Abstract of "A Framework for the Perceptual Optimization of Multivalued Multilayered 2D Scientific Visualization Methods" by Daniel Acevedo Feliz, Ph.D., Brown University, May 2008.

In this dissertation we address an important problem in the visualization of multivalued scientific datasets: To quantify the utility of different visualization methods. This quantification could become the basis for a theory of visualization that would explain and predict how different methods represent data effectively. Having such a utility function defined over a space of visualization methods would facilitate the search for an effective representation of scientific data and, given a problem, optimally use visual resources to solve it. Our hypothesis is that by measuring the perceptual capabilities of some of those methods for simple single-valued cases, and combining that with subjective evaluations of complex multivalued displays using expert visual designers, we can generate a predictive model of utility for a space of visualization methods. Defining this model involves understanding the capabilities of those methods to represent data individually, and quantifying the effectiveness changes when they interact to represent multiple data variables simultaneously. While experiments inspired by psychophysical studies can inform about the expressive capabilities of individual methods, the complexity of the combined displays create an exponentially growing amount of variables to be controlled during the studies. Using critiques from expert visual designers to evaluate such combinations can reduce the experimental difficulties and help create the utility model.

Our contributions include new experimental and computational techniques to evaluate how different visualization methods perform when displaying multivalued scientific datasets in 2D. The work spans several experiments aimed at quantifying the perceptual capabilities of some icon-based visualization methods, and we report on some significant results that help describe the structure of our space of visualization methods and the process necessary to explore it.

Our hypothesis aims at stablishing a basis for a theory of visualization. Our results contribute to that goal by providing several models for the effective use of some 2D visualization methods. We have also produced a rich knowledge base for the design, execution, and analysis of evaluation studies that use expert visual designers as the main participants. We hope the visualization community will benefit from this body of work in its continuing quest for its theoretical foundations.

A Framework for the Perceptual Optimization of Multivalued Multilayered 2D Scientific Visualization Methods

by

Daniel Acevedo Feliz

B. S. Civil Engineering, University of A Coruña, Spain, 1997

M. Sc. Computer Science, Brown University, Providence RI, 2001

A dissertation submitted in partial fulfillment of the
requirements for the Degree of Doctor of Philosophy
in the Department of Computer Science at Brown University

Providence, Rhode Island

May 2008

This dissertation by Daniel Acevedo Feliz is accepted in its present form by the Department of Computer Science as satisfying the dissertation requirement for the degree of Doctor of Philosophy.

Date ______________ ____
David H. Laidlaw, Director

Recommended to the Graduate Council

Date ______________ ____
John F. Hughes, Reader

Date ______________ ____
Leslie Welch (Department of Psychology), Reader

Approved by the Graduate Council

Date ______________ ____
Sheila Bonde
Dean of the Graduate School

Curriculum Vitae

Daniel Acevedo Feliz was born on October 15, 1974 in A Coruña, Spain. He attended University of A Coruña, where he received a bachelor's degree in civil engineering in September 1997. After obtaining a fellowship from Pedro Barrie de la Maza Foundation, he arrives in Providence, RI in August 1998 to attend Brown University as a graduate student in the Department of Computer Science, where he received his master's degree in 2001 for his work in the ARCHAVE project: a virtual reality application for archaeological research. Continuing his studies towards the doctorate degree, Daniel works on many projects in the areas of visualization, virtual reality, user interfaces, human-computer interaction, perception, and visual design, publishing and presenting his work in several journals and conferences. His publications are cited below.

Journal Papers

- Using Visual Design Experts in Critique-based Evaluation of 2D Vector Visualization Methods. Daniel Acevedo, Cullen Jackson, Fritz Drury and David Laidlaw. *IEEE Transactions on Visualization and Computer Graphics, 2007, in review.*

- Subjective Quantification of Perceptual Interactions among Some 2D Scientific Visualization Methods. Daniel Acevedo and David Laidlaw. *IEEE Transactions on Visualization and Computer Graphics, 12(5), September-October 2006. Proceedings of IEEE Visualization/Information Visualization.*

- Discovering Petra: Archaeological Analysis in VR. Eileen Louise Vote, Daniel Acevedo Feliz, David H. Laidlaw, and Martha S. Joukowsky. *IEEE Computer Graphics and Applications. 22(5), 38-50, September/October 2002.*

Peer-Reviewed Conference Papers

- Pop Through Buttons for Virtual Environment Navigation and Interaction. Robert Zeleznik, Joseph LaViola, Daniel Acevedo, and Daniel Keefe. *Proceedings of IEEE Virtual Reality 2002, March 2002.*

- Archaeological Data Visualization in VR: Analysis of Lamp Finds at the Great Temple of Petra, a Case Study. Daniel Acevedo, Eileen Vote, David H. Laidlaw and Martha S. Joukowsky. *Proceedings of IEEE Visualization 2001. San Diego, California. October 2001.* BEST CASE STUDY AWARD.

- CavePainting: A Fully Immersive 3D Artistic Medium and Interactive Experience. Daniel Keefe, Daniel Acevedo Feliz, Tomer Moscovich, David Laidlaw and Joseph LaViola. *Proceedings of ACM SIGGRAPH's I3D 2001, Symposium on Interactive 3D Graphics. North Carolina, March 2001.*

- Hands-Free Multi-Scale Navigation in Virtual Environments. Joseph LaViola, Daniel Acevedo Feliz, Daniel Keefe and Robert Zeleznick. *Proceedings of ACM SIGGRAPH's I3D 2001, Symposium on Interactive 3D Graphics. North Carolina, March 2001.*

- Virtual Reality and Scientific Visualization in Archaeological Research. Eileen Vote, Daniel Acevedo, Martha S. Joukowsky and David Laidlaw. *VAST 2000: Virtual Archaeology between Scientific Research and Territorial Marketing. Arezzo, Italy. November 2000.*

- ARCHAVE - A Three Dimensional GIS for a CAVE Environment. Eileen Vote, Daniel Acevedo, Martha S. Joukowsky, and David H. Laidlaw. *Proceedings of CAA 2000: Computing Archaeology for Understanding the Past. Ljubljana, Slovenia; April, 2000.*

Peer-Reviewed Conference Abstracts, Posters, and Workshops

- Modeling Perceptual Dominance Among Visual Cues in Multilayered Icon-based Scientific Visualizations Daniel Acevedo, Jian Chen, and David Laidlaw. *Poster at IEEE Visualization 2007. Sacramento, California, October-November 2007.*

- Using Visual Design Expertise to Characterize the Effectiveness of 2D Scientific Visualization Methods. Daniel Acevedo, Cullen Jackson, David Laidlaw, and Fritz Drury. *Poster at IEEE Visualization 2005. Minneapolis, Minnesota, October 2005.* BEST POSTER AWARD.

- Color Rapid Prototyping for Diffusion Tensor MRI Visualization. Daniel Acevedo, Song Zhang, David H. Laidlaw, and Chris Bull. *Short Paper on the 7th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI). St. Malo, France, September 2004.*

- Case Studies in Building Custom Input Devices for Virtual Environment Interaction. Joseph LaViola, Daniel Keefe, Robert Zeleznik, and Daniel Acevedo. *IEEE VR 2004 Workshop: Beyond Glove and Wand Based Interaction. Chicago, Illinois, March 2004.*

- Designer-critiqued Comparison of 2D Vector Visualization Methods: A Pilot Study. Cullen Jackson, Daniel Acevedo, David H. Laidlaw, Fritz Drury, Eileen Vote, Daniel Keefe. *ACM SIGGRAPH 2003. Sketches and Applications Proceedings. San Diego, CA. August 2003.*

- Design-by-Example: A Schema for Designing Visualizations Using Examples from Art. Eileen Vote, Daniel Acevedo, Cullen Jackson, Jason Sobel, David H. Laidlaw. *ACM SIGGRAPH 2003 Sketches and Applications Proceedings. San Diego, CA. August 2003.*

- ARCHAVE: A Virtual Environment for Archaeological Research. Daniel Acevedo, Eileen Vote, David H. Laidlaw and Martha S. Joukowsky. *IEEE Visualization 2000, Work in Progress. Salt Lake City, Utah. October 2000.*

To my father, Luis

Preface and Motivation

"The basic flavors were a summing up of the Japanese concept of 'umami', of savoriness, meatiness, mouthwateringness, the bliss-point of any food. 'Umami' is the Japanese fifth taste..."
– Jeffrey Steingarten, in *The Man who ate Everything*

To search for the elements that form the basis of things is natural. There are examples of this in many knowledge disciplines: language and literature ("The Elements of Style"[1]), architecture ("House Thinking"[2]), cooking ("The Elements of Taste"[3]), to name a few.

The goal is the same in all of them: by enumerating, understanding, and describing those elements we have power over them. We can then use them in a controlled way so their combination achieves the purpose we intend in an efficient and effective manner.

The example in cooking is particularly interesting to me, and quite relevant to the theme of this thesis. In "The Elements of Taste"[3], the authors set off to explain the principles behind great taste in a way that anyone could understand. They proceeded to devise a system that included most of the tastes in the modern palate. Not just the four we have been taught we have receptors for, but the ones chefs use to produce their creations.

The same as artists and designers, chefs use their experience and knowledge of the basic components available to develop their recipes. In deconstructing their creative process, "The Elements of Taste" defines fourteen basic tastes: salty, sweet, picante, tangy, vinted, bulby, floral herbal, spiced aromatic, funky, bitter, garden, meaty, oceanic, and starchy. How do taste artists put these together? What are the rules they follow and can we learn them?

This dissertation attempts a similar feat in a completely different area of knowledge, but one also based very much on experience, intuition and the control of a basic set of components with endless combinatory possibilities.

[1] *The Elements of Style*, 3rd edition, by William Strunk Jr., Macmillan, 1979.

[2] *House Thinking*, by Winifred Gallagher, HarperCollins, 2006.

[3] *The Elements of Taste*, by Gray Kunz and Peter Kaminsky, Little Brown and Co., 2001.

Acknowledgements

I would have not completed this thesis without the help and support of many people.

Students, faculty, and staff from the Visualization Research Lab, the Graphics Group and the Computer Science Department in general have made my nine years at Brown and unforgettable experience. I am grateful for the opportunity they all have given me to be part of this family, and I will take the many friendships I have made with me everywhere. I will proudly announce that I once was part of this great group of researchers that helped shape my career, and indeed my life. In particular, I would like to thank the following colleagues and friends for the countless hours they have devoted to helping me with the many gaps in my scientific (and not so scientific) knowledge, and for serving as an inspiration in times of doubt: Daniel Keefe, Cullen Jackson, Eileen Vote, Jason Sobel, Tomer Moscovitch, Joe LaViola, Song Zhang, Liz Marai, Jian Chen, Cagatay Demiralp, Andrew Forsberg, and Bob Zeleznik.

Thank you also to my officemates over the years, Frank Wood, Yanif Ahmad, and Russell Bent, for providing that very much needed a appreciated support on both academic and non-academic endeavors.

I greatly appreciate the effort that my advisor, David Laidlaw, has put into shaping this civil engineer into a (maybe) mildly successful computer scientist. His work ethic and attitude towards his students is something I will try to imitate in my professional life. Always there when I needed him, I cannot imagine a better friend leading me through this journey.

Thank you to the readers of this dissertation, John Hughes and Leslie Welch, for their fantastic comments and helpful critiques. I am also grateful to Andries van Dam, for his help and support during all my time at Brown.

It has been an honor and a privilege to work with all of them.

But life is not only work, specially being from Spain. I really appreciate the time I spent as part of "The Dingers", 2005 Intramural Softball Champions, and "Byte Soccer", 2004 Intramural Soccer Champions.

Thank you to all the members of the "Gapas", my afro-peruvian-latin-rock band. Those late-night jams, and random concerts around town helped immensely in keeping a healthy balance of fun, funk, and work.

Thank you to all my friends and family back in Spain, for their support and understanding. To my mom, Teresa, and my brother Luis Diego and his wife Virginia, thank you for being there for me.

I would like to dedicate this thesis to the memory of my father, Luis Acevedo Martín. He saw me leave home to come to the US and he encouraged me all the way to fulfill this dream. I hope I have made you proud and, although you cannot be here to share this with me, I know you helped me complete it from wherever you are.

Lastly, and most important of all, I would like to thank my wife, María, for letting me follow my dreams and for coming with me to this strange land. Feeling your love and support has made this possible and inspired me to be a better person. Thank you also to my daughter Inés, for providing that last powerful and motivating push to complete this work (in 2 to 3 hour intervals between diaper changes.)

Contents

List of Tables

List of Figures

Chapter 1

Introduction and Contributions

> *"Doctrines and theories are best for weaker moments. In moments of strength, problems are solved intuitively, as if of themselves"*
> – Johannes Itten in *The Elements of Color*

The main goal of this dissertation is to study ways of maximizing the bandwidth of information successfully transmitted by a visualization, while leveraging human competencies so a viewer can understand its visual depiction. In other words, we want to optimize visualization creation by utilizing human visual resources efficiently. To achieve this we quantify and model how human perception explores the types of stimuli present in scientific visualizations. Such a model would serve as a basis for a theory of visualization, one that would explain and predict the way a space of visualizations is organized. This dissertation is an initial step towards the definition of that theory.

A theory of visualization could be described as a theory of effective information representation. This effectiveness is measured by how well that representation allows its users to "detect the expected and discover the unexpected" in their data [Thomas, 2005]. In the process of creating this representation, data are mapped onto visually perceivable units to facilitate their analysis. The goal is to be able to comprehend the data wholistically, through intuitive processing, as opposed to the linear processing required by looking at the raw numerical data.

