Eye tracking analysis complements findings from more traditional measures, such as error rate, and provides actionable eye movement data for design and research.

Pirolli et al. [6] studied the effect of "information scent" on Hyperbolic Tree Browser. With two eye tracker studies, they concluded that the performance of the system was highly affected by proximal cues to the value of distal information.

Johansson et al. [4] studied a new visual quality metric, "acceptable distortion of patterns", in context of parallel coordinate visualizations and noisy data sets. They explored the users' ability to recognize patterns in parallel coordinate visualizations as the data became increasingly noisy. The recognition of patterns in parallel coordinate plots is demanding, as the familiar shapes appear different in the parallel world. They suggest limits for pattern recognition both with respect to level of noise in the data and to the maximum number of variables.

3 Experiment

In an earlier experiment [11], 16 database professionals carried out eight tasks with an SQL query interface and eight tasks with a Parallel Coordinate Explorer (PCE) program [9]. In this study, we used the same experimental setup, but without a comparison to the SQL interface. Eye tracking device was used to capture participants' gaze data, especially their fixation targets and lengths. A *fixation* occurs when the person maintains his or her gaze on a single location for a certain period of time, and it can be used as an approximation for the person's focus of attention [13].

3.1 Participants

Nine participants were recruited for the study via a student organization's mailing list. They received two cinema tickets as compensation for their participation. Potential participants were screened to have at least some experience with data graphics (creating graphs with MS Excel or similar software was deemed adequate experience). For seven participants, the capture of gaze data was successful, and their data was further analyzed. These seven participants (3 female, 4 male) were 24–30 years old.

3.2 Materials

A Java applet version of PCE was used in the tests (available for experimenting at `http://www.cs.uta.fi/~hs/pce/`, Java source available upon request). The applet runs within a Web browser and presents a data set as a parallel coordinate plot. The user can brush and highlight the polylines by creating and combining manipulable graphical constraints on them. The applet and all its features are used with a mouse. The Cars data set [8] is hard-coded in this proof-of-concept prototype. The data set contains the name and eight other attributes of 406 cars manufactured between years 1970 and 1982, originating from the Consumer Reports magazine.

Tobii T60 Eye Tracker integrated into a 17-inch flat screen monitor (1280x1024 pixel resolution) was used to capture eye movement data in the tests. Tobii Studio software (v1.1.25) was used to set up, execute, and analyze the test. The test computer had MS Internet Explorer 7 Web browser running under MS Windows XP operating system. In addition to Tobii equipment, a USB Web camera was used to record audio from the test situation.

The eye movement and mouse data was post-processed in Tobii Studio into fragments containing the tasks, and Areas of Interest (AOI, Figure 2) were defined. Then the fixation data based on the AOIs was computed and with videos clips containing a gaze path overlay were produced.

The fixation filter settings used in the analysis were Tobii Studio default values, where the maximum radius of fixation is 50 pixels and the minimum fixation duration is 100 ms – i.e., fixations less than 50 pixels apart in space and 100 ms apart in time are fused together. We used in the analysis the following eye movement metrics. The last two of these are commonly used [3], whereas the first is specific to our analysis:

Fixations Before: the number of fixations before fixating on a certain AOI.

Fixation Duration Mean: the average duration of fixations on a certain AOI.

Fixation Count: the number of fixations on a certain AOI.

These metrics allow us to inspect how the participants' attention was distributed, assuming that their attention is where they are looking at. For each task, we can easily define the set of AOIs that the user must at least look at to carry out the task. Then we can inspect from the gaze data how many fixations they accumulated before they entered one of the mandatory AOIs (Fixations Before), how their fixations are distributed between mandatory and non-mandatory areas (Fixation Count). and how long are the fixations in these areas (Fixation Duration Mean).

The participants performed the following nine tasks:

0. How many American cars are there in the data set?
1. How many cars have a four or six cylinder engine?
2. What is the average mileage for the six-cylinder cars?
3. How would you describe the cars that weigh over 4500 pounds?
4. What is the origin of the six-cylinder cars that were manufactured in 1971?
5. Which Japanese cars have the best acceleration?
6. What else is common to the most powerful, best accelerating, and heaviest cars in the data set?
7. How many non-American cars are there in the data set?
8. What is the most common number of cylinders for cars manufactured in 1973?

