Evaluating Visualizations: Do Expert Reviews Work?

Melanie Tory

Imager Lab
University of British Columbia

Torsten Möller

Graphics, Usability and Visualization Lab
Simon Fraser University

Visualization research generates beautiful images and impressive interactive systems. Such developments make fascinating demos, but how do we know if they are actually useful for real people doing real tasks? If the interaction is awkward or we have not carefully considered users' needs, even the most well intentioned and technically developed visual display will be ineffective.

Emphasis on evaluating visualizations is growing. User studies of perceptual phenomena related to visualization and comparisons of visualization tools are becoming hot topics in the visualization literature [7][13]. But along the way, researchers are discovering that user study design is rarely straightforward.

When formal user studies fail

Formal laboratory user studies can effectively evaluate visualizations. Recently, many such studies have appeared (e.g., [7]). We support and encourage these experiments, and have carried out some of our own (e.g, [14]). At the same time, we feel they are not ideal for every situation.

Designing and running a controlled experiment requires substantial time and resources. A clear objective, controlled laboratory setting, and strict/limited tasks are essential; otherwise clear conclusions cannot be drawn. Thus, formal laboratory user studies may be inappropriate during an exploratory phase of research when clear objectives and variables may not be defined. Furthermore, formal laboratory user studies often focus on perceptual or simple cognitive tasks. High level cognitive tasks (e.g., thinking, deciding, and exploring ideas) are important activities, yet performance of these tasks is difficult to measure objectively and quantitatively.

Possible Alternatives

In these situations, how should we proceed? Clearly, just asking a few friends their opinions is not sufficient, even if they are knowledgeable. Besides their response being biased, we may miss valuable information if we do not structure our investigation in a meaningful way.

Alternative evaluation techniques have been used successfully in human-computer interaction (HCI), including focus groups, field studies, and expert reviews. These methods tend to produce qualitative results and require fewer participants than controlled experiments. In this paper, we focus on expert reviews. Based on our experience and the experience of others, we argue that they are a valuable way to evaluate visualization techniques.

Expert reviews [8][10] are commonly used to assess interface usability (and other, more general ideas [2]). The method involves only a few usability experts and does not require bulletproof test software or strict performance measures. Five usability experts typically find 75% of a system's usability problems [10]. Compare this with up to 50 participants for formal laboratory user study.

We used expert reviews for two applications. Specifically, we examined *heuristic evaluation* (where experts evaluate a tool with respect to pre-defined heuristics).

Comparing widgets for specifying lights in a scene

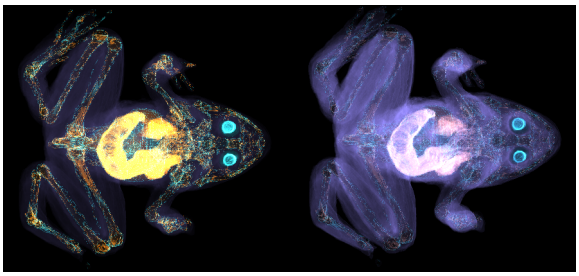

Figure 1: Spectral volume rendering of a frog displayed under two lighting conditions.

We first compared two interfaces for specifying lights in a spectral volume rendering tool. Details may be found in [1]. With spectral volume rendering, changing the light spectrum changes which parts are visible and how those parts appear (e.g., see Fig. 1). We compared a traditional interface consisting of one slider for each light to a new interface called the light dial (see Fig. 2).

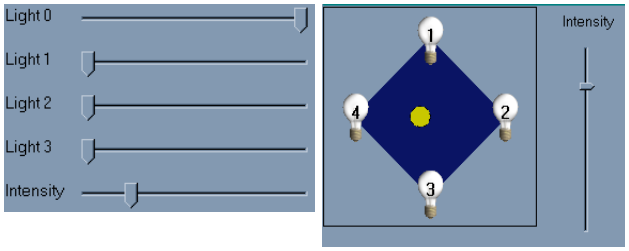

Figure 2: Multiple-sliders (left) and light dial (right) interfaces.

Three usability experts (HCI graduate students) took part. They had no prior knowledge of our project but were familiar with heuristic evaluation. We gave them descriptions of typical end users and tasks (exploring data, matching target images, and returning to previous settings). They reported advantages, disadvantages, and problems with the interfaces, and evaluated the interfaces with respect to several heuristics. They recorded written comments and then discussed them with us.