Figure 1.1 shows an example of that mapping and our vision for this research. This image represents a potential interface for a visualization software utilizing our results. In this case a user would like to visualize together four values from a weather dataset (temperature, pressure, precipitation, and wind speed) all of them represented at the top of the display using the same grayscale representation with low values in black and high values in white. Below those data images are a set of knobs that represent the requirements the user has for each of our four design factors (see below for an explanation of these). Below these knobs

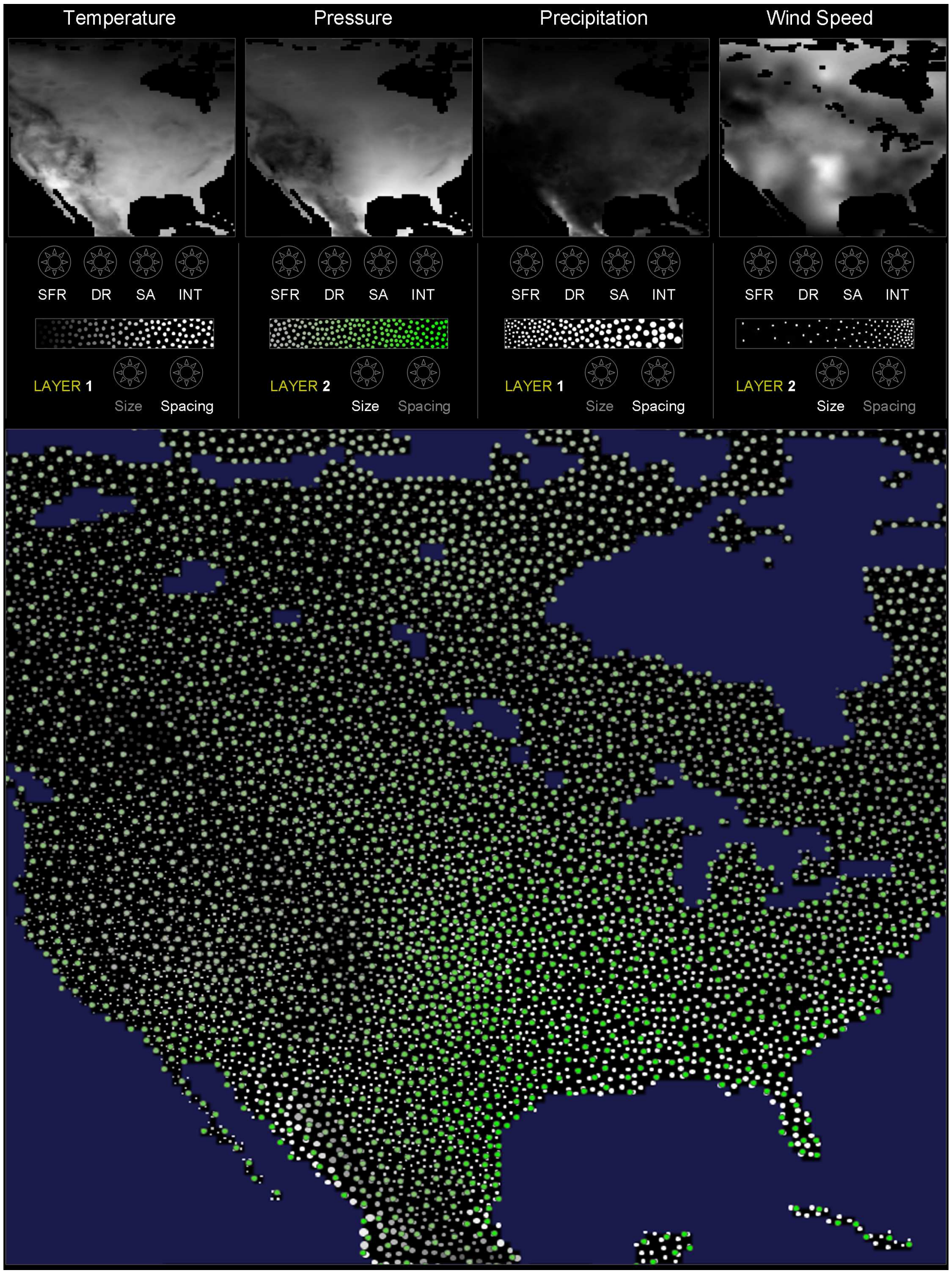


Figure 1.1: Our utility models are based on a set of design factors such as spatial feature resolution (SFR), data resolution (DR), saliency (SA) and perceptual interference (INT). Users should be able to either adjust those knobs and have the mapping parameters (layer order, size, and spacing) change automatically, or fix the latter and observe how the utility values change for each design factor. A more detailed explanation of this figure is included in the text.

are indications of what visualization method has been chosen to represent each data value, what layer this value will be at (numbered from bottom to top), and two more knobs that can control the icon parameters of size and spacing.

We aim to provide users with an indication of what the effects on effectiveness are when they modify the visualization parameters. In our experience, users of visualization software feel overwhelmed by the multiple visualization options available to them. Most packages currently available provide them with a wealth of methods to visualize their data, but they usually provide no guidance at all about which ones are more effective for their goals. Experience, or a visual designer at the user's side, is usually the key to a successful visual representation of the data. Explaining how the elements of the visualization methods work together is the first step to understanding the structure of a space of visualizations, and a crucial one in defining our model of their utility.

To accomplish this we must measure the utility of the different elements that compose our visualization methods, what we call our *visual dimensions*. Having those measurements, our hope is that we would be able to precisely combine these dimensions in such a way that the overall use of their capabilities was optimized to the needs of the data and the requirements of the user. In particular, our hypothesis is:

> Measuring the perceptual capabilities of several icon-based scientific visualization methods for simple single-valued scalar datasets in 2D, and combining that with subjective evaluations of complex multilayered methods representing multivalued datasets, we can generate a predictive model of the perceptual properties of a space of visualization methods.

The key challenge we must overcome for the definition of a general utility model is the extremely fast growth of the number of parameters to explore as we combine visual dimensions to represent multivalued datasets. Although models for single-valued cases have not been developed, and we introduce them here, it can be argued that it is possible to fully study that space. However, starting with two-valued datasets, the complex parameterization of the visualization methods would make exhaustive experiments infeasible. Part of this dissertation's contribution is the definition of a methodology that attempts to describe a utility model for higher-order combinations based on lower-order results.

In order to validate the models we obtain we will perform an evaluation that, using our results, shows an improved and more efficient visualization creation process. Section 6.2 contains the details of this evaluation.

In the rest of this dissertation we will consider a *visualization method* as an abstract function that transforms a scientific dataset into a visual representation to facilitate data exploration. In turn, a *visualization display* is the instantiation of a visualization method.

Here, we are interested in studying visualization methods for multivalued continuous scalar datasets in 2D, using multilayered icon-based methods. Furthermore, the goal of our visualizations is exploratory. We assume our end users want all the data displayed in an unbiased way: they have no preconceptions about more or less interesting areas that should be highlighted or de-emphasized. In the multivalued case, their exploration seeks to understand the relationships among data values.

Defining and exploring the space of possible visualization methods for a given scientific problem has challenged computer scientists, statisticians, geographers, and cognitive scientists for many years; it is still an open research problem. The goal of such models is to describe a searchable space where scientists can find visualization methods that optimally convey the information they require. This dissertation is a modest attempt to generate a utility prediction model of a small subspace. In other words, we seek a function that given a visualization method returns the perceptual capabilities of that method to represent data.

The value of identifying the basic visual dimensions that form visualizations, and their interactions, is that we thus develop a framework to organize knowledge of visualization design and predict behavior of data displays [Cleveland and McGill, 1984]. We want to create a way to get scientists, and visualization users in general, closer to an effective visual representation of their data.

As Watson outlines in [Watson, 2006], to automate the design and creation of visualizations, researchers must identify the particular problem and its constraints, find and capture the heuristics that describe a good solution, and build a tool that finds one or more of those good solutions in the problem space. To these we add, as a first step, the actual definition of the pieces used to build the visualizations, our *visual dimensions*, and, before the search for good solutions, the definition of the measures that characterize the utility of those dimensions, our *design factors*. Once these elements are in place, the process of exploration of the data can really begin by allowing the user to interact with the visualization.

The Visualization Problem: Exploratory Visualization

The basic scientific visualization process involves symbolization, the translation of verbal and numerical information into graphic form [McCleary Jr., 1983], and comprehension, the analysis and understanding of the data presented. Our research is oriented towards developing exploratory data visualization methods, with the goal of visually presenting raw data in a way that prompts visual thinking and knowledge construction [MacEachren and Kraak, 1997]. Understanding and insight are the main goals of scientific data visualization methods, but methods to represent known phenomena (e.g. turbulence in air flow or stress points in a structure) or geared towards performing specific tasks (e.g. finding extrema or identifying a type of turbulent flow) are qualitatively different from visualization methods designed for exploration of the data. Scientists usually utilize the latter during the early stages of their research, when they require visuals that provide a broad understanding of the data being presented. They begin posing hypotheses and asking questions about the data, which lead them towards task-oriented visualization methods for further analysis. Exploratory visualization methods allow them also, in a first approximation, to qualitatively assess the validity of their experimental and data gathering methods. At this stage, visualization is merely a tool to help scientists think about their problem [Hibbard, 2004].

Our focus here is on visualization methods for multivalued scalar scientific datasets in 2D. These datasets are widely used in disciplines such as meteorology, geology, cartography, physics, and engineering. Even when scientists are studying three (or higher) dimensional phenomena, they often rely on 2D slices, such as cutting planes or isosurfaces, to explore and study the datasets. It is usually easier for them to directly understand two-dimensional displays than three-dimensional ones, which would require motion, stereo, or some kind of tracking to be fully perceived. Extensive training and experience is usually required to be able to extract three-dimensional information from a set of 2D slices, a process that doctors master, for example, when analyzing 2D magnetic resonance images of parts of the human body.

With this dissertation we aim to augment the control and understanding of existing visualization methods. We do not aim to create new methods that would replace existing techniques, but to introduce some guidance as to what the perceptual capabilities of those techniques are and how to use them efficiently.

Visual Dimensions

Common practice in scientific visualization is the mapping of scalar quantities to the visual qualities of surfaces containing the data, with color being the predominant example. Other visual qualities that can be used to represent a scalar field on a 2D surface belong not to the surface itself, but to glyphs or icons that can be placed on the surface. Color is again the initial choice for most applications, but size, distribution, and orientation of these icons can also be used to visually represent a scalar field.

The class of visualization methods we are concentrating on includes multilayered Poisson-disk distributed icons where icon size, spacing, lightness, and color saturation can be set to a constant or coupled to data values from a scalar field in 2D. As part of this dissertation we created dedicated software to generate our visualization displays. In it we implemented a Poisson-disk distribution scheme to position the icons according to the underlying data in an efficient way, trying to obtain a uniform distribution of icons with a minimum of noticeable gaps that could be misinterpreted as data features.

Icons have the advantage that they can be layered, increasing the number of variables being simultaneously shown. Even though we are using this methodology to study a very limited space, this framework could be extended to include more complex visual dimensions and even three-dimensional visualization methods.

Design Factors

This dissertation focuses on the creation and evaluation of visualization methods according to a set of design factors. These factors relate to the relative importance of the different variables in the dataset, their relationships, and the quality of visualization needed for each one of them. We define the utility of a visualization method as a function of these factors.

During our experiments we measured and modeled the performance of different methods with respect to our set of design factors. These are important since our visualization problem is the exploration of the data, with no predetermined task in mind. Our factors serve as a characterization of the utility of our visual dimensions. There are many possible ways to do this characterization, but we decided on a manageable set of factors so we could perform our experiments in a reasonable amount of time. Yet the results provide some indication of the expressive power of our visual dimensions and the utility of the visualization methods they form.

1.1 Contributions

This thesis makes contributions towards the quantification of the effectiveness of visualization methods, the identification and exploration of perceptual issues in multivalued visualization, and the definition of a theory of visualization. Dr. Christopher Johnson, in his list of top scientific visualization problems [Johnson, 2004], recognizes all these as some of the major research areas in the visualization field, and stressed the importance of their study in order to advance the state of the art and make visualization grow as a scientific discipline.

This dissertation is further inspired by Fred Brooks's "The Computer Scientist as a Toolsmith II" paper [Brooks, 1996] where he posits that our success as computer scientists must be measured by the success of the users of our applications. In that sense, a successful completion of our research would mean that our visualization software allows scientists to concentrate on data exploration instead of visualization parameter exploration. We facilitate the search for an effective visualization of scientific data by providing a model of performance that, given a set of design goals and a multivalued dataset, can be used to obtain a reasonable initial solution and lead the search through the highly complex space of visualizations.

Furthermore, this dissertation helps clarify how some of the disciplines that take part in the visualization process can be put to work together effectively. The process of effectively representing scientific data involves several disciplines that must be well understood to create useful displays for analysis: data mining, statistical analysis, visual design, perceptual psychology, computer interface design, and human-computer interaction are some of those disciplines. Rarely does a single person have enough expertise in all of these fields to tackle a visualization problem alone, requiring collaborative efforts among a group of experts. In particular, we contribute to advancing the state of the art in three separate disciplines: computer science, perceptual psychology and visual design.

1.1.1 Computer Science Contributions

- We define and quantify a set of design factors that describe the utility of the basic visual dimensions used for exploratory scientific visualization.