Task 0 was always performed first as a practice task but the participants did not know this. The presentation order of the subsequent eight tasks was counter-balanced using a balanced Latin Square design to eliminate order effects. The tasks were selected from the 16 tasks used in the previous experiment (except task 7 that was added for this study), based on the task types (simple selection, multiple selections, and exploratory). All tasks except 3 and 6 had an unambiguous answer.

3.3 Procedure

In the beginning of the session, the purpose and outline of the experiment were explained to the participant. Next, the participant was asked to sign an informed consent form, and fill in a background information questionnaire. At this point, the participant was also told that the used data set included information about cars manufactured between years 1970 and 1982.

The participant was seated before the eye tracking device at a distance of 50 cm and requested to be as stationary as possible throughout the experiment. Next, the eye tracker was calibrated using a standard Tobii calibration set of 9 fixation points. The calibration was repeated as necessary. After the calibration, the Web browser with PCE applet was launched and the participant was asked to describe what they see. Finally, the participant was allowed to experiment with the applet for a maximum of three minutes.

Following this experimentation, the participant was introduced to fundamentals of parallel coordinates: they are a graphical way to represent data sets; they include vertical parallel axes, which represent the attributes; and unlike orthogonal coordinates, parallel coordinates present multidimensional data in one view. Also some basic information concerning the PCE applet was provided: selections on the axes are made by "brushing", the selections can be moved by dragging and resized by dragging the triangles on each end of the selection (see Figure 2); "Background" option makes the unselected cases visible; "Remove" button clears all selections made; and "Combiner" menu provides options, like AND and OR, to combine selections. The participant was given a list of attributes which were to be used through the whole test. Before moving on to the actual tasks the participant was advised to read aloud the task and also to tell the answer aloud. Thinking aloud was not required.

The tasks were presented one by one below the PCE applet in the Web browser (Figure 2, the TASK area). The participant used the "Next task" button to move to the next task upon task completion. Tasks were considered to have started when the task description and the applet were visible and ended when the participant clicked on the "Next task" button. After performing all tasks, the participant was asked to fill out a questionnaire about subjective opinions concerning the applet, its usage, and the tasks. Finally, the participant was provided with the compensation.

4 Results

According to the pre-test questionnaire, the participants examined graphical presentations once a week or less frequently, and formed graphical presentations once a month or less frequently. Only one of the participants said he confidently knows what a parallel coordinate plot is.

Participants were requested to describe what they saw in the parallel coordinate visualization before carrying out tasks. Two of the participants had seen parallel coordinate visualizations before and were familiar with the representation. Other participants described the visualization as follows:

- "Some kind of a hassle in a tabular form . . . some values about cars?"
- "This is hard to find out. . . no way to acquire a quick overview . . . probably there are some variable names up there, and the values are shown at the lines?"

