Using Visual Design Experts in Critique-based Evaluation of 2D Vector Visualization Methods

Daniel Acevedo, Cullen D. Jackson, Fritz Drury, and David H. Laidlaw

Abstract— **We describe an experiment in which art and illustration experts evaluated six 2D vector visualization methods. We found that these expert critiques mirrored previously recorded experimental results; these findings support that using artists, visual designers and illustrators to critique scientific visualizations can be faster and more productive than quantitative user studies. Our participants successfully evaluated how well the given methods would let users complete a given set of tasks. Our results show a statistically significant correlation with a previous objective study: designers' subjective predictions of user performance by these methods match users measured performance. The experts improved the evaluation by providing insights into the reasons for the effectiveness of each visualization method and suggesting specific improvements.**

Index Terms— **Scientific Visualization, Perception, Visual Design, User Study, Two-dimensional Vector Fields, Critical Point, Advection, Design Goals, Evaluation Tasks, Visual Dimensions.**

I. Introduction

THE human visual system is a highly optimized pattern detection and recognition system. Visualization methods leverage this ability to allow efficient data exploration, discovery, and analysis. One goal of exploratory scientific visualization is to display measurements of physical quantities so that the underlying physical phenomena can be interpreted accurately, quickly, and without bias, prompting visual thinking and knowledge construction [13].

This paper addresses the question of how best to evaluate the effectiveness of visualizations methods, and also asks whether a good evaluation method suggests improvements in the visualization methods. Visualization is used in data-intensive domains; data mining, meteorology, geography, transportation sciences, environmental studies, uncertainty analysis, and evolutionary biology are a few examples. In these fields, a common problem for visualization experts is: given a large set of multivalued data and hypotheses scientists would like to address, what visualizations best represent the data and how do we measure their effectiveness?

We hypothesize that using visual design experts to perform critique-based evaluations can let us quantify the expected performance of visualization methods as well as elicit fixes for visual design problems that are often difficult for a domain or visualization expert to articulate. Evaluation of scientific visualization methods is typically either anecdotal, via feedback from or observation of scientific users, or quantitative, via measurement of the performance of relatively naïve users on simple abstract tasks. In this study we add visual design experts to the pool of evaluators (see Table I).

	Pros	Cons
Non-experts	• Easy access • Unbiased opinion • Help on minor or overlooked issues	• Possible subconscious influence of external factors • Little help on fixes/improvements
Visualization Experts	• Easy access • Evaluation of implementation issues • Knowledge of alternative methods	• Too close to development: possible bias towards easily implemented or more familiar techniques
Domain Experts	• Specific knowledge of tasks and goals • Extrapolate to tasks not being tested	• Infrequent access • Expensive/busy • Biased: too much experience with established techniques • No help on improvements
Visual Design Experts	• Capable of translating scientific goals into design goals • Concentrate on overall data readability • Provide guidance on improvements	• Infrequent access • Expensive/busy • No training in scientific goals • No training in visualization techniques

TABLE I

PROS AND CONS OF PARTICIPANT TYPES IN QUALITATIVE SCIENTIFIC VISUALIZATION EVALUATIONS.

Human-computer interaction (HCI) literature has established that usability experts are more efficient participants in heuristic evaluation studies than novices [14] (heuristic studies are performed early in a project's development to find both major and minor problems in a design). Minor issues, however, are missed by experts and can be uncovered only by end-user evaluation in real situations.

Here we propose expertise in visual design as the basis of a visualization evaluation methodology that assesses the effectiveness of scientific visualizations, providing reasons for that effectiveness and suggesting improvements. Our participants, visual designers and illustrators, are experts in evaluating visuals for targeted communication goals; while their results are often appealing and aesthetic, they first must satisfy the communication goals, which in this case means presenting scientific data for effective exploration.

Acevedo is with the R&D Department at Acciona Infraestructuras, S.A., Madrid, Spain. E-mail: acevedo.daniel@gmail.com

Laidlaw is with the Computer Science Department, Brown University, Providence, RI 02912, USA. E-mail: dhl@cs.brown.edu

Jackson is with the Cognitive Systems Group at Aptima Inc., Woburn, MA 01801, USA. E-mail: cjackson@aptima.com

Drury is with the Department of Illustration, Rhode Island School of Design, Providence, RI 02903, USA. E-mail: fdrury@risd.edu

The purpose of this study is to evaluate whether these experts can effectively evaluate scientific visualization methods as compared to a quantitative study performed earlier [11]. I also hope to learn more about how visual designers approach these evaluations. Understanding this process should help us build better evaluation methods, particularly ones that will both judge visualizations on their scientific merits and provide insights into improving their design.

In the present experiment, artists and visual designers graded the vector visualization methods from a previous study [11] on the basis of their subjective estimates of user performance and also verbally critiqued each method's effectiveness. Our hypothesis was that designers would rank the methods similarly to the objective task-performance measures in [11]. We also hoped that the critiques would help us understand why methods work well by identifying which visual attributes within each method worked best for the given tasks. Our results are consistent with our hypothesis.

II. State of the Art

Many researchers are trying to answer the question of what the best ways to craft visualizations are and how to evaluate their success. We summarize here several approaches found in the literature.