We felt interface heuristics could limit results to GUI elements (e.g., is it easy to exit the dialog box?) rather than visualization issues (e.g., how well do the interfaces support exploration?). Therefore, we based our heuristics on standard GUI heuristics [10], generic visualization tasks [11], and visualization tasks specific to our tool.

Our study uncovered many advantages / disadvantages of the interfaces and difficulties and misconceptions people might have. The light dial was usually faster and was better for quickly exploring data and moving back and forth between preset light combinations. However, it evoked the misconception that the spatial arrangement of lights represented their physical arrangement in the scene. Multiple-sliders were better for understanding exact light contributions and resetting them during an undo operation; note that a simple undo function might be a better way of providing this functionality.

We believe the usability experts had far more insight than ordinary user study participants into how the interfaces might be used and what problems might arise. However, only two experts examined how the interfaces would affect data analysis. These two experts specialized in medical imaging and were familiar with data display issues. Because of this, we believe that involving experts familiar with both HCI and visualization is important.

Comparing volume rendering interfaces

We then used expert reviews to compare two interfaces for volume data exploration: a table interface based on [6] (see Fig. 3) and a parallel-coordinates style interface [12] (see Fig. 4). We chose expert reviews because data exploration is a difficult task to assess quantitatively, and because we were in an exploratory phase of our research.

Based on our experience with the light widget study, we included experts with varied backgrounds: two HCI experts with data display experience, a volume visualization expert, a graphic design expert, and one end user (a medical imaging physicist). We expected this group to provide a wide variety of opinions and insight. Experts were given two sample tasks: (1) explore several data sets and (2) search for an identifiable object (a key).

Experts recorded their opinions and ratings in writing, like in the previous study. We also observed participants and discussed their opinions and comments throughout the evaluation procedure. This process generated much more qualitative information about when each visualization tool was useful, what changes might be valuable, and what problems or misconceptions might be common. For example, we could observe difficulties with setting transfer functions (colour and transparency levels) and inquire about why this task was challenging.

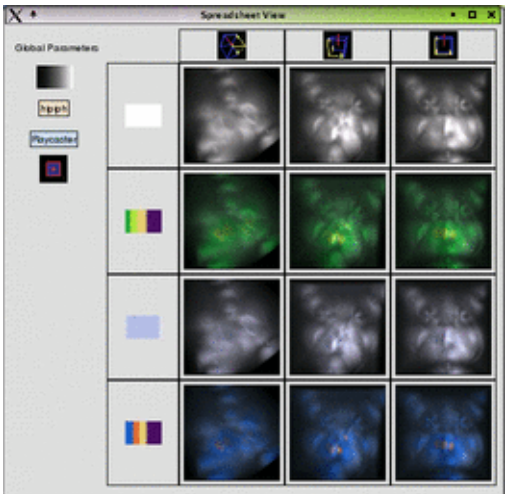

Figure 3: Table interface for volume exploration. The table shows 4 colour transfer functions and 3 rotations. Current values of other parameters (opacity transfer function, data set, renderer, and zoom) are displayed at the top left.

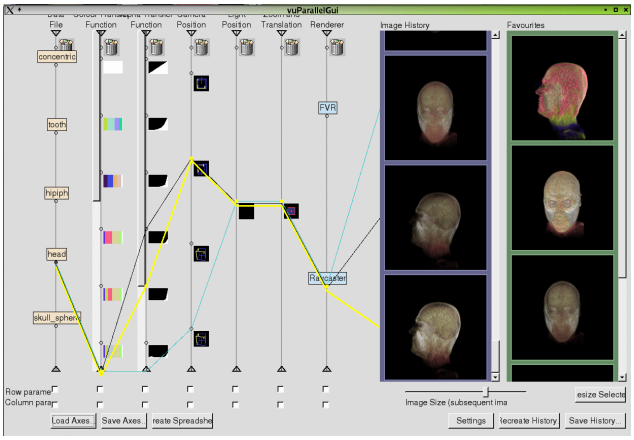

Figure 4: Parallel coordinates style interface for volume exploration. Each volume rendering parameter is shown on one axis. A history bar (blue) records all rendered images, and a favourites bar (green) holds images of interest. A polyline connects each image in the history to the parameters used to create it. The history bar here shows images of a head created with three opacity transfer functions and two rotations.