- We provide a set of novel predictive models for all our design factors that allow for the effective use of visualization methods based on the precise control over the utility of the visual dimensions that form them.

- We hypothesize and evaluate a novel methodology for building these types of predictive models. This dissertation serves as a proof-of-concept showing how to complete these models for individual methods. It is an important milestone in formalizing models for other types of visualization methods and applications.

- We validate the use of visual design experts as evaluators of scientific visualization methods.

- We evaluate a methodology for knowledge modeling where we tried to capture targeted critiques of visualization methods from visual design experts to incorporate them in a quantitative model.

1.1.2 Perceptual Psychology Contributions

- We extend previous experimental results, in spatial feature resolution and data resolution, to cases that are closer to real visualization displays, which have not been tested before. We provide models of how the data resolution characteristics of our methods change as the independent variables of icon size and spacing change.

- We design an experiment and build a detailed model of saliency and perceptual interference between visual dimensions. We are able to predict and control the relative saliency and interference of the various dimensions on a given display by modifying the size and spacing of the icons, their color, and the order of the layers.

1.1.3 Visual Design Contributions

- We evaluate methodologies to gather and model numerical results obtained from subjective experiments with visual design experts.
- This dissertation is a novel attempt at quantifying the critique process used in art and visual design. We are numerically exploring this experience-based technique for information display evaluation.

1.2 Experimental Methodology

Evaluating the effectiveness of visualization methods is difficult because tests to evaluate them meaningfully are hard to design and execute [Kosara et al., 2003].

Our research involves experiments [1] where participants perform subjective perceptual tasks from which we obtain numerical measures of interactions among visual dimensions. These studies are inspired by psychophysical experiments but geared towards our goal of developing visualization methods for effective data exploration. We also perform subjective studies where expert visual design educators critique visualization methods that use those same dimensions. Our research brings together both experimental approaches by using lessons from the latter to inform the design of the former set of studies.

We do not aim to find a single optimal solution that will exactly match the visualization problem's description. Perceptual psychophysicists and cognitive scientists have been studying human perceptual capabilities for decades, and there are still many unresolved problems. Even those problems that have been explored often find conflicting experimental results that make it difficult to elaborate a complete and solid theory on how humans perceive visual dimensions. It would not be realistic for this dissertation to try to exactly quantify all possible interactions among visual dimensions. Even limiting ourselves to four dimensions, there are many elements that affect the reading of visualization displays, such as interaction techniques or display form factors, that we cannot possibly begin to explore if we hope to succeed in our initial goal.

The use of visual design and artistic expertise to develop visualization methods is widely acknowledged in our discipline. The novelty of this dissertation comes from our goal of

[1] All participants in our experiments were recruited, given informed consent forms, and compensated for their participation, according to Brown University's IRB rules and following our approved IRB protocol titled "Quantifying the Benefits of Scientific Visualization Techniques".

quantifying that expert knowledge in a way that we can combine it with perceptual experiments to build our utility model. We have created a framework for evaluating visualization methods through feedback from expert visual designers and art educators. Our framework mimics the art education process, in which art educators impart artistic and visual design knowledge to their students through critiques of the students' work.

Our use of both perceptual studies and visual design expertise is based on their respective interest in, first, understanding the parameters of use of the different perceptual cues and, second, an optimal utilization of those cues to communicate information.

1.2.1 Perception and Cognitive Science

Perceptual psychologists study how we obtain information from the world around us. In the visual domain in particular, they study how the different visual cues reaching our eyes are recognized, organized and transmitted into our brains for comprehension. Cognition, on the other hand, studies how the information gathered by our senses is put together to form concepts we understand. How much information processing is done where, at the eye level or at the cortex level, is up for debate. What is important from our point of view is their interest in recognizing the individual units of information that get transmitted, and what visual dimensions carry that information efficiently.

This thesis diverges from perceptual psychology in that we are interested not in the pure understanding of how the individual information-transporting units work in general, but how they transport the particular types of information present in scientific visualization displays. This creates a fundamental methodological problem.

To understand those visual cues, perceptual psychologists perform experiments in which they try to isolate their effects as much as possible. This provides unbiased clean information about the individual cues. The idea is to build the knowledge of how they work together from the bottom up: exploring how those cues are combined and what pieces of information are transmitted up stream into the brain for their comprehension. Using this methodology allows for a thorough investigation of the basis of our visual information processing system, but creates problems when complex visual stimuli need to be analyzed. In those cases, it is not clear how to apply the results from very controlled experiments where visual dimensions are studied in isolation.

This thesis presents new methodology that tries to obtain the characterization of those visual cues in the context of more realistic visualization displays. Our experimental results are then more closely applicable to practical situations where multiple cues are combined in otherwise unpredictable ways.

1.2.2 Visual Design and Art

At the other end of the spectrum are visual designers and artists. They study how to present the information in such a way that utilizes the expressive capabilities of each cue in an optimal way. They are experts in what could be described as visual rhetoric which, extending the definition of rhetoric, can be defined as the faculty of discovering and utilizing all the available means of visual persuasion to communicate information effectively.

Their methodology is based on experience and critiques rather than formal psychophysical experiments. After years of study and practice, a visual designer will begin to understand how size gradients, for example, effectively communicate some particular type of information change. They will learn how this visual dimension works in combination with all other visual elements present in their toolbox, e.g. different colors or shapes will affect the reading of those size gradients. Different compositions of the visual display will affect them too. Expert artists and visual designers routinely use this experience-based knowledge to create great art and effective information displays [Grossberg, 2006].

Studying the results of perceptual experiments is part of the visual designer's training, but quickly moving from those to more complex situations allows them to understand the space in which they are operating. In this process, constant critiquing of the effective use of the different cues is the key to their learning process. Given a goal for a particular design, they evaluate how the elements used combine to present the information required. This makes them strive for an efficient use of resources, since any extra information presented, even if redundantly representing a particular message, might create ambiguity in the display and diminish the overall effectiveness. Expert visual designers know that human perception is very good at noticing very intricate patterns and, if a redundancy is present, it could be interpreted as a separate piece of information instead of a reiteration of an already presented one. Along these lines, there is a term in art and visual design called *economy of line*. It means that, when creating a charcoal or pen-and-ink drawing, the least amount of line should be used to show the pose. The expressive power of a well drawn single line is huge. In our case, we will know we have an efficient visualization method when the expressive power of our visual dimensions is used in the right amount to convey the message in our datasets.

In this dissertation we introduce new methodology to gather expert knowledge from visual designers and artists. We utilize these experts to critique visualization methods and evaluate how efficiently those methods use different visual elements. We seek to incorporate their capacity of analyzing the full compositional elements in a display and of evaluating the efficient use of visual cue combinations.

There are, however, two big challenges here. The first is to be able to make their thought process explicit so as to obtain the necessary information we need to build a model of their knowledge. Performing the evaluations in the form of critiques should help in this process, although there is not such thing as a unique and accepted critique methodology, which makes data analysis extremely difficult.

The second challenge is consistency. Our characterization of 2D visualization methods acknowledges that the input we get from the designers is directly targeted at the needs of scientists, and is not about artistic qualities, visual appeal, or aesthetics. However, different experience levels would normally lead to very different critiques of the same visual displays. Our approach to solving this challenge is to use experienced educators that are used to teaching design concepts. They are used to concentrating on the problem at hand, abstracting from aesthetic considerations when they have to focus on what the final goal of the design is; while their results are often appealing and aesthetic, they first have to satisfy the given communication goals.

Finally, one of the main advantages of introducing this type of subjective experiment is the fact that expert designers can not only evaluate the efficient use of visual dimensions, but they can also tell us *why* a method does or does not work and, in most cases, how to fix it by moving along the axes formed by the visual dimensions used in it.

1.3 Research Elements

We have accomplished the contributions outlined above by completing the following elements of the research:

- The development of an interactive software environment for creating the visualization displays needed for this research. It provides users with a text-based interface to create and manipulate multivalued multilayered visualization methods. We developed it in collaboration with Fritz Drury, professor at the Illustration Department at the Rhode island School of Design (RISD), who advised us on what visual dimensions to implement first and how to organize their interplay in the software. The basic software design is an extensible framework for visualization methods in 3D, and includes other non-icon-based visual elements such as color planes and streamlines. It also includes support for vector and tensor-based datasets.

- The design and implementation of a study in which art and illustration experts evaluated six 2D vector visualization methods. We found that these expert critiques mirrored previously recorded experimental results [Laidlaw et al., 2005]; 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; i.e., designers' subjective predictions of user performance by these methods match their measured performance. The experts improved the evaluation by providing insights into the reasons for the effectiveness of each visualization method and suggesting specific improvements. *This was published as a Sketch in ACM SIGGRAPH'03 [Jackson et al., 2003], and has been submitted for publication at IEEE Transactions on Visualization and Computer Graphics Journal [Acevedo et al., 2007b].*

- The design and implementation of a study comparing 2D scalar visualization methods using expert visual design educators as subjects. Based on the experience of the previous study, and after the development of our visualization software, we conducted an initial study to evaluate the utility of 2D visualization methods in terms of a set of design factors, which were subjectively rated by expert visual design educators. We successfully characterized a total of 33 visualization methods using 11 different visual dimensions and 6 different design factors for representing single-variable continuous scalar datasets. This study raised the question of using expert designers, specifically educators, versus non-expert designers as in the previous experiment. The level of understanding of the tasks to be performed and the profusion of comments about why and how to improve some methods increased dramatically in this second experiment. Although not empirically evaluated yet, educators seem to be better subjects for evaluating visualization methods than non-experts. We have not yet utilized expert non-educators to complete our sample. *This was published as a Poster in IEEE Visualization'05 [Acevedo et al., 2005] and received the Best Poster Award at the conference.*

- The design and implementation of an evaluation of a parameterized set of 2D icon-based visualization methods where we quantified how perceptual interactions among visual dimensions (size, spacing, icon lightness, and saturation) affect effective data exploration. In the previous experiment, the difficulty and number of the tasks required, the high variance of the responses obtained, and the small subset of visual dimension combinations tested made our results difficult to generalize. This current experiment improved the tasks by making them more accessible to non-experts, lowering the variance between participants. Of course, this moves away from the critique-inspired methodology towards more quantitative perceptual tasks but, as mentioned before, keeping in mind the application of the results and the type of visualization display we will create. This experiment presents the basic methodology for modeling perceptual interactions among visual dimensions. *This work was published at IEEE Transactions on Visualization and Computer Graphics and presented at the IEEE Visualization'06 [Acevedo and Laidlaw, 2006].*

- An experimental quantification of how factors such as icon size, spacing, layer order and color affect the relative saliency and interference among five different 2D scalar visualization methods: saturation, lightness, orientation, size, and spacing. We define saliency as the perceived dominance of some visualization method over another when representing scientific data. Saliency can be used to visualize the importance of some variables over others: designers may want some variables to dominate the composition while others should recede to the background. Our experiment also recovers the perceptual interference among methods, which we define as the amount of distraction a method creates when users are trying to read another method present in the same display. *This work has been accepted for publication as a Poster in IEEE Visualization'07 [Acevedo et al., 2007a].*

- An experiment to subjectively measure legibility changes in multivalued visualization methods. We used expert visual designers for this experiment so we could rely on their expertise to evaluate many different combinations in a short period of time. We presented them with an interactive application showing two-valued scalar datasets from brain MRI. They critiqued our four visualization methods (size, spacing, lightness and saturation) for how well they maintained legibility of data features when they were combined to show two data variables simultaneously.

1.4 A Hypothesis for our Outcome

Let us reiterate our hypothesis here:

> Measuring the perceptual capabilities of several icon-based scientific visualization methods for simple single-valued scalar datasets in 2D, and combining that with subjective evaluations of complex multilayered methods representing multivalued datasets, we can generate a predictive model of utility of a space of visualization methods.

We successfully characterized the utility of individual methods and obtained predictive models for their use. We even obtained predictive models of their relative saliency and perceptual interference when used in pairs. We also validated the hypothesis that using expert visual designers to subjectively evaluate scientific visualization methods can yield significantly comparable results to objective task-based quantitative studies. After all these encouraging results, our next step, utilizing expert visual designers to evaluate complex combinations of visualization methods, was not successful in validating our initial hypothesis.

There are two possible reasons for this outcome. The first is that our hypothesis is partially false, since we could not disprove the null hypothesis for its final statement ("...combining that with subjective evaluations of complex multilayered methods representing multivalued datasets..."). The second possible reason is that our methodology for evaluating such hypothesis did not have the power to capture a significant result when, in fact, there is one. Let us briefly look at these two scenarios here. A more extensive explanation is provided in Chapter 6 of this dissertation.

Our objective was grandiose: to explain how visualizations work in general, independent of the application. To try to accomplish it, we constrained the problem to 2D cases and to a very specific set of methods, so our experiments were feasible. This made us switch the characterization of our research to become a proof-of-concept: we would demonstrate how to accomplish the final goal through our exploration of this small sample of the full space.

Our problem was, as it might be clear to the reader, ill-posed. Even trying to study very few of those visual dimensions created an uncontrollable exponential growth in the number of combinations that should be explored to reliably describe the space they form. Even assuming this exploration was done, an effective solution to a given multivalued visualization problem might not exist and, even if it does, it usually would not be unique.

However, we based our research on the assumption that the potential solution to the visualization problem depends continuously on the perceptual characteristics of the individual dimensions that form such a result. Thus, we were attempting to obtain a mathematical model that explained and predicted how those characteristics work in complex situations.

