Presented at the IEEE Symposium on Information Visualization (InfoVis 2003), Seattle, October 19-21, 2003

Visualization of Large-Scale Customer Satisfaction Surveys Using a Parallel Coordinate Tree

Dominique Brodbeck
Macrofocus GmbH
dominique.brodbeck@macrofocus.com

Luc Girardin
Macrofocus GmbH
luc.girardin@macrofocus.com

Abstract

Satisfaction surveys are an important measurement tool in fields such as market research or human resources management. Serious studies consist of numerous questions and contain answers from large population samples. Aggregation on both sides, the questions asked as well as the answers received, turns the multi-dimensional problem into a complex system of interleaved hierarchies. Traditional ways of presenting the results are limited to one-dimensional charts and cross-tables. We developed a visualization method called the Parallel Coordinate Tree that combines multi-dimensional analysis with a tree structure representation. Distortion-oriented focus+context techniques are used to facilitate interaction with the visualization. In this paper we present a design study of a commercial application that we built, using this method to analyze and communicate results from large-scale customer satisfaction surveys.

Categories and subject descriptors: H.5.2 [Information Interfaces and Presentation (e.g., HCI)]: User Interfaces - *Graphical user interfaces (GUI)*; H.4.3 [Information Systems Applications]: Communications Applications - *Information browsers*; J.1 [Administrative Data Processing] - *Marketing*.
General terms: Design, Human Factors.
Additional Key Words and Phrases: parallel coordinates, focus+context, hierarchical data, satisfaction survey.

1. Introduction

Business process automation, together with the insight that backwards-looking financial metrics are not a sufficient indicator for the state of an enterprise, have led to the emergence of performance measurement management systems such as balanced scorecards [Kaplan and Norton 1992]. Tools to monitor and interpret all the different indicators collected across the enterprise are needed. One of the important aspects that is measured by such systems is customer satisfaction, based on the assumption, that the perceived service quality is a key factor for customer acquisition and retention. The most direct way to measure customer satisfaction is to ask the customers themselves. When done in a statistically sound way, then the results are accurate, detailed, and enjoy high acceptance by the stakeholders of such a survey.

Public transport in the metropolitan area of Zurich, Switzerland is provided by a heterogeneous mix of private companies and government agencies. To fuse them into a coherent public transport network, they are governed by a management holding (ZVV) that is responsible for strategic marketing and financing.

Service quality in the case of public transport is of high importance, as it not only influences the image of the service providing companies, but also the basic attitude towards public transport. In order to assess, analyze and monitor service quality, ZVV performs a large-scale study of customer satisfaction for the whole public transport network every year. The study forms the basis for performance-based financial compensation of the network partners, as well as to spot concrete service quality problems. Transparent and complete communication of and access to the results is therefore essential, to support decisions makers in the public transport area and rationalize their decisions.

So far, the results were presented as one-dimensional charts and cross-tables, both during oral presentations as well as in the form of a thick folder, containing the details that are deemed relevant. Next to the obvious problems that one expects from such an approach to represent a large and complex data set (i.e. lack of overview, inefficient access to details, numbers and tables causing cognitive overload), there is the problem that it is limited to predefined tables and charts, and therefore can not answer questions that were not anticipated.

We were given the task of developing an interactive visualization application - SurveyVisualizer - that addresses these problems. The following design objectives were defined:

- **Show the data at various levels of detail and aggregation:** Provide an overview of structures and trends, allow comparison of analysis elements (e.g. young people vs. seniors). Display details in context.
- **Quick access to all the results:** Provide possibility to select elements of all the different analysis groups (e.g. customer segment, market region, service provider). Ensure the integrity of the numeric values (statistical significance).
- **Easy to use and understand by non-experts:** Stakeholders should be able to use the application for interpretation and analysis of the survey results, without any knowledge of statistics or complex data analysis software. The application should be simple and expressive enough to be used “live“ in strategic business meetings.