Our results showed that having both interfaces available at the same time was beneficial since they served different purposes. The parallel coordinates interface was useful for understanding what display options were available,

undoing operations, and manipulating display settings. The table was useful for quickly exploring many settings and for comparing settings side-by-side. In both interfaces, the small images were a disadvantage; evaluators suggested providing an option to interact with an enlarged image. Many additional results may be found in [12].

Challenges

Our biggest challenge was recording and analyzing all the observations. We plan to try recording the sessions on video (to ease the burden of notetaking) and experimenting with qualitative data analysis software tools. Nonetheless, the feedback we obtained was highly valuable and directly applicable to design. Furthermore, separating important major recommendations from the far more prevalent minor ones was fairly straightforward.

Other experiences with expert reviews

Other visualization researchers have also had positive experiences with expert reviews. Freitas *et al.* [3] uncovered advantages and disadvantages of a hierarchy visualization tool using heuristic evaluation. Gabbard *et al.* [4] found expert reviews valuable for evaluating early prototypes of a battlefield visualization. Most notably, Jackson *et al.* [5] evaluated flow visualization techniques using graphic design experts. Results of the expert review agreed with results of a formal laboratory user study, indicating that results from expert reviews are reliable.

When should we use expert reviews?

While expert reviews can provide quick and valuable insight into usability problems, they should not be used exclusively and should not replace user studies. Expert reviews often identify different problems than tests with end users, and may miss important issues. Furthermore, some of the "problems" found by experts may not present difficulties for end users [10]. Results of an expert review will depend on experts' qualifications, and opinions of experts are subjective and may vary [2]. For these reasons, we believe expert reviews should complement formal user studies. One possibility is to have experts evaluate early prototypes (formative evaluation), and then end users evaluate a refined version (summative evaluation) [4].

Conclusion

With a few modifications, expert reviews can generate valuable feedback on visualization tools. We recommend:

- Including experts with experience in data display as well as usability, and
- Developing heuristics based on visualization guidelines as well as usability guidelines.

Conducting an expert review is far faster than a formal user study, but can generate more useful feedback than "asking a few of your friends". Our experience, and that of others, indicate that expert reviews are valuable, particularly early in development when requirements may not be well established. We therefore encourage researchers to consider using the technique. At the same time, expert reviews should not be used exclusively, since experts may not fully predict actions of end users. Furthermore, we encourage more experimentation with this technique, particularly to develop a good set of visualization heuristics and to compare it with other methods. While our initial experience seems promising, much more experience is needed to determine detailed strengths and weaknesses of evaluation techniques in the field of visualization.

Acknowledgements

Thanks to our expert reviewers for their valuable feedback. This work was supported by the Natural Sciences and Engineering Research Council (NSERC).

References

[1] S. Bergner, T. Möller, M. Tory, and M.S. Drew, "A Practical Approach to Spectral Volume Rendering", *IEEE Trans. Visualization and Computer Graphics* (in print).

[2] *The Delphi Method*, http://www.iit.edu/~it/delphi.html.

[3] C.M.D.S. Freitas, P.R.G. Luzzardi, R.A. Cava, M.A.A. Winckler, M.S. Pimenta, and L.P. Nedel, "Evaluating Usability of Information Visualization Techniques", *Proc. Symp. Human Factors in Computer Systems*, 2002.

[4] J.L. Gabbard, D. Hix, and J.E. Swan II, "User-Centered Design and Evaluation of Virtual Environments", *IEEE Computer Graphics & Applications*, 51-59, Nov./Dec. 1999.

[5] C.D. Jackson, D. Acevedo, D. Laidlaw, F. Drury, E. Vote, and D. Keefe, "Designer-critiqued Comparison of 2D Vector Visualization Methods: A Pilot Study", *SIGGRAPH Sketch*, 2003.

[6] T.J. Jankun-Kelly and Kwan-Liu Ma, "Visualization Exploration and Encapsulation via a Spreadsheet-like Interface", *IEEE Trans. Visualization and Computer Graphics*, vol. 7, no. 3, 275-287, 2001.