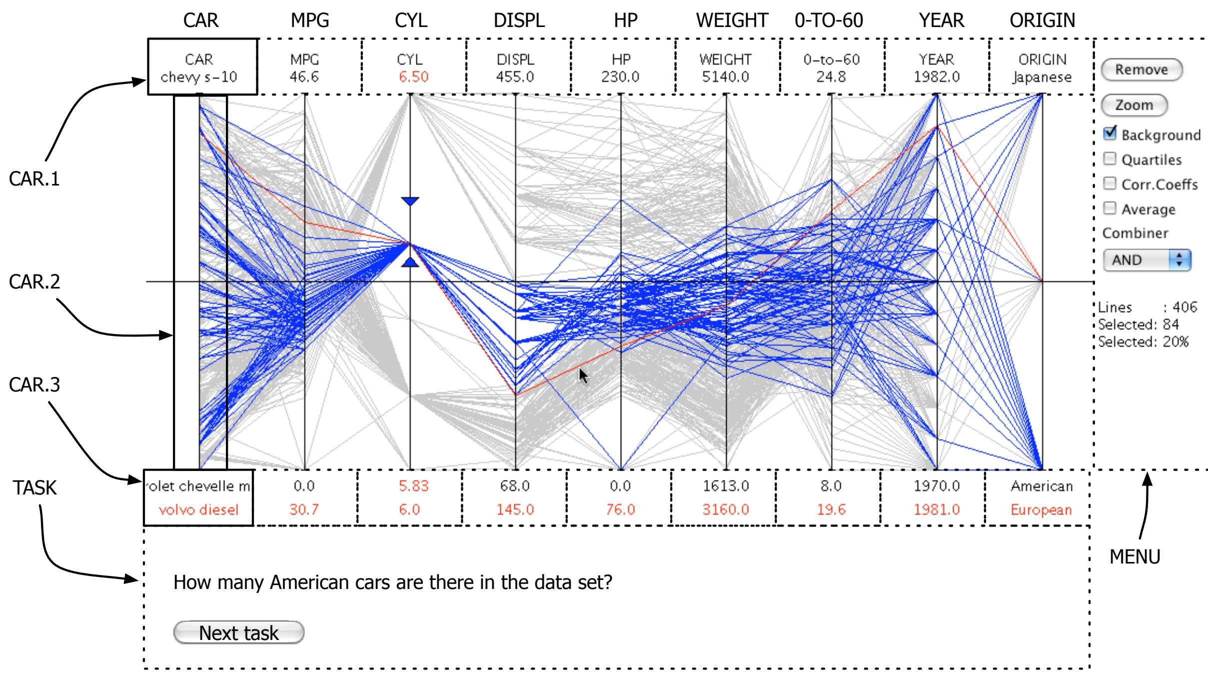

Figure 2: Areas Of Interest (AOI) used in the analysis and their naming conventions. Each area is named after the axis and divided into three parts: top of the axis (1), axis itself (2), and bottom of the axis (3), e.g. CAR.2 is the first vertical axis from the left. In addition, shows a visual query where all the six-cylinder cars have been highlighted (triangles at the CYL axis) and the one with the smallest engine displacement has been selected (a Volvo diesel, values shown below the axes labels).

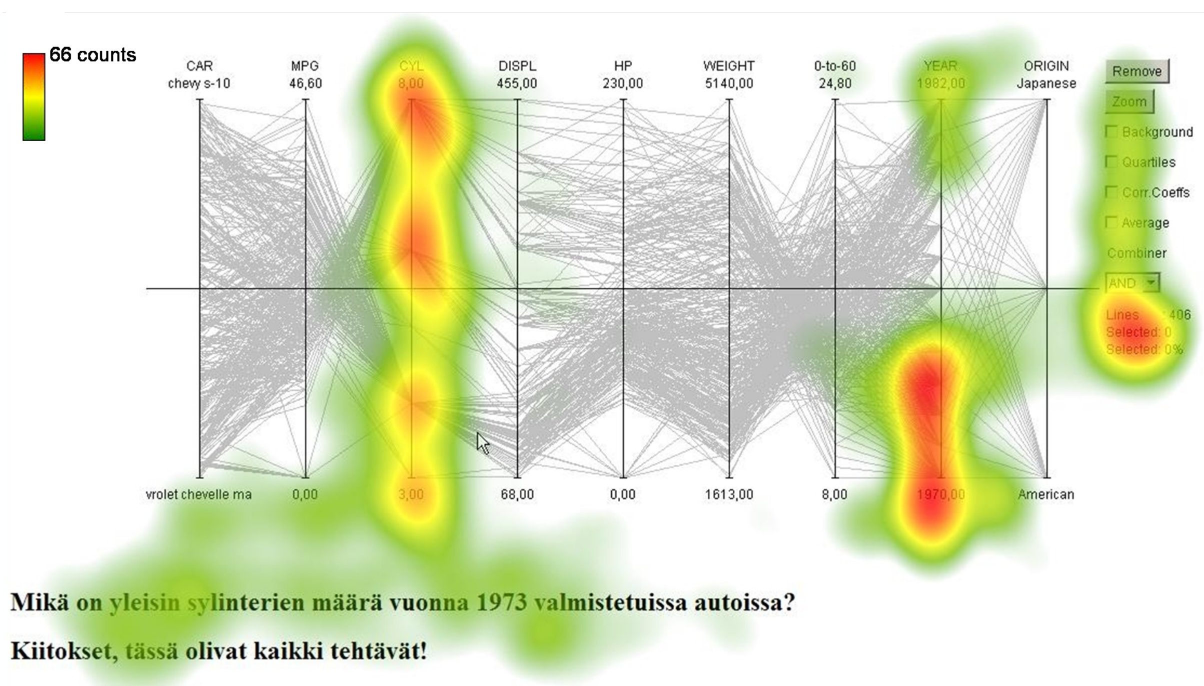

Figure 3: Heatmap of fixation counts for task 8, “What is the most common number of cylinders for cars manufactured in 1973?”