Furthermore, we believe that our strategy of combining perceptual and artistic knowledge to build our model was the right choice to achieve that goal. Through psychophysical studies we can quantify how users perceive different aspects of the data when using different methods to represent it. It also seems reasonable to engage the expertise of visual designers and artists, since they are living proof that it is possible to present very complex information in an effective way. They use the same perceptual capabilities we measure through our experiments, yet they are able to combine it and analyze it in such a way that they can solve complex visual problems based on their experience, without the aid of equations that map out the space in which they are moving.

In summary, we believe the hypothesis is valid and can be successfully evaluated. If this is correct, our approach to engaging expert visual designers and capturing their knowledge could be at fault.

Given our experience throughout all these experiments, we believe the main reason for our outcome was the initial assumption of independence from the data. Our set of design factors, and the exploratory characterization of the visualization problem, allowed us to abstract as much as we could from defining a specific task for our visualization displays. Yet, from the beginning of our collaboration with designers, they complained about not having a specific use for the data being represented.

Furthermore, the fact that we were trying to obtain a general model meant we could not fixate on specific instances of our datasets. We tried to accomplish this by presenting our experimental stimuli using several different datasets and asking designers to subjectively evaluate methods for a general dataset. This was acceptable, for the most part, when exploring single-valued visualizations. But, the moment real perceptual interactions came into play during our last experiment, using two-valued brain MRI datasets, the problem of abstracting from the specific data combinations became clear.

Methods that would seem to work for most cases would be found to fail for particularly unfavorable combinations of datasets. Also, the similar spatial distribution of the data across values lead to the question of whether results from this experiment would apply for datasets with non-correlated spatial distributions. Participants commented that, most likely, the opposite situation would also be controversial depending on the particular distribution of values for all data variables.

In summary, without including the data values' spatial distribution characteristics and defining a general metric for data variable interactions, we were unable to model the utility of our space of visualization methods. At the end of this dissertation we offer some hope for accomplishing this by describing a possible way of adding the data characteristics as an extra set of axis in our visualization space.

The potential for a great contribution at the end of this research was not our only motivating factor. The experiments we have conducted add up to be an important methodological framework with which other visual dimensions can be explored. Through those experiments, as we will present in this dissertation, much has been learned that the visualization, perception, and visual design communities can build upon. The overall results do not add up to a full fledged model to be plugged into a visualization software, but our individual experimental results can help non-expert users in their search for an effective visualization, by providing some indication about probable directions of improvement for their visualizations, and by shedding some light as to what methods to use in what situations.

1.5 Organization of this Dissertation

Chapter 2 introduces some definitions and notation that will be used throughout this dissertation.

After that introductory chapter, Chapter 3 explores the extensive literature related to this project and how we shaped our investigations based on previous work.

Chapters 4 and 5 form the main portion of this dissertation. They describe the experiments we have conducted and their results. First we go through our initial studies using visual designers to evaluate scientific visualization methods. After that we present our psychophysical experiments to measure perceptual interactions among visual dimensions.

Chapter 6 presents the design and outcome of our last experiment, where we attempted to use expert visual designers to evaluate visualizations of two-valued scalar datasets. Given the outcome we explained before, this chapter also includes the results of an informal evaluation of our existing models presenting real datasets to scientists for exploration.

This will be followed, in Chapter 7, by a final discussion of our results, the impact of this dissertation, what the future lines of research would be, and some conclusions.

Chapter 2

The Elements of Visualization

This chapter clarifies the scope of the thesis and puts each of the components of our research in context: the type of visualization problems we are dealing with, the type of visual dimensions we consider, and our definition of design factors. Although it introduces very basic concepts, the discussion in this chapter helps in understanding our research plan and the decisions made along the way.

Every visualization process starts with a question about some characteristic of the dataset a scientist has just compiled and needs to study [Springmeyer et al., 1992]. After performing some filtering of the data, and maybe some statistical analysis or data mining, it becomes clear that looking at a set of numbers does not help the scientist understand what the data contains: a visual representation is needed.

The first step is figuring out how to translate the numbers into visual entities so it is easy to explore the relationships among all the variables in the dataset. The components used to create that representation are our visual dimensions. Color, shape, size and movement are examples of some of those dimensions.

There are many ways of combining those dimensions to show the data, and scientists need some guidance in deciding which mapping, from numerical data to visual dimensions, is appropriate for his or her visualization problem. Our design factors characterize the capabilities each of the dimensions have for representing data.

Let us review these three pieces of the visualization process (the problem, the visual dimensions and the design factors) one by one.

2.1 The Visualization Problem

The visualization problem has two distinct components: the goal for the creation of a visual representation of a dataset and the type of dataset being visualized. We will discuss these in the following subsections.

2.1.1 The Goal of the Visualization

What is the goal of the visualization display? This is the first question we must answer when we want to transform the set of numbers that form our dataset into a visual representation. Maybe the goal is to check the dataset for problems or obvious errors. Maybe we want to highlight some parts of it, like extrema or areas below a certain threshold. Maybe we are searching for a specific pattern that indicates some interesting phenomena are happening. All these examples would require, at first glance, a dedicated visualization design that would translate the numerical data into a visual representation that fulfills the requirements. We can classify these visualization problems into two main categories: explanatory problems and exploratory problems.

In cases where the goal is to show specific characteristics of the data and we know how to find them, or when we want to show the results of an experiment that revealed some unexpected patterns, our visualization problem is explanatory. There are certain things in the data that require the viewers attention, and the visualization method used should lead the viewer to them.

On the other hand, there are occasions in which the end-user just wants to see whether the data coming out of the experiment *looks OK* or if there are some errors in it. In this case the approach is one of exploration. The end-user wants all the data presented in front of her in an unbiased way. There are no preconceptions about more or less interesting areas that should be highlighted or blurred. These exploratory visualization problems are the ones that we are addressing in this dissertation.

In some sense, the lack of a clear task to be performed by our visualization users makes our job more difficult. The fact that they just want to visually absorb everything the data has to offer without creating biases is a big challenge. For example, local maxima of a dataset can be marked using visually salient icons. Their numerical values can even be displayed beside the icons. There can certainly be a discussion about the design of such icons, the placement of the numerical value, etc., but the visualization problem is clear.

The goal of exploratory visualization is to gain insight into how the data are spatially organized. In the multivalued case, exploration seeks an understanding of the relationships among data variables. Once these are presented, the visualization user will begin asking more explanatory questions, derived from the insight gained and requiring, in general, a different type of visual display that helps support his or her arguments.

2.1.2 The Type of Dataset Being Visualized

Once we have the main goal the visual display must fulfill, we must take into account the scientific problem we are trying to address. In other words: what type of dataset are we

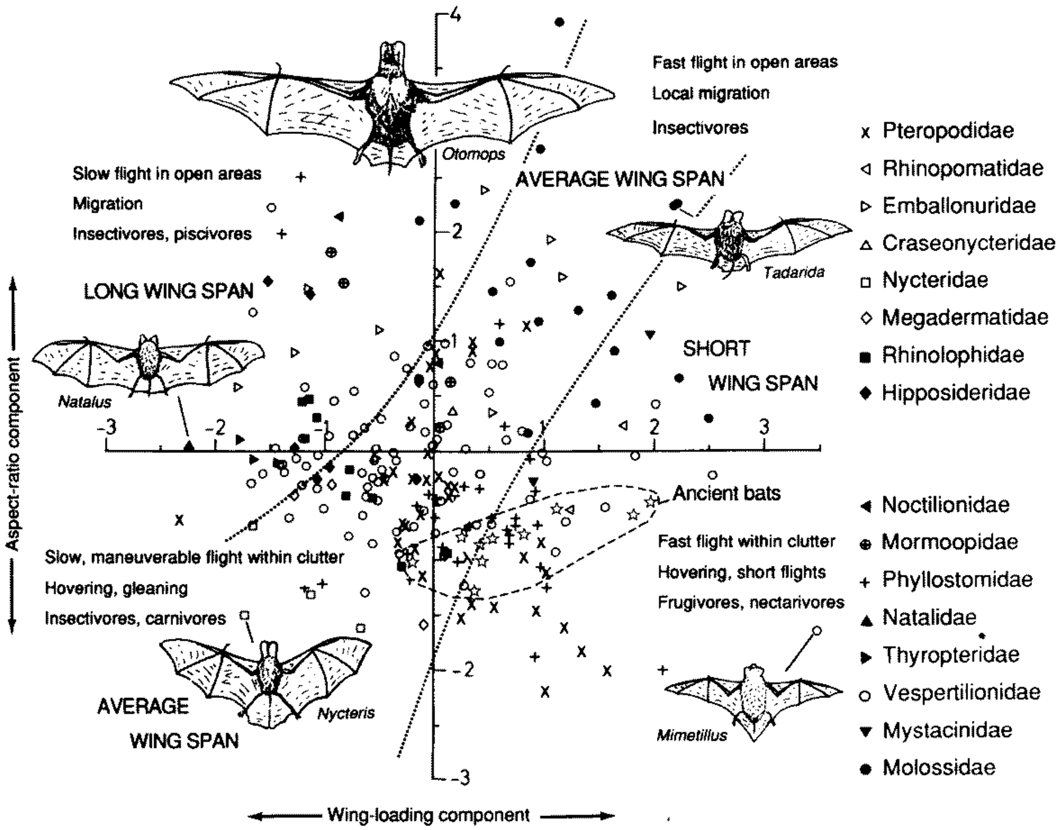


Figure 2.1: An information visualization display showing and example of non-spatial data from [Norberg and Rayner, 1987] presented in a Cartesian grid designed ad-hoc for this visualization. It displays different anatomical and flight-related variables for many species of bats to try to discover a correlation between size of the bats, their flight speed and their behavior. Shape and location are used as visual dimensions in this case.

dealing with?

Information Visualization vs. Scientific Visualization

There is much debate in the community about the distinction between these two types of visualization [Rhyne, 2003]. The annual IEEE conference on visualization is divided in two to distinguish between research in one area or the other.

Information visualization deals with datasets that do not have an inherent spatial component or that, having one, represent abstract non-physical data. On the first case, a visual representation of those datasets must be made in an abstract space delimited by some of the variables present in the data (see Figure 2.1). The second case is more debated since it has clear spatial reference, such as the geographical area indicated by the map in Figure 2.2, but includes non-physical information (in the same figure, the types of conflicts are indicated

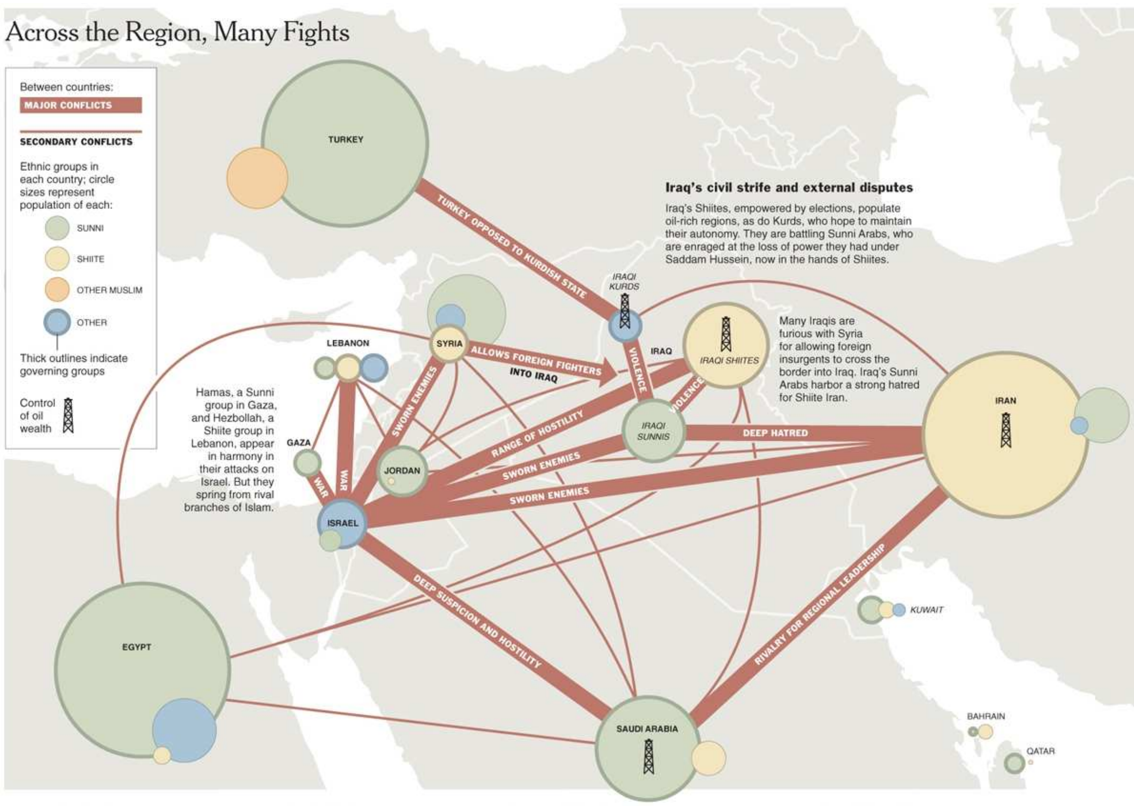


Figure 2.2: An information visualization display from the New York Times that shows several types of data using size, color, and outlines on top of a map, which provides the spatial information component (Copyright 2006 The New York Times Company)

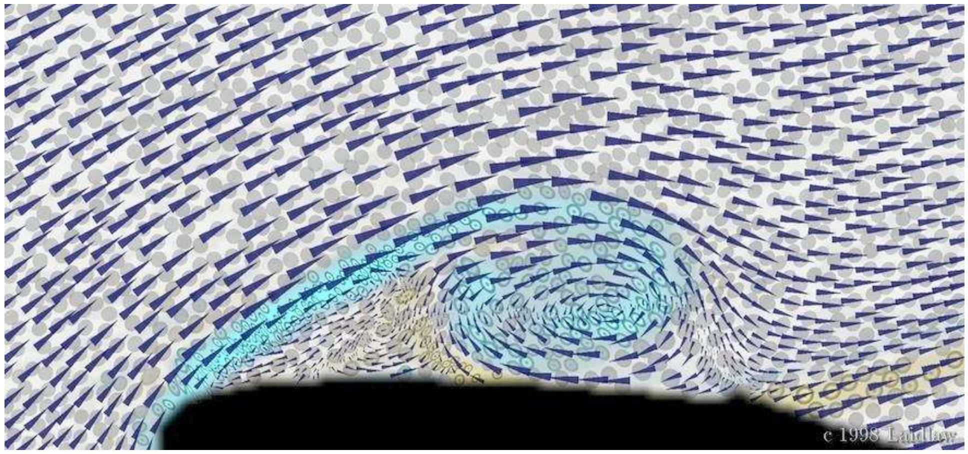


Figure 2.3: Visualization of experimental 2D flow past an airfoil [Kirby et al., 1999a]. Six different variables of the flow are visible at every point in this image. It shows relationships among the values that can verify known properties of this particular flow or suggest new relationships between derived quantities.

by lines.)