In the following section we will provide details about the survey and the data that is collected, describe the visual encoding that we chose, and explain the different interaction techniques that tie everything together. We will explain the design decisions that we took to meet the above objectives, and discuss the lessons learned.

2. The Survey

The data for the survey is collected through the use of questionnaires that are distributed to the customers of the public transport network. The questionnaire contains 89 questions such as:

> *If you think of cleanliness, how well does the bus line that you are using fulfill your expectations in terms of the inside of the vehicle?*

The questions - also called quality criteria - are then aggregated into 23 quality dimensions (e.g. network quality, ticketing, cleanliness, security, reliability). They represent the level of satisfaction with a whole group of questions pertaining to a particular issue. The quality dimensions themselves are further aggregated into three different customer satisfaction indices, reflecting the different areas of responsibility (tasks for which A: the management holding is responsible, B: the service providers, and C: the operators). The first two indices have two definitions, reflecting different aggregation methods. Aggregation is performed as a weighted average, where the weights are determined by an underlying statistical model. The hierarchical structure that is defined by this process is shown in Figure 1.

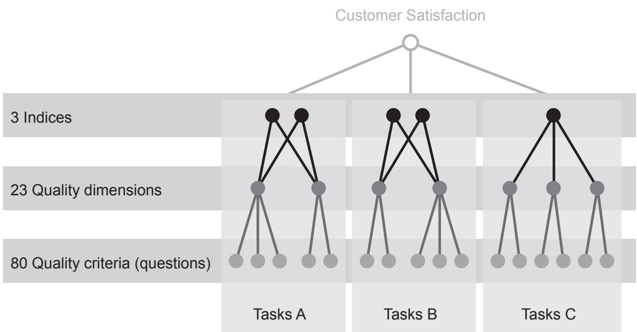

Figure 1: The data structure of the survey. 89 questions are aggregated into 23 quality dimensions, and then into 3 satisfaction indices (the first two with alternate definitions), adding up to 8’424 values ((89+23+5) nodes * 72 attributes) that need to be represented and accessed for each year of the study.

In addition, the questionnaire contains a number of demographic questions that are used to associate the responses with six different analysis groups. The groups are further subdivided into a number of analysis elements (shown in parentheses): benchmarks (various global averages), market segments (geographic regions), service providers and operators (individual companies and government agencies), customer segments (age and activity groups), means of transportation (bus, tram, train). The analysis groups encompass a total of 72 analysis elements, some of which are also structured as hierarchies (e.g. buses total > city buses, commuter buses, long-distance buses). These analysis elements are the actual data objects that are visualized. Individual questionnaires are not shown directly, they are used as raw material for the aggregation.

Roughly 5’000 questionnaires are collected for one survey. After all the aggregation and averaging is done, the final spreadsheet contains close to 10’000 values. Surveys are repeated yearly and have been conducted for the past three years.

3. Visual Encoding: The Parallel Coordinate Tree

Hierarchies are a common strategy to structure large amounts of information into nested manageable chunks. We find them for example in tables of contents, file directories, or corporate organization charts. Hierarchical structures can be conceptualized as inverted trees with the most important, common, or original node at the top, and several other nodes below each node in a recursive way. Trees have two properties: the structure of how the nodes are linked, and the content or the attributes describing each node. There are many ways to visually encode trees, depending on whether the emphasis lies on the structure or the content of the tree.

Tree structures are commonly represented as node-link diagrams. While node-link diagrams are very intuitive, they don’t scale well for large number of nodes or for unbalanced trees. Enhancements that have been suggested include distortion-oriented techniques [Lamping et al. 1995], use of interaction and dynamic pruning [Kumar et al. 1997], or animated 3-dimensional representations [Robertson et al. 1993] to compress more nodes into the available screen space. An additional problem of node-link diagrams is the representation of node content. The obvious approach is to use primary visual cues like shape and color to encode node attribute values, or more sophisticated icons [Kleiberg et al. 2001]. This is limited to only a few dimensions however, before the design space becomes overloaded.