- “This is some kind of line diagram . . . every vertical and horizontal line is some kind of a measurement? Seems to be data about cars.”

- “This is a mess, I do not understand it at all!”

- “There is some kind of a graph of each car . . . according to those characteristics?”

As an example of a fixation heatmap, Figure 3 shows the heatmap for task 8, or “What is the most common number of cylinders for cars manufactured in 1973?”. Areas where users looked at are darker (ranging from green to red via yellow in color).

Figures 4–6 show the values of three aspects of fixation data: the median number of fixations occurring before the participant fixated within one of the AOIs for the first time, the fixation duration means (in seconds) in an AOI, and the median count of fixations within an AOI. These metrics are represented as AOI by Task balloon plots (package gplots, [7]) where numbers are replaced with correspondingly sized circles.

In Figure 4, the gray background in a cell indicates an AOI that must be first attended to solve a task, i.e. in making the initial query. In Figures 5 and 6 the gray background shows all the AOIs that must be attended to in a course of the task. In exploratory tasks, T3 and T6, that encompasses most of the areas of interest.

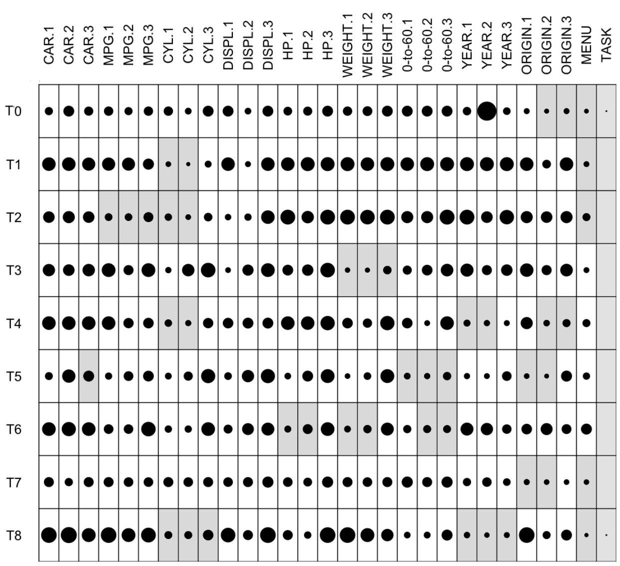

Figure 4: Balloon plot for AOI by Task: Medians of fixations before entering an AOI (smaller value is better). The mandatory AOIs for a task have a grey background.

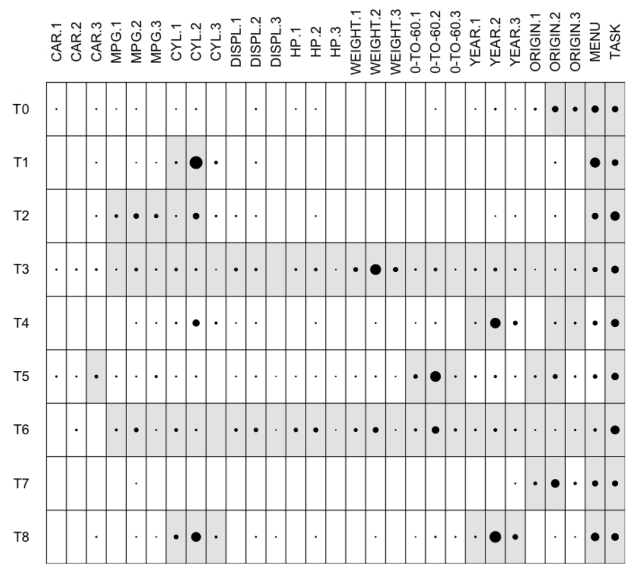

Figure 5: Balloon plot for AOI by Task: Fixation duration means in an AOI. The mandatory AOIs for a task have a grey background.