One possible strategy is to use perceptually based rules that let the user perceive data features efficiently without confusion from conflicting visual cues. Among the perceptually based cues proposed are preattentive visual cues such as color and orientation [7] and Gestalt principles of grouping according to proximity, similarity, or closure; even motion has been used to disambiguate visual elements [23] or to animate time-varying data.

Another possible strategy is to use design-based rules that incorporate aesthetics and task-driven solutions [5], [17]. Some of the design-based rules proposed use inspirations from art to drive visualization solutions, for instance, using an Impressionist style, or using brush strokes from a particular work of art [7], [10], [22].

On the more practical topic of visualization synthesis, many current systems provide environments for generating visualizations: AVS [21], Vis5D [8], VTK [16], IRIS Explorer [6] and IBM's Data Explorer [12] with PRAVDAcolor. None of these systems, however, provide any visual design guidance or suggestions for evaluating the resulting visualizations. They all provide plenty of knobs to modify and adapt your visualization methods, but rarely do they give any advice about how to turn the knobs to create a more effective visualization. A notable exception is PRAVDA, which was developed with perceptual principles in mind to provide users the most effective color maps depending upon the type of data being visualized [15].

In order to evaluate and compare visualizations, there must be a way to classify their possible goals and analyze how different methods fulfill those goals. Much research has been done in this area [2], [3], [18], mostly in the information visualization arena. In addition, in the expert systems field, Xu et al. [24] have tried to capture the designer's intent in order to customize the visualization and design sketches and reach a valid solution efficiently.

Once the goals of a scientific visualization are defined, two evaluation methods are currently used: anecdotal and quantitative studies. Anecdotal studies are usually conducted by asking scientists to evaluate the visualizations on their scientific merits; the feedback this method elicits is usually very specific to the visualization and scientific problem at hand. Quantitative studies, on the other hand, explore the performance of the visualization methods on generic tasks, and their results may well be generalizable to other methods within the same scientific domain. Neither evaluation method addresses visual design issues such as, for example, possible confounding effects due to simultaneous contrast, or the visual organization of the displays that might unexpectedly create highlighted areas when even attention to the whole display is intended.

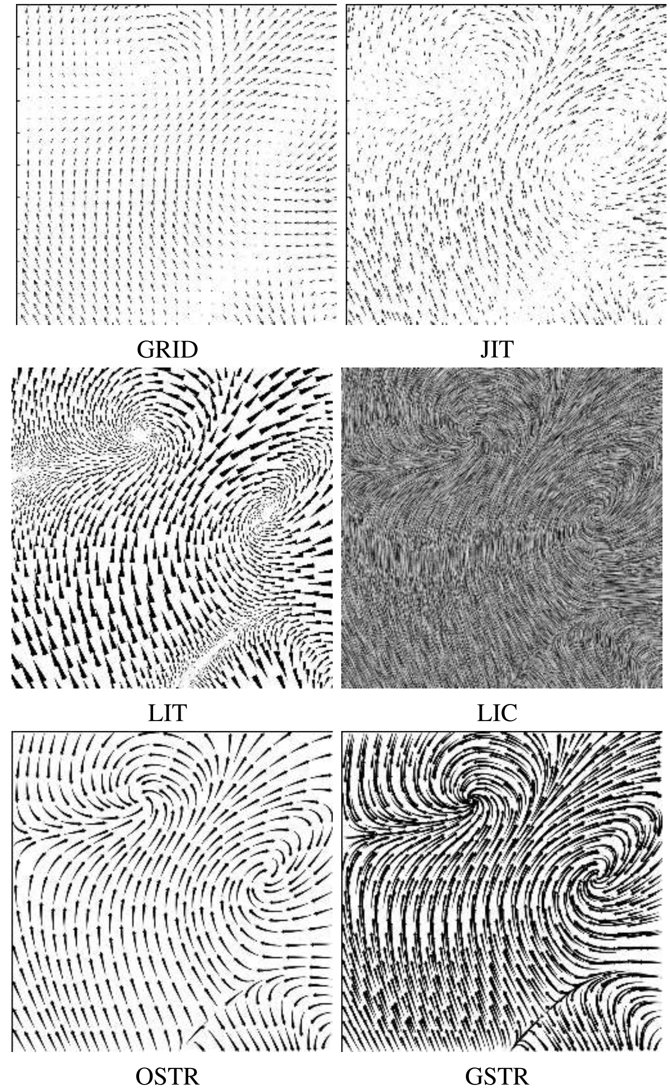


Fig. 1. The same vector field as visualized by the six visualization methods critiqued by the designers.