When the dataset contains information that has a clear spatial component and involves physical phenomena, we have a scientific visualization display (see Figure 2.3). Even using the same spatial context as Figure 2.2, the origin of the data and their characteristics provide the visual display different qualities (see Figure 2.4)

The distinction is by no means clear cut. The information in the examples of Figures 2.1 and 2.2 is not un-scientific, but the data are qualitatively different from the examples of Figures 2.3 and 2.4. This dissertation is aimed at developing better, perceptually effective scientific visualization methods.

Data Continuity

Another key characteristic of the data is whether they are continuous or discrete. Continuous data can be queried and visually represented at every point in the region of space where it resides. Discrete data, on the other hand, correspond to measurements at specific spatial locations. The data to be visualized might come from the interpolation of values gathered in those discrete points; the temperature readings used to create Figure 2.4 were obviously discrete, but the dataset being visualized is the interpolated one, making it continuous.

There are also continuous data being visualized discretely (see Figure 2.5) but the intention of the visual display is to have the user do the interpolation.

Data continuity can sometimes serves as a distinction between information and scientific visualization displays. Since physical quantities such as temperature or pressure can only be measured in discrete locations but exist in every point in space, we consider them continuous

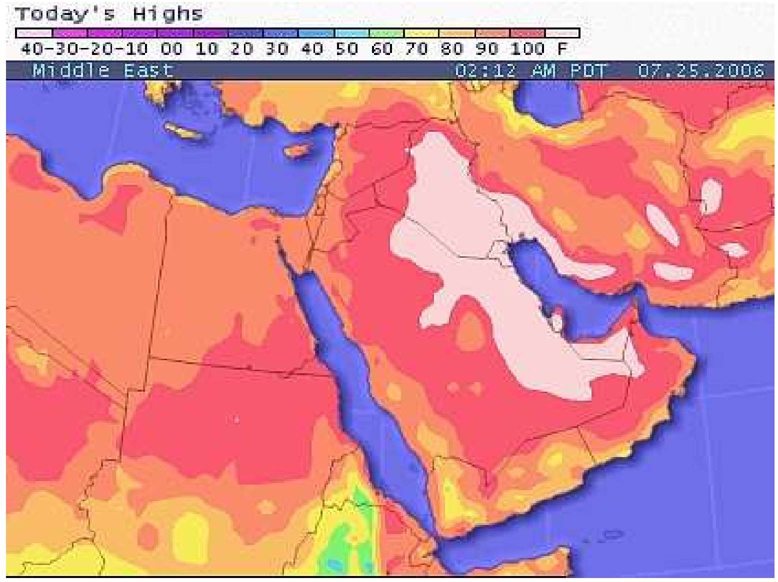


Figure 2.4: Example of a very simple scientific visualization display using the same spatial reference as Figure 2.2.

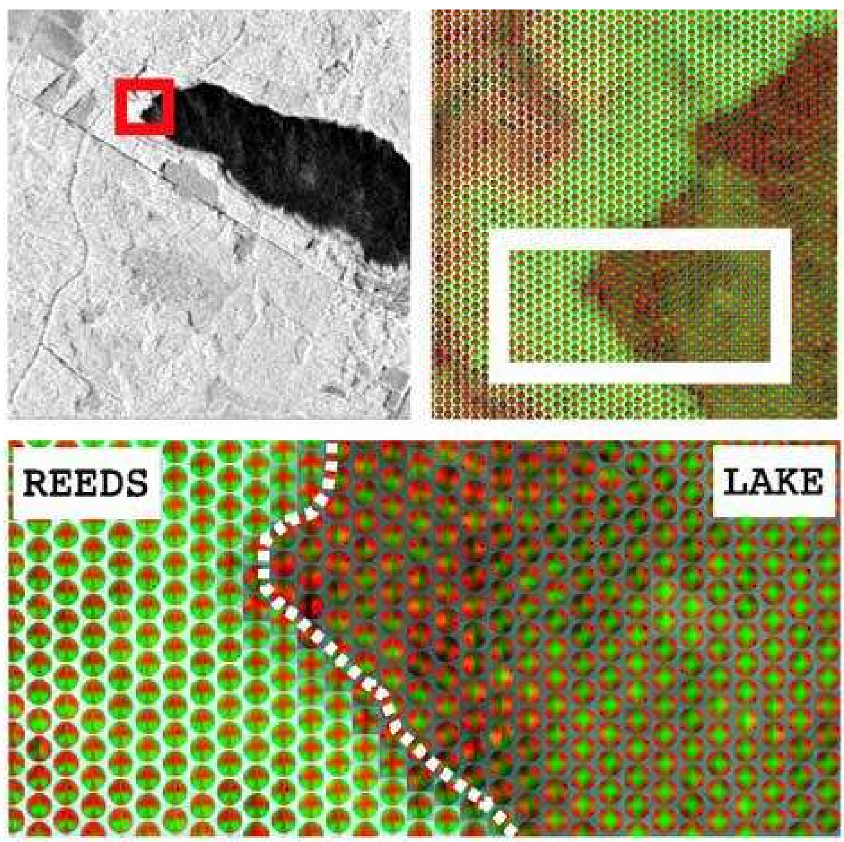


Figure 2.5: Example of continuous data being visualized using discrete glyphs. In this case SAR polarimetric response patterns indicate different types of surface cover. A coherent change in response pattern between the lake surface and reed beds can be detected [Woodhouse et al., 2002].

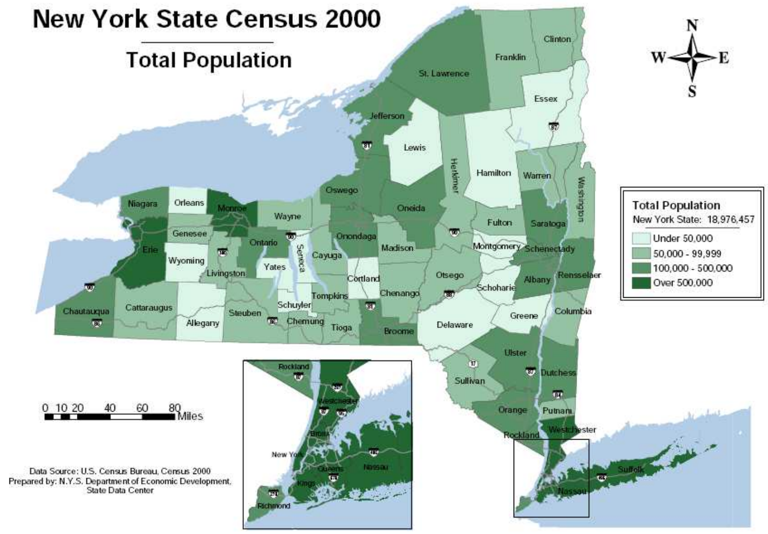


Figure 2.6: Example of discrete non-interpolable data. This is also an example of information visualization data which clearly has a spatial component.

and scientific visualization methods are responsible for visualizing them. Data variables measured discretely but that cannot be interpolated, like number of people living in an area (see Figure 2.6), are the responsibility of information visualization methods.

For this thesis we will consider only 2D interpolated continuous datasets. Since our visualization methods are icon-based, we will be discretizing the interpolated datasets in order to facilitate layering of multiple variables in the same display. Figure 2.5 accomplishes this by creating a complex glyph that incorporates multiple variables. Some of our design factors deal with this loss of spatial feature resolution due to this discretization. Perceptually measuring this quality for our visualization methods will allow us to optimize our visualization results depending upon the importance of the different requirements: e.g. is exploring multiple variables more important than higher information frequency?

Data Characteristics

There are two more data characteristics that will define the visualization problem. The first one differentiates between qualitative and quantitative data. The former include elements of different classes that may or may not be ordered and that have no inherent numerical relationship among them. Quantitative data, on the other hand, maintain a mathematical relationship among all data elements.

There are three types of quantitative data depending on the number of values each data element has. In a scalar dataset only one value exists at each location in space where data are present. A vector dataset contains elements with as many values as the dimensionality

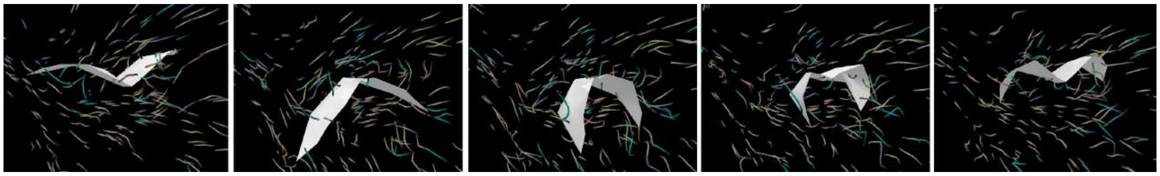


Figure 2.7: Particle flurries in a Virtual Reality visualization of air flow around a bat in flight [Sobel et al., 2004]. Over a short time, the particle animation gives a synoptic visualization of the main features of the three dimensional vector field.

of the space they are in, i.e. in 2D, vector elements have two components; three in 3D, and so on. A 2nd order tensor dataset contains elements with more values per spatial location, e.g. four values per element in 2D, and nine in 3D.

The last characteristic of the data that defines the visualization problem is the space in which the data live. Whether it is 1D, 2D, 3D, or more, the visualization display must be adapted to the requirements of the data space. Time can also be included as a dimension here. Dynamic datasets require a very different treatment than static ones since correlations across time-steps and maintaining temporal coherence in the visualization method become keys to avoid distracting the user with artifacts not really belonging to the data.

Summary

The previously described characteristics of the visualization problem shape the visualization methods that can be applied to it. The discussion up to this point is intended to briefly introduce the vast number of visualization challenges that datasets in all those different spaces pose for the creation of effective visualization methods.

For this dissertation we chose to constrain our model to continuous scalar datasets in 2D. This might seem, on first glance, a simplification of the problem, given that there are many researchers that have already tackled much more complex types of datasets and developed successful visualization methodologies. Some examples of complex datasets successfully being visualized are shown in Figures 2.7, 2.8, and 2.9, where different visual techniques, display form factors, and interactions, combine to form the various visualization methods.

Our choice is based on the fact that the basic components of all those visualizations visually interact in ways that are still not clearly understood. Almost all great visualization methods must be iterated upon until a solution is reached, which should effectively show the information required and minimize perceptual issues. We want to analyze those issues from the ground up. By constraining ourselves to more manageable datasets, we can be more thorough in the analysis of perceptual interactions, eliminating from the experiments a multitude of dependent variables. This is not to say that those variables, such as the type of display used, the interaction techniques, etc., are not important, but they should

Figure 2.8: Three different visualizations of the diffusion tensor magnetic resonance imaging data of a brain. This tensor field data can be explored using an immersive VR environment such as a CAVE (left) [Zhang et al., 2003], a fishtank-VR setup (center) [Demiralp et al., 2006], or a physical model created through color rapid prototyping (right) [Acevedo et al., 2004].

Figure 2.9: Virtual Reality is again used here to visualize archaeological excavation data. Colors indicate the type of artifact while size and quantity indicate other abstract variables present in the dataset. From left to right, a view of the excavation site, an overview of the full dataset, and two different moments of a user interacting with the system. In this case, the need to perform spatial correlations were the key to using a VR environment [Acevedo et al., 2001; Vote et al., 2002].

be added to the experiment once other basic visual characteristics are well understood.

2.2 The Visual Dimensions

Each of the types of data detailed in the previous section will potentially require a different visualization method. The choice of method must be made based upon their strengths and weaknesses, keeping in mind the underlying pledge that every visualization method must display the data truthfully and avoid misleading the viewer. Fred Brooks summarized this point beautifully in a talk given during IEEE SIGGRAPH 2003:

"Visualize to inform, not to impress. If you really inform, you will impress."