[7] R. Kosara, C.G. Healey, V. Interrante, D.H. Laidlaw, and C. Ware, "User Studies: Why, How, and When?", *IEEE Computer Graphics & Applications*, 20-25, July/Aug. 2003.

[8] R.L. Mack and J. Nielsen, "Usability Inspection Methods: Executive Summary", Reprinted in R.M. Baecker, J. Grudin, W.A.S. Buxton, and S. Greenberg, eds., *Readings in Human Computer Interaction: Toward the Year 2000*, Morgan Kaufmann Publishers, San Francisco, 170-181, 1995.

[9] L. Nigay and F. Vernier, "Design Method of Interaction Techniques for Large Information Spaces", *Proc. Advanced Visualization Interfaces*, 37-46, 1998.

[10] J. Preece, Y. Rogers, and H. Sharp, *Interaction Design: Beyond Human-Computer Interaction*, John Wiley and Sons, New York, 407-424.
[11] B. Shneiderman, "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations", *Proc. of IEEE Symp. Visual Languages*, 336-343, Los Alamos, Sept. 1996.
[12] M. Tory, S. Potts, and T. Möller, "A Parallel Coordinates Style Interface for Exploratory Volume Visualization", *IEEE Trans. Visualization and Computer Graphics*, (in print).
[13] M. Tory and T. Möller, "Human Factors in Visualization Research", *IEEE Trans. Visualization and Computer Graphics*, vol. 10, no. 1, 72-84, Jan./Feb. 2004.
[14] M. Tory, T. Möller, M.S. Atkins, and A.E. Kirkpatrick, "Combining 2D and 3D Views for Orientation and Relative Position Tasks", *Proc. CHI 2004*, 73-80.

Sidebar: Conducting a heuristic evaluation

This sidebar describes the basic procedure for conducting a heuristic evaluation. The steps are based on well documented HCI methods, with modifications based on our own experience. For more details of the standard HCI approach, see Mack and Nielsen [8] or Preece *et al*. [10].

Step 1: Preparation

Careful preparation leads to better results and shows the experts that you respect their time.

Develop descriptions of a few typical users and tasks.

Providing experts with these descriptions will help them determine how well the tool will meet users' needs. (Usability experts may not be familiar with your target users.) The descriptions should be specific rather than general; for example, "Nancy is a physics professor who wants to illustrate the structure of an atom to her first year physics class..." instead of "typical users are physicists who study atomic structures...". A visualization tool may work well for some physicists and tasks, but may be useless for Nancy. Although it seems counterintuitive, focusing on specifics can produce better broad results because you can relate the results to a theme.

Determine your objectives and choose a set of heuristics that test how well your tool meets those objectives.

Published heuristics are a good starting point. Interface design heuristics (e.g., [8],[10]) can be valuable, but do not focus on visualization. Therefore, we also recommend using visualization guidelines (e.g., [3],[9],[11]).

Select a set of experts.

A typical process includes approximately five evaluators. Choose usability experts with:

- strong communication skills
- experience conducting usability inspections, and
- experience with data display (not just usability)

Choosing people you know may enable you to better assess their experience; however, they must be independent of the development team and be willing to give honest opinions. Including end users as well as usability experts can provide a different perspective. Establish a good rapport with the evaluators so they will be comfortable talking with you.

Step 2: Conduct the evaluations

Have experts work independently.

Working independently allows experts to form and express their own opinions, and allows you to observe the experts and talk to them while they work. Asking evaluators to think aloud or probing them with questions while they use the software may help generate more detailed feedback.

Do not place too much emphasis on the heuristics.

Although the heuristics provide guidance, they should not dominate the process. Since the review is informal, focus on qualitative issues. Ask questions and encourage evaluators to discuss their opinions in detail.

Remain neutral and do not defend your visualization tool.

Asking unbiased questions will encourage honest responses. If evaluators are critical, encourage them to discuss the reasons for their opinion and how they might change the tool. If maintaining a neutral position is difficult, have someone else conduct the interviews.

Take copious notes.

If notetaking is difficult and interferes with the conversation, have another person take notes. Videotaping the session and taking notes from the video may also help, but should not replace notetaking during the event.

Step 3: Analyze the results

Review your notes soon after the evaluations, while memory of the event is still clear. Compare responses of the evaluators, identifying common themes and areas of disagreement.