As a possible solution to these issues, Tory and Möller [19] recommend using both domain experts and quantitative studies. We extend the state of the art for evaluation methodology by exploring the use of visual design experts as our evaluators, potentially reducing the reliance exclusively on quantitative task-specific studies. Expert visual designers know from experience the limitations of each visual dimension, and we hope to collect and use that knowledge to guide the creation of new visualization methods. In order to show that designers can contribute in this

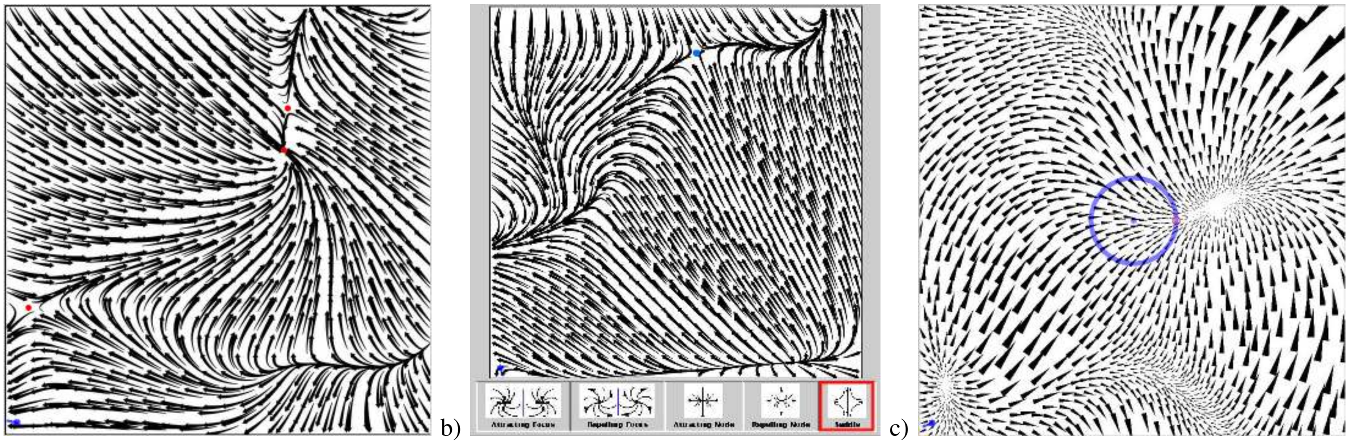
a) b) c)


Fig. 2. Sample stimuli for the three experimental tasks. The solutions for each, marked in red, were provided to participants during the subjective critiques. Their goal was to judge how accurately and quickly a real user of the visualizations would perform these tasks for each method. (a) Counting task with three critical points visible, (b) Type ID task with a saddle-type point (marked in blue), and (c) Advection task in which a small red circle indicates the location to which a particle in the center of the large blue circle would advect.

way, we designed the following experiment.

III. Evaluation of Vector Field Visualization Methods

A. Methodology

In order to evaluate the efficacy of our designer critiques, we modeled our study on a previous quantitative user study [11] comparing six 2D vector field visualization methods on three different tasks using expert and novice scientists. Having designers evaluate the same six visualization methods, using the same tasks as in the previous study, let us validate our designers' ability to evaluate scientific visualizations effectively.

In [11], users were asked to evaluate the merits of the six visualization methods shown in Fig. 1:

- GRID: icons on a regular grid.
- JIT: icons on a jittered grid [4].
- LIT: icons using one layer of a visualization method that borrows concepts from oil painting [10].
- LIC: line-integral convolution [1].
- OSTR: image-guided streamlines (integral curves) [20].
- GSTR: streamlines seeded on a regular grid [20].

With these methods, users were asked to perform three tasks designed to mimic generic tasks fluid-flow experts would use to investigate a vector field (Figure 2):

- *Counting Task:* Choosing the number and location of all critical points (CP) in an image.
- *Type ID Task:* Identifying the type of a CP at a specified location.
- *Advection Task:* Predicting where a particle starting at a specified point will advect.

Seventeen users were run through the 90-minute computer-controlled experiment [11]: five were fluid-flow experts and 12 were first- or second-year applied math graduate students with little previous experience in computational fluid dynamics. Details of the results are given in [11].

In the present study, visual designers were asked to judge the six visualization methods on their ability to convey the information necessary for a user to complete the three tasks accurately and quickly. Figure 3 shows one of our visual designers critiquing the

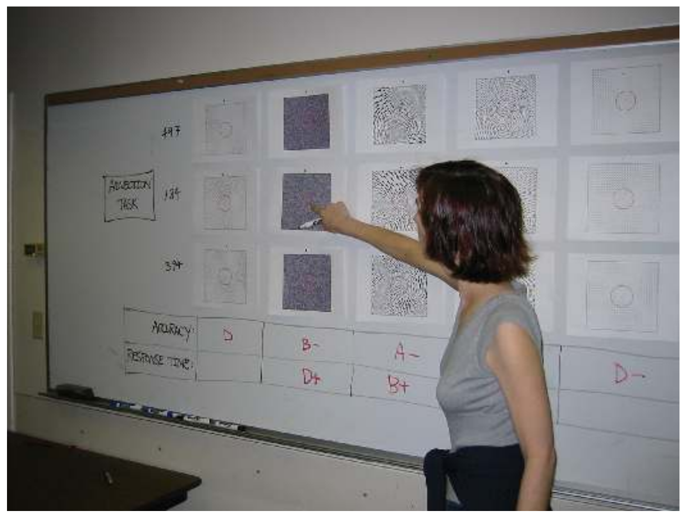


Fig. 3. During the study, designers rated the different methods subjectively, based on accuracy and time to perform the task. They could also appraise how the visual dimensions used in each method would affect their performance.

six methods. The experiment took an average of 60 minutes. Six experts, who were compensated for their participation, judged all six methods for all tasks (within-subjects design). As a training exercise, all designers took the objective computer-based study first. Participants could ask the experimenter for any necessary clarification during the experiment.