Alternatives to node-link diagrams that emphasize node content, are enclosure-based representations, namely Tree-Maps [Johnson and Shneiderman 1991]. A node attribute is directly mapped to area by recursively subdividing the available screen space. This method works very well to visualize attribute values of a large number of nodes. Color and other perceptual cues can be mapped to the shapes, but these approaches suffer from the same problem that only a very small number of attributes can be encoded simultaneously for each node in this way.

The hierarchy of the survey described in the previous section, represents a balanced tree with uniform depth, meaning that all the nodes have a similar number of child nodes, and all the branches from the root node to the leaf nodes have the same length. While the structure is important, it is not overly complex. The attributes however, corresponding to the analysis elements of the survey, are high-dimensional (72 dimensions, one for each analysis element), but they are homogeneous in that they all describe perceived customer satisfaction on a scale from 0 to 100. Our challenge was therefore to represent the tree structure of the hierarchy, while at the same time allow for multi-dimensional analysis of the node attributes.

In order to achieve this, we chose a Cartesian node-link layout as a basis to represent the tree structure. Tree depth is mapped onto the vertical direction, and node position onto the horizontal direction. We represent each node in this layout by its own vertical axis. This naturally creates a parallel coordinate system [Inselberg and Dimsdale 1990] for each level of the hierarchy. To layout the axes, we start at the bottom level, arranging them laterally and equally spaced. The nodes of the next level up are positioned cen-