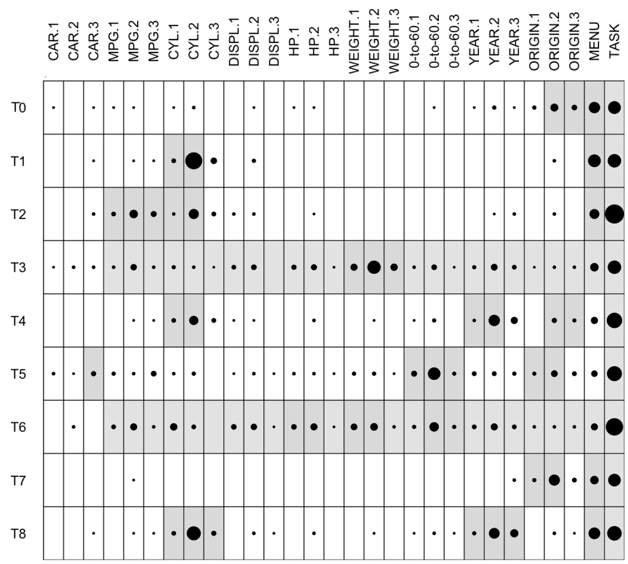

Figure 6: Balloon plot for AOI by Task: Medians of fixation counts in an AOI (larger value is better). The mandatory AOIs for a task have a grey background.

Table 1 presents a summary of the fixation data per task: the total sum of fixation durations on a task, the sum of durations on the mandatory areas, and the percentage of time spent in mandatory areas.

Table 1: Summary of fixation data per tasks.

Task	Total fixation time	Time in mandatory AOI's	% of total
T0	369,6	277,5	75,1
T1	777,5	653,3	84,0
T2	656,3	586,7	89,4
T3	825,3	438,5	53,1
T4	790,2	436,7	55,3
T5	788,3	495,6	62,9
T6	1106,6	644,2	58,2
T7	299,7	252,5	84,3
T8	857,9	788,9	92,0

5 Discussion

The descriptions of the parallel coordinate plot given by the participants prior to interacting with it are in line with our earlier experiences. Those who had seen the visualization before did not have problems recalling the idea, and some of those who saw it for the first time considered it "messy" or "a hassle".

The fixation data in Figures 3–6 shows convincingly that the participants did mainly look at the correct, task-specific AOIs. As an example, this is apparent when looking at the data for task 8, or "What is the most common number of cylinders for cars manufactured in 1973?". In order to complete the task the user has to look at the AOIs of YEAR.1–3, CYL.1–3 and MENU. The year 1973 has to be selected from the YEAR.2 area, the cylinder number has to be brushed from the CYL.2 area, and finally, the amount of cars having this number of cylinders has to be read from the MENU area. In addition, it might be necessary to check the axes names (YEAR.1, CYL.1) and to verify the limits of current selections (CYL.1/CYL.3 and YEAR.1/YEAR.3). As an overview, Figure 3 shows the fixation heatmap for task 8 over all participants. It is evident that their visual attention is where it should be.

Figures 4–6 shed more light on the visual perception process for task 8. First, Figure 4 shows that the important areas were found quickly. For task 8 (as for essentially all tasks), the balloons with grey background are small, meaning that the gaze did not wander around in other areas before hitting the areas that needed to be attended to in order to solve the task. Figure 5 shows that the middle part of the axes drew longer fixations than either the top or bottom area, represented by the bigger balloons. This reflects the fact that the interactive part in the middle required more intense attention than the static values at the ends, which could be quickly read. Finally, Figure 6 shows that the fixations in the important areas were not only longer; there were also more of them, yet another indication that participants paid attention to where they were supposed to.

One interesting detail is our rehearse task (T0) which all participants did as their first task, outside the counterbalancing. Data for the task 0 is included in Figures 4–6, and it shows that the participants paid attention to the correct parts of the visualization right in the beginning, after very short training. This supports the idea that the immediate usability of parallel coordinate visualizations is quite good.