Designers evaluated the methods using printed images from three different datasets simultaneously. This allowed them to critique a visualization method on its own expressive capabilities and not on its specific instantiation for a dataset. (The training on the computer helped here.) The methods for each task were rated separately using letter grades (GPA-style: F, F+, D-, D, D+, C-, C, C+, B-, B, B+, A-, A, A+) according to two measures:

- How well the method would let a user perform the given task accurately.
- How well the method would let a user perform the given task quickly.

Finally, after the critique was completed, designers were asked to create a new visualization of a given data set that would enable

	Counting Task			Type ID Task	
	Accuracy (count)	Accuracy (distance)	Response Time	Accuracy	Response Time
R Square	0.941	0.956	0.676	0.722	0.679
F	63.921	87.579	8.341	10.401	8.457
p	0.001	0.001	0.045	0.032	0.044

	Advection Task (full)		Advection Task (no LIC)	
	Accuracy	Response Time	Accuracy	Response Time
R Square	0.615	0.249	0.852	0.389
F	6.381	1.324	17.211	1.911
p	0.065	0.314	0.025	0.261

TABLE II

LINEAR REGRESSION RESULTS BETWEEN DESIGNER GRADES AND NUMERICAL RESULTS FROM [11].

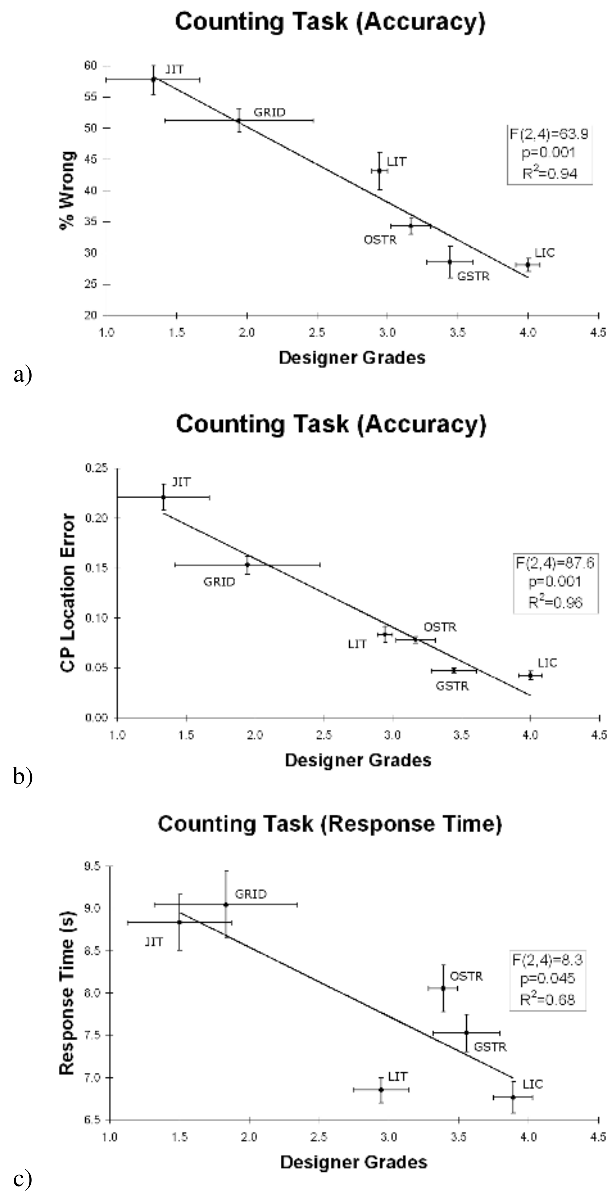


Fig. 4. Regression analyses for the critical-point-counting task, with plots for counting accuracy (a), location accuracy (b), and response time (c). Standard error bars are plotted. Regression test results are also shown. All regressions are significant at $\alpha = 0.05$.

users to perform all three tasks quickly and accurately.

B. Numerical Results and Discussion

We posed two hypotheses at the onset of this study: first, that designer ratings would be similar to the quantitative performance measures for each task in the previous study [11], and second, that the designer critiques would provide additional insight into the merits of each method and how to improve them. Table II summarizes the linear regression results between the designer grades from the current study and the numerical results from [11]. In order to perform the correlation analysis, the letter grades correspond to the following numerical values: F (0), F+ (0.33), D- (0.66), D (1), D+ (1.33), C- (1.66), C (2), C+ (2.33), B- (2.66), B (3), B+ (3.33), A- (3.66), A (4),and A+ (4.33). Our results show a good correlation across methods and tasks between the subjective evaluations of the designers and the performance measures of the previous study. Although some discrepancies are present, we consider that these results validate our hypotheses.

We look first at the critical-point-counting task. Apart from their response time rating, designers could give only one rating for the other two accuracy variables measured: accuracy of finding the correct number of critical points and accuracy of placing the critical-point markers precisely on their locations. Participants in the objective study performed these last two tasks simultaneously.