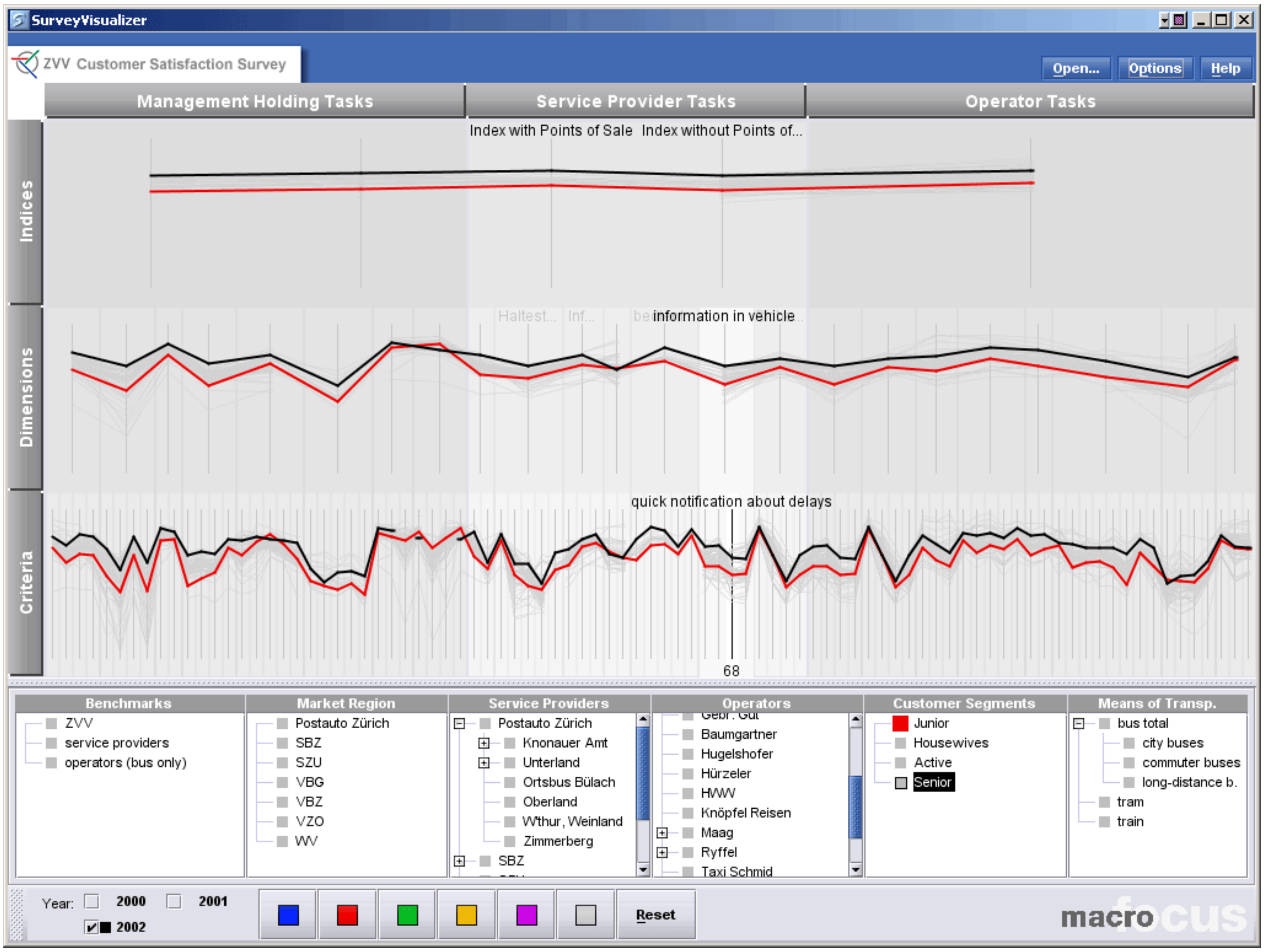

Figure 2: The SurveyVisualizer application. The top view contains the Parallel Coordinate Tree. The tree structure is shown as a combination of a node-link diagram and an enclosure-based layout. Each level forms a horizontal parallel coordinate system. The values of a particular analysis element are shown as polygonal lines on all levels simultaneously (red and black lines). The second view at the bottom shows the different analysis groups (from left to right: benchmarks, market regions, service providers, operators, customer segments, means of transportation). Both views are coordinated and elements can be selected in either one of them.

tered above the child nodes that they represent. We continue this way until the top is reached. The hierarchical structure is emphasized by recursively underlaying each node and its children with a grey box whose intensity varies with the depth of the node (from grey at the root to white at the leaf level). The background coloring changes dynamically as the structure is probed with the mouse pointer (see section 4).

The root node is ignored, as it would be a very inefficient use of space if it was assigned its own level. We first inserted the root node as just one additional index on the first level (without child nodes). While this did work, it did not add enough interesting information and thus was left away for simplicity reasons.

A particular analysis element is now represented by connecting all the values on the respective node axes of a level with a polygonal line. Missing values (caused by insufficient statistical significance) are shown as gaps, and cases where both neighboring values are missing, are represented with a short horizontal dash (see black line at bottom left in Figure 2). All the analysis elements are always shown as grey lines in the background. This provides an overview of the ranges and spreads of the individual values for each node, and facilitates the detection of outliers. Clicking on a background line, selects that analysis element and paints it in a different color (see next section).

We call this visual encoding a Parallel Coordinate Tree (Figure 2). It combines the advantages of a familiar tree layout with the multidimensional analysis capabilities of parallel coordinates.

The data from different years can be combined and displayed simultaneously to analyze changes of satisfaction values over time. The lines of the older studies are then painted with decreasing saturation, creating a "fading out" effect. This method scales to about four to five different studies, which is sufficient for the purpose of this particular application. Figure 3 shows the results from two years. The top level tells us that satisfaction has increased for this particular group of tasks. Examining the more detailed level below, reveals that this was mainly due to a significant increase in the satisfaction with the handling of complaints. Looking at the