Going from numbers to pictures is usually the first step in the exploration of any scientific dataset, and choosing the right tool for the job can be the difference between success or failure of a scientific enquiry. As an example of this, Tufte suggests how the space shuttle Challenger's catastrophic launch in January of 1986 could have been avoided if only the available data had been presented correctly [Tufte, 1997]. The visualization problem, which includes the goal of the visualization and the type of data, will partly determine this choice. Nevertheless, the available choices are many and their differences in terms of visualization utility are not well understood, especially when used in combination.

The visual dimensions are the toolset we use to create the visual representation of the data. There are non-visual dimensions we could be using to represent data, such as sound or haptic interfaces, but this dissertation is aimed at a small subset of just visual dimensions.

We will define these dimensions as the elements that, more or less independently, can be used to create a visual representation of a scientific dataset. This is to say, and for the purposes of this dissertation, that continuous scalar data variables must be mappable to them.

Note that even when only a few of these dimensions get mapped to data variables in any particular visualization method, all of them will be present in the final visualization display. For example, it is obvious that even when size is not used to represent any data variable in the dataset, all icons must have a certain size, either constant or randomized, across the display.

For practical reasons we chose to limit our experiments to five visual dimensions: icon lightness, icon size, icon spacing, icon orientation and icon color saturation (see Fig. 2.10). This decision was made to decrease issues due to participant fatigue during experiments and to provide a relevant sampling of the space of visualization methods. The reasons to choose these five particular dimensions are diverse. Size and spacing are elements that received a highly varied set of reactions during our initial experiments [Acevedo et al., 2005].

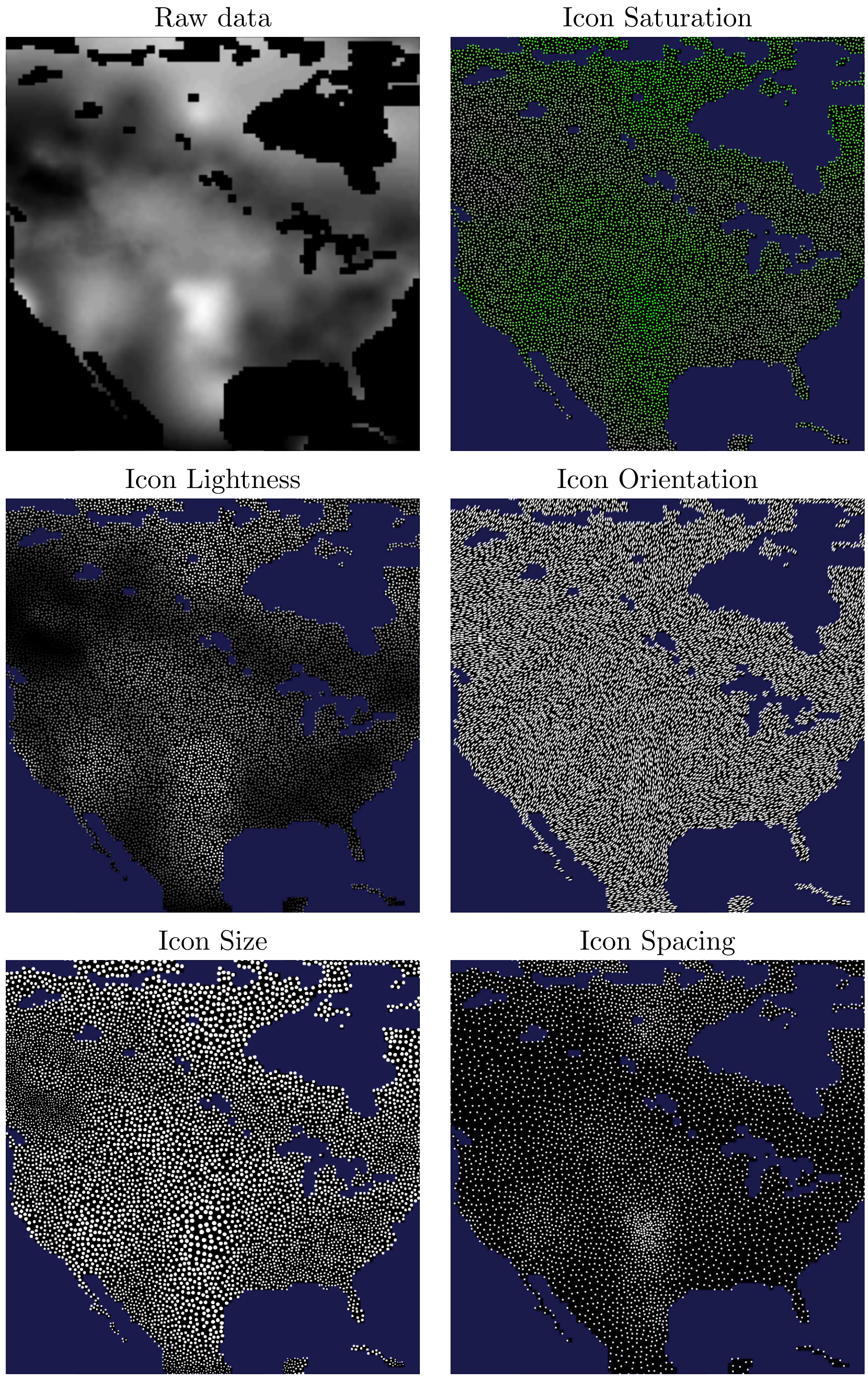


Figure 2.10: Visual dimensions. We are quantifying and modeling the utility of icon saturation, lightness, orientation, size, and spacing when representing scalar datasets in 2D. These five dimensions are demonstrated here representing the same single-valued dataset shown at the top left.

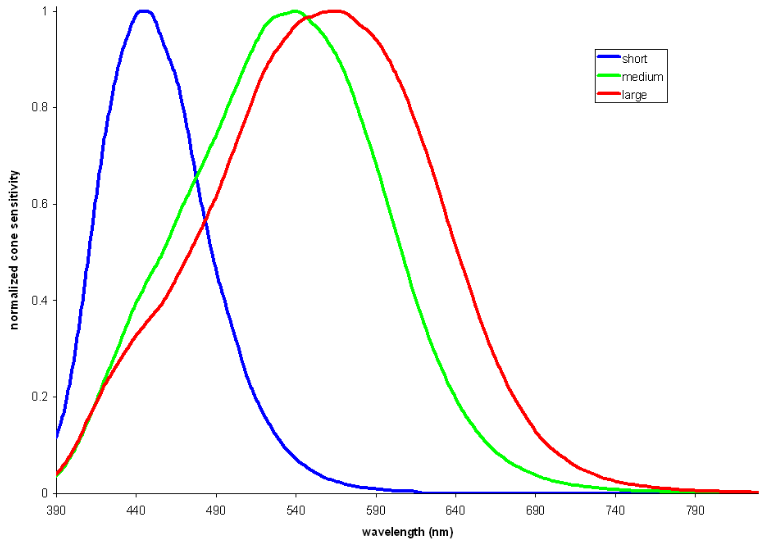
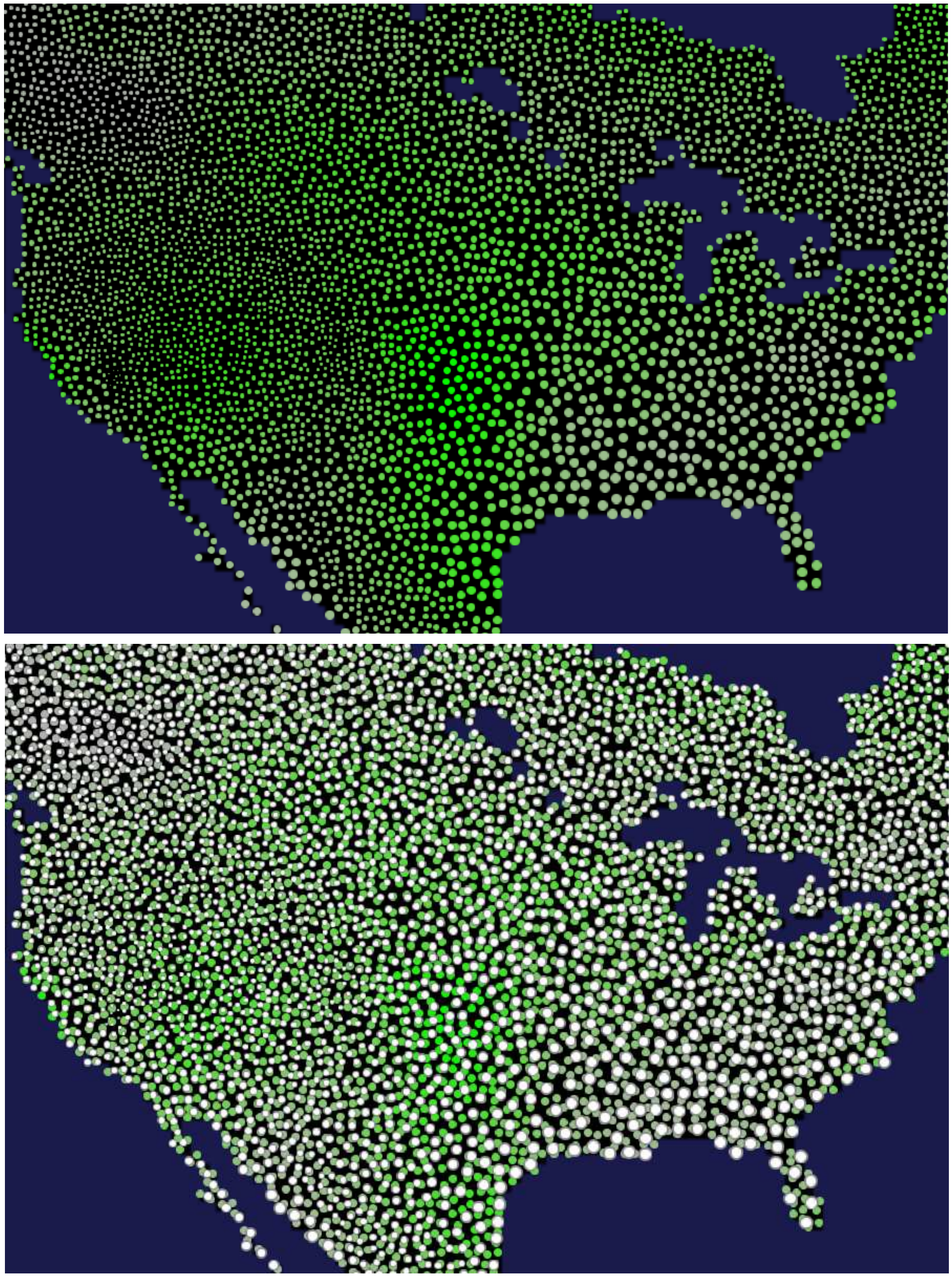


Figure 2.11: Cone response sensitivities normalized as a proportion of the peak response for each type of cone cells. Our eyes are most sensitive to "yellow green" wavelengths around the middle of the spectrum. In fact, most of our light sensitivity lies between 500 to 620 nm – roughly from "blue green" to "scarlet"

Also, very few studies have been published exploring these two elements together [Wolfe, 1998]. Icon lightness was chosen because it is an element that has been studied in depth, allowing us to compare our results with previous experiments. Icon orientation has had a lot of attention in the perceptual psychology literature related to texture discrimination but its use for scientific visualization is limited to a few studies. It is, from our experience during the first experiments with visual designers, a difficult element to perceive as a scalar, albeit a very salient one. Finally color saturation provides our link with the use of color for visualization. Including hue would bring in a lot of different issues related to the use of color spaces that would complicate the main focus of this dissertation. Saturation, being a very much neglected visual dimension for scientific visualization, provides a convenient middle ground to introduce color in our experiments and explore its capabilities for visualization use. In particular, we decided to choose a green hue (with a 0.6 lightness value) because our eyes are more sensitive to light around that wavelength (see Fig. 2.11). We use CIELab color space to generate our visualizations, so we can control that constant lightness is achieved throughout the range of saturation at this hue.

Apart from these five dimensions mentioned, another one that we will consider will be the number of layers a visualization method utilizes. For example, if two data variables have to be represented in the same display and we want to use size and color saturation, these can be accomplished on one or two layers, as Figure 2.12 shows. We are including the



a)
b)

Figure 2.12: Examples of a two-valued dataset visualizations using size and saturation in a single layer (a), and in two layers (b). The data mappings do not change from one to two layers. Note the gray borders on the icons of the top layer in (b). Our experiments will quantify which of these representations works best depending upon what the visualization requirements are.

number of layers as a visual dimension in our models. Furthermore, to differentiate among layers and avoid, as much as possible, simultaneous contrast issues, the icons on our top layer will have a small gray border at half the lightness value of the icon color.

To summarize, our analysis of interactions among visual dimensions represents the study of how independent these dimensions are of each other. Since these dimensions represent the axes of a space of visualization methods, finding and quantifying those interactions would lead to an understanding of how orthogonal those axes are. An ideal space would have orthogonal axes that could be used independently of each other when creating visualization methods.