Figure 4(a) shows the regression analysis for mean percentage correct in counting the critical points, and also the mean designer grades. It is clear that the designers' pattern of performance matches the quantitatively collected performance measure for this task very well ($R^2 = 0.941, F = 63.9, p = 0.001$). Figure 4(b) shows the regression analysis for mean critical-point-location error and the mean designer grades. Again, the designers' pattern of performance matches the quantitatively collected performance measure ($R^2 = 0.956, F = 87.6, p = 0.001$). Last, Fig. 4(c) shows the regression analysis for the mean time to complete the critical-point-location task and the mean designer grades. Once again, the designers' pattern of performance matches the quantitatively collected performance measure ($R^2 = 0.676, F = 8.3, p = 0.045$).

As can be seen from the graphs in Fig. 5, the designer grades closely matched the pattern of performance in the original quantitative user study for the critical-point-type identification task, both accuracy ($R^2 = 0.722, F = 10.4, p = 0.032$) and response time ($R^2 = 0.679, F = 8.5, p = 0.044$).

However, for the advection task, the designer grades did not quite match the previous experiment's pattern for accuracy ($R^2 = 0.615, F = 6.4, p = 0.065$). Also, no regression model fit the designer ratings to the quantitative response time measure ($R^2 = 0.249, F = 1.3, p = 0.314$).

This last discrepancy can be explained by looking at one visualization method in particular: line integral convolution (LIC). As seen in Fig. 1, LIC shows no information about the vector field direction, and this is detrimental in performing the advection task. In order to compensate for this known problem, a direction icon was placed at the lower-left corner of the image to let users extrapolate the field's direction across the entire image. The time needed for this extrapolation contributed to the large increase in completion time for this method in the previous user study. Most designers viewing this method for the advection task suggested

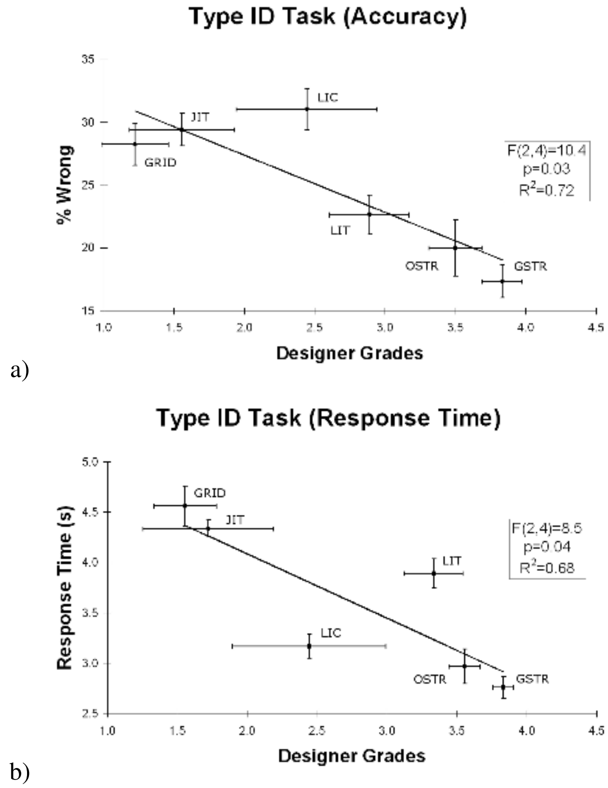


Fig. 5. Regression analyses for the critical-point-type identification task, with plots for type accuracy (a), and response time (b). Standard error bars are plotted. Regression test results are also shown. Both regressions are significant at $\alpha = 0.05$.

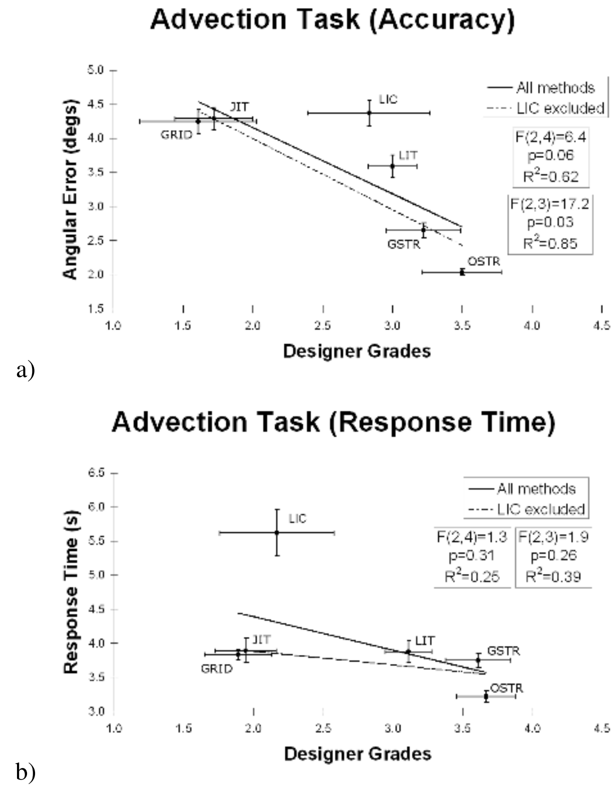


Fig. 6. Regression analyses for the advection task, with plots for accuracy (a), and response time (b). Standard error bars are plotted. LIC results were removed for the "LIC excluded" regression. Test results are also shown, and only the "LIC excluded" regression for accuracy is significant at $\alpha = 0.05$.

adding direction icons sparsely throughout the image; having seen this easy fix, they tended to grade the completion time for this method leniently, resulting in the poor correlation between the two sets of data for this task. Removing this polemic method from the regression analysis yields significant results for accuracy ($R^2 = 0.852, F = 17.2, p = 0.025$), but does not improve the response time regression results ($R^2 = 0.389, F = 1.9, p = 0.261$). Figure 6 shows the linear regression plots for this task with and without LIC.