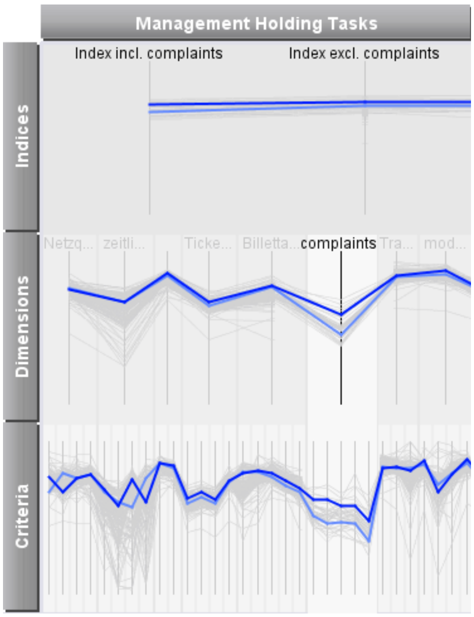

Figure 3: Results from studies of different years can be combined using a "fading out" effect (older years are colored with less saturation). Changes over time become quickly apparent on all aggregation levels.

individual questions at the bottom level, confirms that this seems to have been the case equally for all aspects. We further see that there are some interesting effects more towards the left, where the changes cancel each other.

A second view below the Parallel Coordinate Tree is used to visualize the analysis groups and its constituting analysis elements. The hierarchical structures of the analysis elements are represented as indented outlines. The two views are coordinated in terms of "selecting items to items" and "selecting items to navigating views" (selecting an item in the tree navigates the outline to the corresponding item) according to the taxonomy of [North and Shneiderman 1997].

4. Interaction Techniques

To navigate the Parallel Coordinate Tree, we chose a distortion-oriented technique [Leung and Apperley 1994]. A bifocal lens [Apperley et al. 1982] can be applied in both directions to emphasize either a particular level of the tree, or a specific group of questions. The lens offers predefined focus areas, corresponding to the headings at the top and along the left side of the tree. Clicking on one of these heading buttons, moves the focus to the selected range. The transition between these predefined foci is animated, to maintain context and temporal coherence, and to avoid the problem of change blindness [Rensink et al. 1997; Nowell et al. 2001]. Figure 4 shows the lens active in both directions. Note how the trend of the values outside of the focus area is still perceivable.

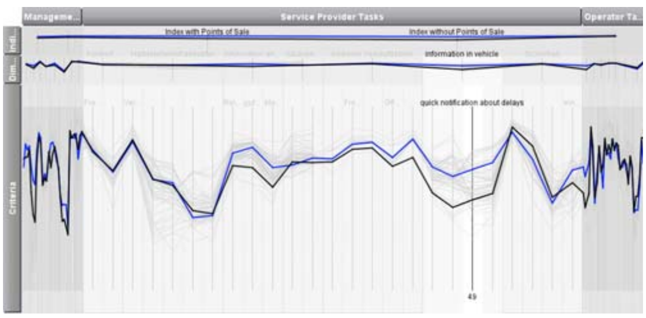

Figure 4: Navigating the tree using a bifocal lens. Details of the customer satisfaction of a particular company (black line) can easily be compared to the benchmark of the overall network (blue line). The context outside of the focus area is maintained.

A common technique to coordinate two views is brushing and linking [Becker and Cleveland 1987], where users can select objects in one view and the corresponding objects in all the other views are also automatically selected. Over the course of using and building many applications that support this technique, we developed a refined approach. We distinguish two modes of brushing and linking interaction (selecting, painting) that are coordinated among the two views, plus a transient exploration mode (probing):

- **Probing:** This mode is used to reveal more details about a node (i.e. full axis labels, numeric values, enclosing node structure). Probing is a transient operation. Moving the mouse pointer over a node, highlights that node. As soon as the mouse pointer is moved away, the highlighting disappears. It is used like a flashlight that examines a dark room.
- **Selecting:** This mode is used to mark analysis elements that are of short-term interest, in order to further examine or perform operations on them (e.g. explore numeric values for individual questions, dimensions or indices). Clicking on an analysis element selects it and marks it in black.
- **Painting:** This mode is used to mark analysis elements that are of long-term interest, in order to use them as references for comparisons (e.g. compare various service providers against a benchmark). Selected analysis elements can be painted by pressing one of the predefined color buttons in the toolbar at the bottom of the window. Analysis elements remain painted until they are reset explicitly.