The data also nicely reflects the characteristics of the data and the tasks. For tasks 3, 5, and 6, Figure 6 shows that there are fixations over most AOIs, even if the counts for the task-critical areas are the highest. For tasks 3 and 6 this follows naturally from the formulation of the tasks, which calls for a free-form characterization of the result set and encourages exploration of several attribute values. In the case of task 5 this is an indication of the conceptual challenge caused by the acceleration attribute, leading participants to look for confirmation for their interpretation from the values of the other attributes.

This experiment and earlier research provide evidence that even the first-time users of parallel coordinate visualizations quickly learn the concept and are able to make at least simple visual queries. But why does this visualization technique work so well when users initially describe it as complex and even unappealing? One explanation offered by cognitive psychology [14, p. 41] is in the mechanisms of visual search: the use of elementary shapes (lines) and their orientation (vertical axes vs. often non-vertical data items), motion (as the constraint 'sliders' are moved), and the spatial grouping (patterns formed by the selected polylines). The parallel coordinate visualization takes advantage of these basic mechanisms. Another, interesting perspective is provided by evolutionary psychology [5][14, p. 36]: our brain is adapted to the conditions of our most long-term and primary environment of evolutionary adaptedness: the African savanna. We may be extremely sensitive to predators lurking behind trees – or data items appearing behind parallel axes.

Conclusions

Visualizations based on parallel coordinate plots are often regarded as an expert-only technique; too difficult for a casual user to take advantage of. The experiment described in this paper – as well as the earlier experiment – contradict this assumption, and show that the immediate usability of parallel coordinate visualizations is surprisingly good. Parallel coordinate visualizations should certainly be among those tools that are offered to office workers once the mainstream software houses realize the benefits of information visualization approach.

Acknowledgments

This work was supported by the Academy of Finland (project 1129300).

References

[1] A. Inselberg. The plane with parallel coordinates. *The Visual Computer (Special issue on Computational Geometry)*, 1:69–97, 1985.

[2] A. Inselberg. *Trends in Interactive Visualization*, chapter Parallel Coordinates: Interactive Visualization for High Dimensions, pages 49–78. Springer, 2009.

[3] R. J. K. Jacob and K. S. Karn. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, J. Hyönä, R. Radach and H. Deubel (Eds.), pages 573–605. Elsevier, Amsterdam, 2003.

[4] J. Johansson, C. Forsell, M. Lind, and M. Cooper. Perceiving patterns in parallel coordinates: determining thresholds for identification of relationships. *Information Visualization*, 7(2):152–162, 2008.

[5] S. Kanazawa. The Savanna Principle. *Managerial and Decision Economics*, 25:41–54, 2004.

[6] P. Pirolli, S. K. Card, and M. M. Van Der Wege. The effect of information scent on searching information: visualizations of large tree structures. In *AVI '00: Proceedings of the working conference on Advanced visual interfaces*, pages 161–172. ACM, 2000.

[7] R Development Core Team. R: A language and environment for statistical computing. `http://www.R-project.org/`, 2009.

[8] E. Ramos and D. Donoho. The 1983 ASA data exposition dataset: Cars. `http://lib.stat.cmu.edu/datasets/cars.data` (accessed 27 Feb 2009), 1983.

[9] H. Siirtola. Direct manipulation of parallel coordinates. In *IV'2000: Proceedings of the International Conference on Information Visualization*, pages 373–378. IEEE Computer Society, 2000.

[10] H. Siirtola. *Interactive Visualization of Multidimensional Data*. PhD thesis, University of Tampere, 2007.

[11] H. Siirtola and K.-J. Räihä. Interacting with parallel coordinates. *Interacting with Computers*, 18(6):1278–1309, 2006.

[12] B. Spence, M. Witkowski, C. Fawcett, B. Craft, and O. de Bruijn. Image presentation in space and time: errors, preferences and eye-gaze activity. In *AVI '04: Proceedings of the working conference on Advanced visual interfaces*, pages 141–149. ACM, 2004.

[13] R. Stiefelhagen, J. Yang, and A. Waibel. Estimating focus of attention based on gaze and sound. In *PUI '01: Proceedings of the 2001 workshop on Perceptive user interfaces*, pages 1–9. ACM, 2001.

[14] C. Ware. *Visual Thinking for Design*. Morgan Kaufmann Publishers, 2008.