This mismatch between designers' opinions and the original performance measures illustrates the fact that we cannot expect this methodology to yield perfect correlations with objective studies. The nature of subjective evaluations is such that we rely on the expertise of our participants to be as consistent as possible, but not perfect. Overall, we obtained good correlations across methods and tasks between both experiments.

C. Design Issues and the Development of a New Method

Apart from those numerically significant results that validate their evaluations, participants provided additional design insights into how to improve the visualization methods to potentially yield quick and accurate information on the vector fields in the three given tasks.

JIT was rated as the "worst" method because its elements were "too small." OSTR, on the other hand, was possibly the "best" method, although sometimes "very sharp turns don't give a sense of movement as well as others." GRID, like JIT, has elements that are "too small to be effective," and "the regularity of the grid induces a false sense of structure that is difficult to ignore." LIC is "OK" but is perceptually "too even" with "not enough contrast," and its elements "don't provide a good sense of flow direction," which is key for some tasks. "Its good sense of tactility connects the user with the concept of flow," but this aesthetic appreciation did not affect the participants' scores, which concentrated on task performance. LIT and GSTR were both "good representations for doing the advection task," but LIT had elements that were "a little small" and GSTR was a bit "scary" to look at, since the visual elements seemed to "pass over each other." Comments about the size were also common, indicating this dimension as the first candidate for modification in order to increase the effectiveness of most methods.

In addition to these critiques, we asked the designers to design a new visualization method for a sample vector field data set that would address all three tasks. Figure 7 shows one of these designer-created visualizations; this image was created by hand using tempera paint, charcoal, and pencil. As you can see, this designer added direction icons, used streamlines to suggest field structure and thus aid in identifying particle advection, placed icons around the critical points for easy identification, and put dots on the critical points to make them easy to locate. It is interesting to see how some of the comments above are exploited in this particular solution. The tactility of LIC, for example, is retained, while its directional ambiguity is solved through small additions. We found that participants designed to the tasks presented and missed the implicit task of understanding the overall vector field structure and features.

Fig. 7. After the experiment, designers were asked to create a new visualization design that would outperform the six methods presented. This image shows one of the results. Black field lines help in the advection task, and white marks indicate direction of the vector field. Critical points are clearly marked by large white dots and critical point type is indicated by the surrounding arrows.

IV. General Discussion

Our results validate our initial hypotheses, but they leave some open questions. In this study, designers were asked to replicate tasks that were quantitatively posed, because we were comparing with earlier quantitative results. As can be seen from Fig. 7, a task-oriented design query yields, naturally, a highly explanatory visualization method in which answers to all three tasks are explicitly depicted. We surmise that assigning a task that requires a more holistic understanding of the datasets will bring out the best in the designers, and that the results will be more effective the greater the designers' expertise.

While the initial study [11] found no differences between experts and non-experts in performing the quantitative tasks, our subjective tasks may elicit some differences among participants with different levels of expertise, as suggested in the HCI literature [14]. In particular, we believe that the participants' visual design expertise is key to providing the types of comments they did during our experiment. Our characterization of vector field visualization methods acknowledges that the input we get from the designers is directly targeted at the needs of scientists, and does not concern artistic qualities, visual appeal, or aesthetics. However, different experience levels on the designers' part may lead to very different critiques of the same visual displays. We believe our next step should be to use experienced educators who teach design concepts. They are used to concentrating on the problem at hand, abstracting from aesthetic considerations when they must focus on the final goal of the design; while their results are often aesthetically appealing, they first must satisfy the given communication goals and teach their students how to do so.

Finally, since the ratings obtained from designers are largely qualitative and do not provide the numeric values necessary to design a visualization method, it seems clear that combining objective and subjective experiments using designers will lead to better, more directly usable results, confirming the hypothesis from Tory and Möller [19]. This combination of quantitative and qualitative studies would yield both numeric performance estimates and guidance on what aspects of different visualization methods help or impede performance on certain tasks.

The experiment presented here lays out a way of using expert visual designers as evaluators of mappings between the data and the visual dimensions that form our visualization methods. This methodology is limited to the specific tasks presented and is difficult to extrapolate to more exploratory visualization goals. Although this is a common issue in any type of user study, we believe the use of visual designers can help us bridge that gap. In particular, by concentrating on improving the overall readability of the data and suggesting improvements, visual designers are providing us general guidelines to ultimately use our visualization methods effectively. It is clear that much more research is needed and that methodological questions, along with the mathematical basis of the eventual model for this design space, must be explored and studied further.