5. Conclusion

We described SurveyVisualizer, an application to communicate and interpret the results from a large-scale customer satisfaction survey. It provides the overview of the structure of the survey and the trends of the results. Details are embedded in this context and can easily be accessed and compared. SurveyVisualizer is a very effective tool to replace the traditional tables-and-charts-based way of presenting such survey results. We do not conclude this on the basis of formal user studies, but rather from anecdotal user feedback and the fact that it enjoys encouraging commercial success.

We introduced the Parallel Coordinate Tree, a visual encoding to represent multi-attribute hierarchies. Preliminary experiments have shown, that this method also works with somewhat unbalanced trees with up to a few hundred leaves and about 5 hierarchical levels deep. Another topic for future work is how the ability of parallel coordinates to handle non-uniform scales can be used, to visualize surveys with less homogeneous attribute ranges than in the present example. This makes it an interesting candidate for visualizing other problems in areas such as accounting, OLAP analysis, or product catalogs. A combination with integrated dynamic queries to allow for deeper analyses, is envisioned.

6. References

APPERLEY, M.D., TZAVARAS, I. AND SPENCE, R. 1982. A Bifocal Display Technique for Data Presentation. In *Proceedings of Eurographics'82*, Conference of the European Association for Computer Graphics, pp. 27-43.

BECKER, R. AND CLEVELAND, W. 1987. Brushing Scatterplots. *Technometrics, 29 (1987)*, pp. 127-142.

INSELBERG, A. AND DIMSDALE, B. 1990. Parallel Coordinates: A Tool for Visualizing Multidimensional Geometry. In *Proceedings of IEEE Conference on Visualization '90*. pp. 361-378.

JOHNSON, B. AND SHNEIDERMAN, B. 1991. Tree-Maps: A Space-Filling Approach to the Visualization of Hierarchical Information Structures. In *Proceedings of IEEE Symposium on Information Visualization '91*, pp. 275-282.

KAPLAN, R. AND NORTON, D. 1992. The Balanced Scorecard - Measures that Drive Performance. *Harvard Business Review*, Jan-Feb 1992, pp. 71-79.

KLEIBERG, E., VAN DE WETERING, H., AND VAN WIJK, J. J. 2001. Botanical Visualization of Huge Hierarchies. In *Proceedings of IEEE Symposium on Information Visualization 2001*, pp. 87-94.

KUMAR, H. P., PLAISANT, C., AND SHNEIDERMAN, B. 1997. Browsing Hierarchical Data with Multi-Level Dynamic Queries and Pruning. *International Journal of Human-Computer Studies, 46(1)*, pp. 103-124.

LAMPING, J., RAO, R., AND PIROLLI, P. 1995. A Focus+Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies. In *Proceedings of CHI '95*, ACM Conference on Human Factors in Computing Systems. ACM, New York.

LEUNG, Y. K. AND APPERLEY, M. D. 1994. A Review and Taxonomy of Distortion-Oriented Presentation Techniques. *ACM Transactions of Computer-Human Interaction, Vol 1, No. 2. June 1994*, pp. 126-160.

NORTH, C. AND SHNEIDERMAN, B. 1997. A Taxonomy of Multiple Window Coordinations. University of Maryland, College Park, *Dept. of Comp. Science Tech. Report #CS-TR-3854*.

NOWELL, L.T., HETZLER, E. G., AND TANASSE, T. 2001. Change Blindness in Information Visualization: A Case Study. In *Proceedings of IEEE Symposium on Information Visualization 2001*, pp. 15-22. 2001.

RENSINK, R. A., O'REGAN, J. K., AND CLARK, J. J. 1997. To See or Not to See: The Need for Attention to Perceive Changes in Scenes. *Psychological Science, 8*, pp. 368-373.

ROBERTSON, G. G., CARD, S. K., AND MACKINLAY, J. D. 1993. Information Visualization Using 3D Interactive Animation. *Communications of the ACM, 36(4)*, pp. 57-71.