2.2.1 Definitions and Nomenclature

A visualization method takes a scientific dataset and produces a visualization display. We define a space, $\mathbb{V}$, of scientific visualization methods. In general, our space includes layered iconic representations of 2D multivalued data. The visual dimensions that are present in each layer are:

- Icon color hue (v_0)
- Icon color saturation (v_1)
- Icon color lightness (v_2)
- Icon transparency (v_3)
- Icon orientation (v_4)
- Icon size (v_5)
- Icon spacing (v_6)

A visualization method, $v \in \mathbb{V}$ maps data values to visual dimensions. We can combine multiple layers, which we will denominate l_k. Each one of these layers will contain all 7 visual dimensions defined above. The subscript $_k$ of the layer indicates its order in the final visualization, from back to front:

$$v = \{l_0, l_1, l_2, ...\} \in \mathbb{V}$$

where,

$$l_k = \{(m_0, m_1, ..., m_6), (r_0, r_1, ..., r_6)\}$$

Each component of l_k refers to one of the 7 visual dimensions v_i:

$$m_i = \begin{cases} 0 & v_i \text{ is not mapped} \\ d_i & v_i \text{ is mapped} \end{cases}$$

$$r_i = \begin{cases} c_i \in \mathbb{R} \in [0,1] & m_i = 0 \\ (b_i, e_i) \in \mathbb{R}^2 \in ([0,1],[0,1]) & m_i \neq 0 \end{cases}$$

where d_i is the index of the data variable mapped into v_i.

Figure 2.13 shows two examples of the use of this nomenclature on the previous two visualization displays. Note that background color, shape and border of the icons can also be controlled but we do not need to include them explicitly in the parameterizations: background is always black, icons are always circular, and borders are always gray at half the lightness of the icon color, with a width between 20% (for large 10-pixel icons) and 25% (for small 2-pixel icons) of the the icon's diameter.

2.2.2 Icon Placement

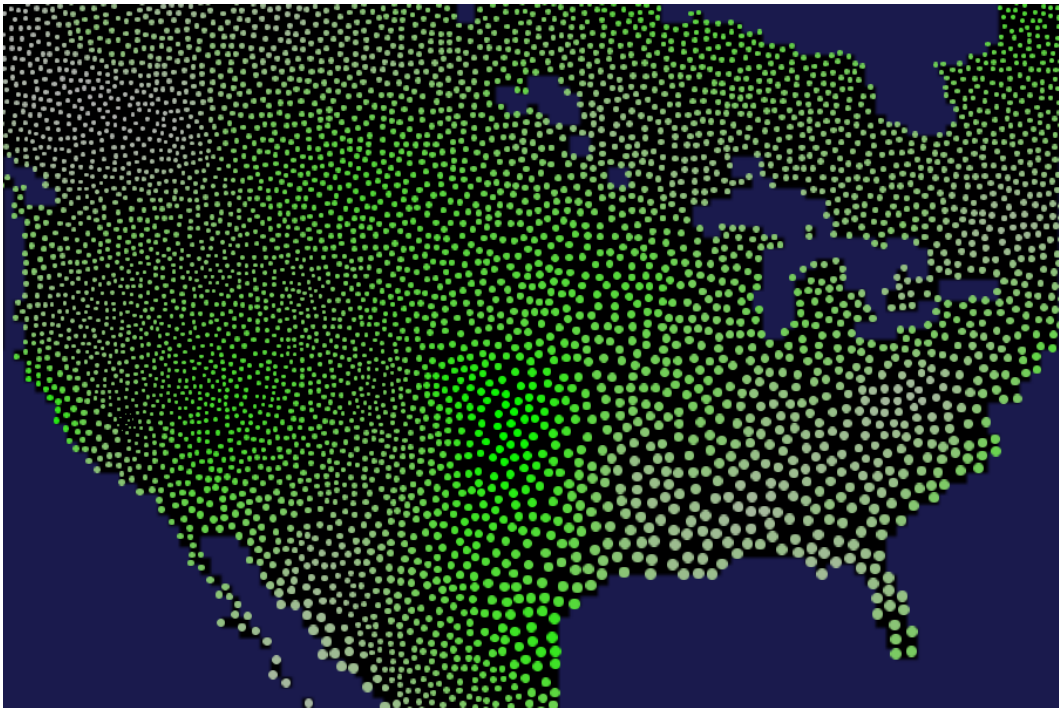
We use a Poisson disk distribution [McCool and Fiume, 1992] to randomly place icons on the display. This distribution can be defined as the limit of a uniform sampling process with a minimum-distance rejection criterion. Successive points are independently drawn from a uniform distribution over the 2D domain where the data exists. If a point is at least distance r from all points in the set of accepted points, it is added to that set. Otherwise, it is rejected.

In our case the points correspond to the center of our icons, and the choice of r comes from the radius of the icons plus half the spacing value. When either or both of these dimensions are mapped to a data variable, the corresponding value is evaluated at the center point of the icon and r is determined to check for overlaps with previously placed icons.

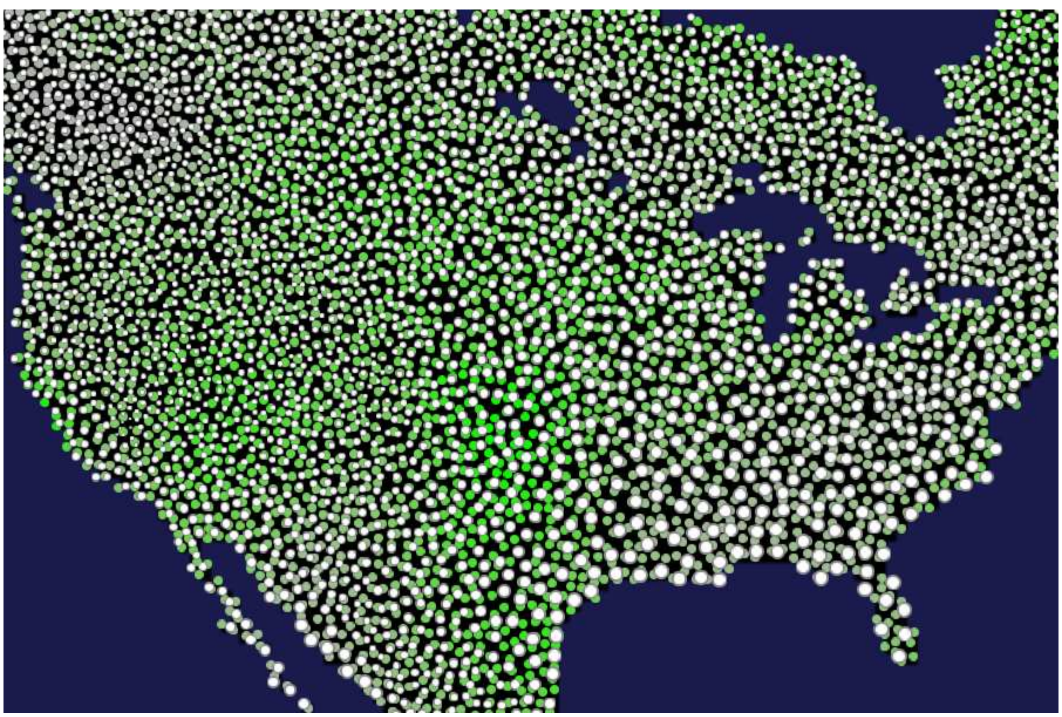
While a straightforward implementation of the algorithm would require a certain number of failures (overlaps) in a row as a stopping condition, that would not guarantee that we are evaluating all pixels in our display. This could lead to many areas remaining untested and obvious holes in the visual display. We remedy this by first randomizing a full list of pixel center locations and going through it in order.

Further, we super-sample our domain (2 times is sufficient to obtain a significant improvement and not extend the running time too much) and perform all our overlap calculations in floating point coordinates, instead of pixel space. Limiting the size of our icons to a minimum diameter of 2 pixels, along with this super-sampling scheme, avoids the appearance of a regular grid bias in our modified location-sampling algorithm using just pixel center locations. Once we have a full set of icon locations in floating point, we let our graphics engine perform the necessary antialiasing for display.

Finally, we chose this placement algorithm instead of a regular grid for two reasons. The first is that clear creases would appear when either size or spacing were tied to a data variable. This would confuse users of the display by creating false features in the display that do not correspond to the data. Further, even when size and spacing were constant



$v = \{l_0\} = \{(0, 1, 0, 0, 0, 2, 0), (0.33, (0, 1), 0.6, 0, 0, (2, 10), 1)\}$



$$v = \{l_0, l_1\} = \{\{(0, 1, 0, 0, 0, 0, 0), (0.33, (0, 1), 0.6, 0, 0, 8, 0.5)\}, \{(0, 0, 0, 0, 0, 2, 0), (0, 0, 1, 0, 0, (2, 10), 1)\}\}$$

Figure 2.13: The same two displays from Fig. 2.12. Their respective full parameterizations are shown here.

across the display, having a regular grid would impose a structure to the data that is not present.

We believe, and our expert visual designers confirmed this hypothesis, that having this quasi-optimal random placement leads to less confusion that a regular grid would. In their own words: "It is hard to get away from this horizontal and vertical organization to try and extract non-aligned structures in the data".

2.2.3 EVOLVIS: A Visualization Language

To create the actual visual displays for this dissertation we developed a language to describe scientific visualizations of multivalued, two-dimensional datasets. Our goal was to create a language with which a user could quickly create complex and precise data-driven visualizations as well as facilitate their modification during the iterative design process. We called this language and the rendering system to display it EVOLVIS.

EVOLVIS is a general tool that can combine three types of basic visual elements – discrete icons, color planes, and streamlines– into layers. A text file fully describes the resulting method and controls the layering of the different elements, their appearance, and their spacing, including the mapping assignments of any of these visual dimensions to data variables. In addition the language supports extensions to accommodate scalar, vector, and tensor data in 2D and 3D.

For the purposes of this dissertation, only icon elements will be used, although color planes were also used for some of the initial studies with expert visual designers. Other visual dimensions that are possible with our language include border width and color, texture mapping, and more complex icon shapes that can be controlled by the data. Figure 2.14 shows some examples of the possibilities EVOLVIS provides.

2.3 The Design Factors

In order to represent the exploratory nature of our visualization methods we must establish a set of design factors that can be used to characterize the utility of a given visualization method and do not constrain our model to any particular task.

Once we have these factors, we will be able to quantify how the different visual dimensions express them and how, when we combine those dimensions for multivalued visualizations, perceptual interactions among them affect the overall utility.

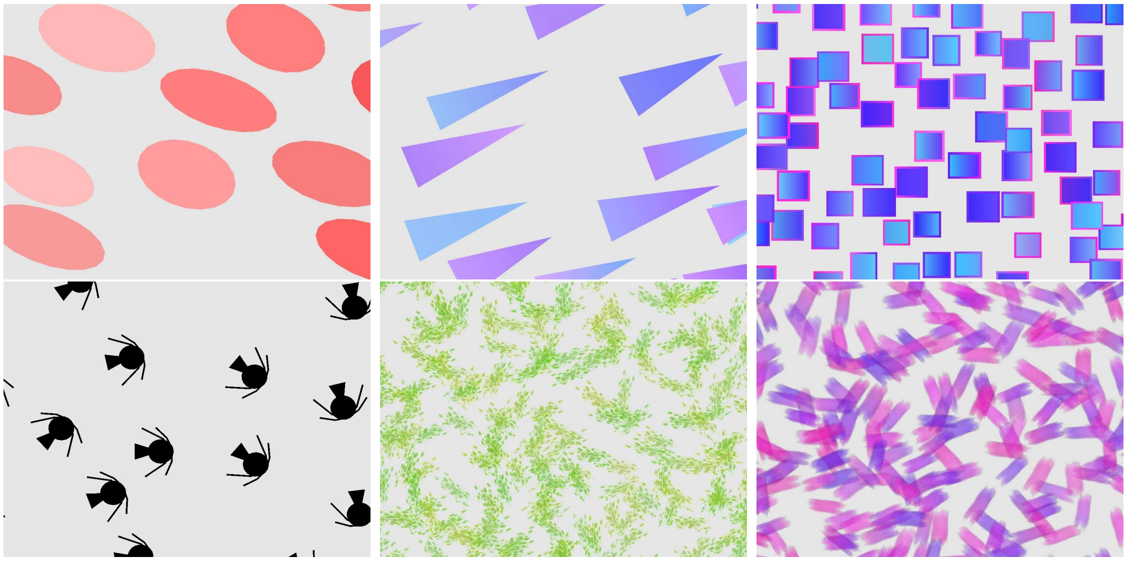


Figure 2.14: Different discrete icons. The top three show each basic primitive (ellipse, triangle and rectangle) provided in the language. The bottom left image shows a composite icon created using a combination of primitives. The last two images show how texture mapping the icons can generate arbitrary shapes.

2.3.1 Definitions and Nomenclature

Developing upon the nomenclature from the previous section, $\mathbb{W}(v)$ provides an evaluation of a visualization method, $v \in \mathbb{V}$. It produces a vector of values, each of which quantitatively characterizes the visualization method with respect to a specific design factor.

$$\mathbb{W}(v) = (w_0(v), w_1(v), w_2(v), ...)$$

In order to generate our utility model we need to specify the type of design factors we will be accepting. We consider four different ones:

- *Spatial Feature Resolution (w_0):* The size of the features a method is able to accurately represent.
- *Data Resolution (w_1):* The number of different values of a data variable a method is able to accurately represent.
- *Saliency (w_2):* How much a method pops-out among the rest of the methods present in a visualization display.
- *Perceptual Interference (w_3):* How much one method affects the accurate reading of another.

We derived these factors from our experience creating scientific visualizations for our collaborators in many disciplines and from our study on designer-critiqued visualization methods [Acevedo et al., 2005]. In that study we utilized a superset of factors that, given that experience, we have narrowed down to these three described here.