The cost of training designers in the scientific goals of the visualization methods is more than recovered by their ultimate contributions. In this experiment we had participants perform first the computer-based study from [11] to train them on the tasks and methods they were going to critique later. This training step increased the overall time they took in performing the full experiment. However, we believe a much shorter and informal explanation of the tasks and scientific goals would be sufficient to engage these experts and have them effectively critique visualization methods. As with any other interdisciplinary endeavor, care must be taken about the language used to explain those goals to ensure a correct understanding, but the experience visual designers have in effectively communicating information visually allows them to quickly grasp the more general goals and concentrate on clarifying the details of the particular scientific problem.

V. Notes on Visual Design Experts as Evaluators

From our experience performing this study and other collaborations with visual designers, we can provide some specific guidelines to conduct this type of evaluation. Although the results we have presented are statistically sound, these guidelines are gathered from our own experiences and should not be taken as proven rules. We provide them to help readers develop their own studies and learn, on their own, the difficulties and nuances of including these subjective evaluations as part of their testing pipelines.

- *Experience is good:* A big part of learning "good" visual design is creating many solutions to a problem and critiquing them. The skill set of a experienced designer is much more broad, in general, than that of a design student. This helps us because they are able to foresee problems with visualization methods for datasets that are not directly shown to them. As we mentioned before, we further believe that educators would be better at critiquing scientific displays than just highly experienced professionals, if only for their ability to explain what they do to others.
- *Illustrators learn new problems quickly:* During our collaborations, we have worked together with illustrators, industrial designers, painters, sculptors, and 3D artists. We have found that illustrators are more easily engaged in critiquing scientific displays than any other type of visual designer. It could be that their attention is focused on the individual pieces of information that could be extracted from

a display and optimizing the visual elements to combine those pieces effectively. Although this might sound like a general definition for visual design, we believe illustrators have a much more broad spectrum for their subject matter. In other words, their minds are much more open to any kind of information communication task, instead of focusing on specific types of it.

- *How many experts are needed?* We have found that just bringing one expert visual designer to evaluate our methods, improves tremendously our efficiency and effectiveness in creating a useful tool for scientists to use. For the current study we were able to recruit six designers and that allowed us to perform statistical analysis on their results. However, the goals of this particular experiment were very specific. As with any collaboration, the more familiar each side of the team is with the other field, the better the results are. It takes time to develop this familiarity and confidence and, if one expert visual designer is engaged in multiple evaluations of scientific visualization methods, his or her effectiveness will increase over time. The differences among experts illustrated by the horizontal standard error bars in our graphs show their level of discrepancy for the different tasks. This experiment was the first time most of the designers critiqued this type of displays. We believe further experiences like this would narrow those error bars without having participants fall into the biases posed in Table I.
- *Do not abandon quantitative evaluations:* As mentioned in the last point, we put both types of evaluation in series with each other. Most of the time with a loop around the subjective evaluation with designers first. This improves the chances of success of the quantitative study, not in terms of just validating its hypothesis, but in terms of engaging the end-users to utilize the tool and being effective at their job with it.
- *Non-experts can detect minor issues easily:* This is clearly acknowledged in the HCI community [14] and we have not performed the present study using lay people as participants. Our sense is that they would recognize obvious flaws and, sometimes, pick up minor issues that experts would overlook but, in general, they would probably not provide suggestions for fixing design issues.

Another upcoming paper from our group develops the details of these collaborations in more detail [9].

VI. Conclusion

The number of options available to solve a visualization problem is far too great for a full analysis of the design space, and expert visual designers can help us explore this space more efficiently.

The main result of this study is that designers can evaluate scientific visualizations effectively: they provide extra information, such as reasons for the good or bad performance of visualization methods, that participants knowledgeable in the specific scientific field cannot give us. We successfully correlated their subjective critiques with previous studies and we obtained new insights into how different methods work, which will need to be evaluated in subsequent studies.

Acknowledgments

We would like to thank all the participants in our experiment, Katrina Avery for her thorough editing, and all the members of the Visualization Research Lab at Brown University. This work was partially supported by the National Science Foundation grants CCR-0086065, CCR-0093238, CCF-0324306, and CNS-0427374.

References

[1] B. Cabral and L. C. Leedom. Imaging vector fields using line integral convolution. In J. T. Kajiya, editor, *Computer Graphics (SIGGRAPH '93 Proceedings)*, volume 27, pages 263–272, Aug. 1993.
[2] S. M. Casner. A task-analytic approach to the automated design of graphic presentations. *ACM Transactions on Graphics*, 10(2):111–151, April 1991.
[3] M. Dastani. The role of visual perception in data visualization. *Journal of Visual Languages and Computing*, 13(6):601–622, December 2002.
[4] M. A. Z. Dippé and E. H. Wold. Antialiasing through stochastic sampling. In B. A. Barsky, editor, *Computer Graphics (SIGGRAPH '85 Proceedings)*, volume 19, pages 69–78, July 1985.
[5] S. G. Eick. Engineering perceptually effective visualizations for abstract data. In *Scientific Visualization Overviews, Methodologies and Techniques*, pages 191–210. IEEE Computer Science Press, 1995.
[6] D. Foulser. IRIS explorer: A framework for investigation. *Computer Graphics*, 29(2):13–16, May 1995.
[7] C. G. Healey, J. T. Enns, L. Tateosian, and M. Rempel. Perceptually-based brush strokes for nonphotorealistic visualization. *Transactions on Graphics*, 23(1), January 2004.
[8] W. Hibbard and D. Santek. The VIS-5D system for easy interactive visualization. In *Proceedings IEEE Visualization 1990*, pages 129–134, 1990.
[9] D. Keefe, D. Acevedo, J. Miles, F. Drury, S. Swartz, and D. H. Laidlaw. Scientific sketching for collaborative VR visualization design. *IEEE Transactions on Visualization and Computer Graphics*, 2008. In Press.
[10] M. Kirby, H. Marmanis, and D. H. Laidlaw. Visualizing multivalued data from 2D incompressible flows using concepts from painting. In *Proceedings of IEEE Visualization 1999*, pages 333–340, 1999.
[11] D. H. Laidlaw, M. Kirby, C. Jackson, J. S. Davidson, T. Miller, M. DaSilva, W. Warren, and M. Tarr. Comparing 2D vector field visualization methods: A user study. *In IEEE Transactions on Visualization and Computer Graphics*, 11(1):59–70, January-February 2005.
[12] B. Lucas, G. D. Abram, N. S. Collins, D. A. Epstein, D. L. Gresh, and K. P. McAuliffe. An architecture for a scientific visualization system. In *Proceedings IEEE Visualization 1992*, 1992.
[13] A. MacEachren and M.-J. Kraak. Exploratory cartographic visualization: Advancing the agenda. *Computers and Geosciences*, 23(4):335–343, 1997.
[14] J. Nielsen. Finding usability problems through heuristic evaluation. In *Proceedings of CHI*, pages 373–380, 1992.
[15] B. E. Rogowitz and L. A. Treinish. How not to lie with visualization. *Computers in Physics*, pages 268–273, 1996.
[16] W. J. Schroeder, K. Martin, and W. E. Lorensen. The design and implementation of an object-oriented toolkit for 3D graphics and visualization. In *Proceedings IEEE Visualization 1996*, pages 93–100, 1996.
[17] H. Senay and E. Ignatius. A knowledge-based system for visualization design. *IEEE Computer Graphics and Applications*, 14(6):36–47, November 1994.
[18] R. R. Springmeyer, M. R. Blattner, and N. L. Max. A characterization of the scientific data analysis process. In *Proceedings of the Second IEEE Visualization Conference*, pages 235–242, October 1992.
[19] M. Tory and T. Moller. Evaluating visualizations: Do expert reviews work? *IEEE Computer Graphics and Applications*, 25(5):8–11, Sep/Oct 2005.
[20] G. Turk and D. Banks. Image-guided streamline placement. In *Proceedings of SIGGRAPH 96*, pages 453–460. ACM SIGGRAPH, 1996.
[21] C. Upson, T. F. Jr., D. Kamins, D. H. Laidlaw, D. Schlegel, J. Vroom, R. Gurwitz, and A. van Dam. The application visualization system: A computational environment for scientific visualization. *IEEE Computer Graphics and Applications*, 9(4):30–42, July 1989.
[22] E. Vote, D. Acevedo, C. Jackson, J. Sobel, and D. H. Laidlaw. Design-by-example: A schema for designing visualizations using examples from art. In *SIGGRAPH 2003 Sketches and Applications*. ACM SIGGRAPH, 2003.

[23] C. Ware and R. Bobrow. Motion coding for pattern detection. In *Proceedings of the 3rd Symposium on Applied Perception in Graphics and Visualization*, pages 107–110, Boston, MA, 2006.

[24] X. W. Xu and R. Galloway. Using behavioral modeling technology to capture designer's intent. *Computers in Human Behavior*, 21(2):395–405, March 2005.

Daniel Acevedo received his PhD degree in computer science from Brown University, and a BS degree in civil engineering from University of A Coruña in Spain. He is currently leading the Visualization and 3D Modeling Group of the R&D Department at Acciona Infraestructuras, S.A. in Madrid, Spain. His research centers on scientific visualization, human perception, human-computer interaction, and on the creation of training and control tools for large civil engineering operations.

Cullen Jackson received the PhD degree in experimental psychology from Brown University, and a BS degree in computer science from Trinity University. He was a postdoctoral research associate in the Computer Science Department at Brown University, and he is currently a Senior Cognitive Scientist and Team Lead at Aptima, Inc., in Woburn, MA. His current work involves assessing the impact of technology on human performance for training and operational applications.

David H. Laidlaw received the PhD degree in computer science from the California Institute of Technology, where he also did postdoctoral work in the Division of Biology. He is an associate professor in the Computer Science Department at Brown University. His research centers on applications of visualization, modeling, computer graphics, and computer science to other scientific disciplines. He is a senior member of the IEEE and the IEEE Computer Society.

Fritz Drury received a Master of Fine Arts degree from Yale University and a Bachelor of Arts with distinction in Art History from Stanford University. He is a professor in the Department of Illustration at Rhode Island School of Design. He is currently at work on a text: Drawing: Structure and Vision, to be published later this month by Prentice Hall. He is a painter based in New York City, currently at work on a group of large-scale oil paintings defining contemporary mythologies. His research in includes study of visual psychology in application to fine and applied art.