2.3.2 Spatial Feature Resolution

This design factor is aimed at giving the visualization user some control over how much information is lost when the icon-based representation is used. Size and spacing between icons are the main dimensions that affect this factor, but our visual system processes visual dimensions differently. Some dimensions could be easier to interpolate perceptually than others, so, for example, larger spacing and smaller icons could provide similar spatial feature resolution results for icon orientation than for icon color saturation using closer together and larger icons.

Important features of a dataset might be lost if the sizes of those features are beyond the capabilities of a certain visualization method. When icons are used for representing continuous data, it is unavoidable that gaps will be present in the final display. These gaps are what make a multilayered visualization possible but, at the same time, they create a challenge. There is a trade off between showing a higher spatial resolution (smaller features) and providing a comparative view with other layers in the visualization (larger holes to see through the layers).

We also limit the range of some visual dimensions so we can visualize smaller features. Size and spacing do not have specific limits for their maximum values. Since icons can potentially be made as large as the full display size, this would clearly limit the available spatial feature resolution. We will provide numeric values for our available ranges when we describe the different experiments.

We measure this factor as the size of the smallest feature a method can represent, measured as cycles per degree. With this measure we try to abstract from the actual physical size of the stimulae used for our experiments, because we roughly know the distance our participants stand from the screen when they perform our experimental tasks (either on paper or on the computer screen).

2.3.3 Data Resolution

This design factor refers to the number of different values that should be visible for a given variable. Although we are dealing with continuous scalar data, scientists often bin their data for easier comprehension.

In our experiments we measure the number of levels a visual dimension has by counting *jnd* (just noticeable difference) units. This means that, although we are not explicitly binning the data for display, we are quantifying how many bins will be perceived. Our result is the limit of the number of bins possible for a certain visual dimension in a given range.

Some dimensions have a limited range, such as saturation, lightness or orientation, while size or spacing do not have specific limits. As we explained for the spatial feature resolution design factor, we limit these dimensions so we can provide higher spatial feature resolution, hence constraining the range and the perceivable bins. This limitation of the range has consequences for the data resolution itself, since there is a limited number of jnd's perceivable in a given range. Measuring data resolution for different ranges will be useful so we can combine visual dimensions using ranges that conflict the least between them.

We measure this factor as the total number of levels (jnd's) visible.

2.3.4 Saliency

With this factor we address the level of importance a data variable must have among the other variables present in the visualization. There might be times when the user wants to highlight a particular variable and keep others as a context. There might also be cases when all variables must be visualized at the same level of importance, leaving the highlighting and backgrounding of some of them for a later stage of exploration.

We measure this factor using direct comparison between methods and asking participants which method dominates the composition. Along with the next factor, we utilize the time it takes participants in our experiments to recognize one dimension in the presence of another. The faster of the two to be recognized is the more salient.

2.3.5 Perceptual Interference

This factor addresses the difficulty of reading a given dimension when others are present. It characterizes how much more difficult a given visual dimension makes the reading of another one. We measure this using the increase or decrease in time that participants take to recognize the particular dimension with respect to a baseline time. More details of this factor will be explained in Section 5.2.

2.3.6 Capturing Designer Critiques

As we mentioned before, the advantage of utilizing expert visual design educators as subjects for some of our experiments is that they can provide reasons for the success or failure of a certain visualization method. During our experiments we will try to capture this information numerically by asking them for estimates on how much a design factor would change if a certain change in one of the visual dimensions is performed. We can indicate this as the derivative of our vector of design factors $\frac{d\mathbb{W}(v)}{dv}$:

$$\frac{d\mathbb{W}(v)}{dv} = (\frac{dw_0(v)}{dv}, \frac{dw_1(v)}{dv}, \frac{dw_2(v)}{dv})$$

where,

$$\frac{dw_j(v)}{dv} = (\frac{\partial w_j(v)}{\partial v_0}, \frac{\partial w_j(v)}{\partial v_1}, ..., \frac{\partial w_j(v)}{\partial v_{11}})$$

Some of these derivatives will be obtained from the analysis of videotapes recorded during the experiments and from interactive sessions where experts modify visualization methods according to different requirements. It is usually easier to let participants explain their reasoning as they perform the experimental tasks than providing them with a form to fill out. This is a more difficult method to obtain quantitative results, but it allows participants not used to numerically critique designs to feel more confident about their decisions.

2.4 Summary of this Chapter

In this chapter we have defined some of the basic components of the visualization process: the visualization problem, the visual dimensions, and the design factors used to characterize our utility model. For each of these we have defined the scope of this dissertation and introduced the nomenclature and mathematical notation we will utilize throughout the thesis. The purpose of this chapter was to characterize the different components of the research and make clear what our assumptions and limitations are.

The next chapter will put our framework in perspective with respect to the state of the art in visualization research, as well as perceptual and design literature.

Chapter 3

Literature Review

Our work is related to three main research areas: visual design, visual perception and data visualization. We will address them separately in this chapter.

In this thesis we are combining all three of these disciplines to facilitate the synthesis of effective visualization methods. Visual design informs our work through techniques for critiquing and works related to image composition and how visual components work together to convey a message. Visual perception, on the other hand, is our main source of low level characterizations for our visual dimensions. The experimental techniques used by perceptual psychologists help us design our own studies, targeted towards more higher level practical applications. Finally, we are trying to contribute to the advancement of the field of scientific visualization, and there are many researchers whose work inspires and complements our own. We summarize here the main sources of knowledge from all three of these areas of research and relate it to our own work.

3.1 Visual Design

Our main point of connection with visual design is the quantification of how the different dimensions that form our basis for visual data communication perform and interact together. There are many authors that have approached the problem of classifying those dimensions and providing guidance for their use, but we are providing a bottom-up approach that numerically quantifies individual performance first and moves on to combinations and their interactions.

Visual designers and artists are trained on how to communicate messages visually. In our case the message is a scientific dataset. We have previously researched, and continue to pursue, the idea of using artistic techniques for scientific visualization [Laidlaw, 2001; Vote et al., 2003; Laidlaw et al., 2004; Kirby et al., 2004; Keefe et al., 2005]. Our experience in this area, and our ongoing collaboration with the Rhode Island School of Design, helped us select the set of visual dimensions that form the means by which we communicate our message. This is, indeed, a very active area of exploration in the field of visualization in general [Watson, 2006].

In looking at art and design critique, there are many principles of design and composition expert critics look for when critiquing a certain piece. Rhythm, repetition, balance, proportion, scale, variety and unity are some examples [Sayre, 1995]. In our case, some of those come defined by the data itself, hence are not subject to evaluation. For example, rhythm and balance in a visualization display are, for the most part, controlled by how the values are distributed across the spatial range of the data. A certain balance could still be judged based on an overall visual balance of the display (whether there are areas that attract attention or not), but it could be argued that finding those areas is precisely the goal of the initial data exploration.

Another aspect of the critique looks at the set of visual elements an art work uses to convey the message intended. There is no agreement among researchers about what is the set of visual dimensions that can be used as a basis to create visualization methods. Ontologies about what visual dimensions are most commonly used served as inspiration for us to come up with a testable set. There are many publications used in art and design schools that deal with specific visual dimensions, but Wallschlaeger and Busic-Snyder in [Wallschlaeger and Busic-Snyder, 1992] provide a very comprehensive classification of the different elements involved in the communication process. Their work spans visual principles for architecture, art and design, and demonstrates the commonalities among those disciplines in this context. Our approach is similar in the sense that we are applying these concepts to an area that makes use of them [Swan et al., 1999], but has not had many researchers studying the formalization of their use. Although Wallschlaeger and Busic-Snyder provide a very clear description of each element (color, shape, texture, etc.), they fail to formalize the interaction among them and the issues arising from their simultaneous use, a key component in our research.

In the classification and analysis of visual dimensions for data representation, one of the first and most cited works outside of the academic literature for art and design is Bertin's *Semiology of Graphics* [Bertin, 1983]. Our approach is very similar to his in that we are trying to characterize the capabilities of each of our visual dimensions individually, and then build up a model of how they perform in combination. He acknowledges that any combination of dimensions is possible but he dedicates very few pages to formalizing the use of those combinations. Our studies are designed to gather knowledge and provide a basis for a formal model for the effective combination of visual dimensions. Our work also presents an opportunity to address a main criticism of Bertin's work, that he lacks experimental results for his factual presentation of visual properties, by providing quantifiable evidence of his theories.

Bertin describes in extensive detail the associative and selective qualities of what he calls *retinal variables*: location, size, value, texture, color (hue), orientation and shape. Compared to our visual dimensions, he combined saturation and hue into a single variable (color), and used texture in a similar way that we use icon spacing. In the case of the location visual dimension, the datasets we will be working with have an inherent spatial component which prevents us from using the location of the visual elements as a carrier of any other information. An example of this are the points on a map: they indicate precise geographical locations, so distortions from the real locations would change the map itself and modify the message (cartograms are an exception to this, in which the visible space changes its meaning [Bertin, 1983, p.120]).

A significant difference with Bertin's work is that we are dealing with continuous datasets instead of discrete ones. In his book, the main examples are centered around map displays, although charts and information visualization displays are also discussed. Maps are very close to the kind of display we are going to be analyzing. In our case, datasets are assumed to be continuous across the spatial range of the data variables. Although maps can contain that type of dataset (e.g interpolated temperature or precipitation readings in meteorological maps), there are many discrete variables (e.g. labor statistics or population maps) that are not usually interpolated. Our research will extend Bertin's results to continuous data.

Many researchers have followed and applied Bertin's work, and map making is one area that has used his work and inspiration extensibly. MacEachren presents an excellent summary of previous research in cartographic visualization [MacEachren and Kraak, 1997]. He expanded Bertin's visual variables to include crispness, resolution, transparency and arrangement. He also divided Bertin's color into hue and saturation for a total of 12 visual variables. Although his classification is better supported by experimental references from map makers and perceptual scientists, we miss some discussion about the specific use of each variable, both individually and in combination (combinations of hue, lightness and saturation are briefly presented). He provides clues towards the generation of rules for map-making but does not go as far as presenting such rules.

Cartographic data visualization is an area that uses similar datasets to the ones we utilize. Techniques and classifications of visual elements for map communication form an important basis for our model and our choice of dimensions for the experiments. Along these lines, MacEachren also describes the three main components of map communication as the data, the graphical elements, and the user. He contends that a characterization of all of them must be obtained to create effective maps. We are constraining our research to a very specific type of data and a predefined set of visual elements that can potentially be used to represent that type of data. The end-user's characterization is represented by our set of design factors.

Finally, one of the most cited works on effective visual design for communicating scientific data has to be Tufte's series of three books on information visualization [Tufte, 1983, 1990, 1997]. In his books he discusses many examples of cases where bad choices in the design of visualization displays caused problems and misinterpretations. He also presents alternatives for how to fix those issues and introduces alternative designs that, for the most part, are effective in conveying the original ideas. It has been said that with Tufte's work bad practice has been uncovered. Even though this is a very valuable step to take in any scientific field, the recognition of existing flaws, Tufte does not take the next step and tries to formalize his views into a coherent comprehensive model. It is very difficult to connect all the extensive advice given in his work and categorize it in a homogeneous way. It seems clear that was not the intent of his work, and that is where our approach fits in. We are trying to build that model and, for that, we are starting the same way Bertin did in his work: from the ground up. The examples Tufte introduces present multiple combinations of visual dimensions and other factors that are very difficult to isolate. We are starting with the study of visual dimensions in isolation and trying to quantify the expressiveness of each of them as they get combined in increasingly complex situations. In summary, Tufte does an excellent job at critiquing and analyzing finalized visualization displays, but fails at exploring how much or how little each component of the displays participates in its success or failure.

3.2 Visual Perception

At the other end of the spectrum are perceptual psychologists and psychophysicists. They are interested in studying how our eyes and brains perceive, process, and store visual information. For that, they utilize very basic visual displays that are designed to trigger very low level responses on the viewer. This helps them isolate how individual pieces of information are processed and build a model for how we perceive the world around us. In our case we are quantifying how visual dimensions are perceived by visual experts and measuring the effect that their combination has in the perception of scientific datasets.

Ware [Ware, 2004] provides an excellent reference towards the understanding of all perceptual processes involved in information comprehension. Color, texture, form, and motion are the main elements discussed in his work, beginning from the physiological elements involved in perceiving each of those, up to a series of recommendations for their use in displaying abstract information. Ware takes a broad approach at information visualization and, although continuous data are discussed in the book, they are not its main focus. He provides a very good introduction to the theory of integral and separable dimensions for visual attributes, but provides little quantifiable evidence for his classification. Our work provides such evidence for the displaying of continuous scalar data and for how separable dimensions combine to form complex visualization displays.

Along these lines, we have found little experimental evidence about the perception of combinations of visual dimensions. Callaghan studied how hue and lightness interact in a texture segregation task [Callaghan, 1984]. She also compared, in pairs, hue with form and line orientation [Callaghan, 1989]. Although she reached interesting conclusions about which variables dominate and when they interfere, her stimulae were limited to two levels of the visual variable being analyzed (e.g. horizontal and vertical for the oriented lines), while the potentially interfering variable was randomized or kept constant. In general, given that our data are continuous, more than two levels of each visual variable will be displayed. We have not found any studies for the interaction of more than two visual variables. Note that, in our displays, all visual dimensions present in our language must be set. Even when only a single data variable is being mapped to a single visual dimension, there are a whole set of other dimensions that can potentially interfere with it.

Our experiments are very much inspired by Carswell and Wickens's work [Carswell and Wickens, 1990] in which they classify different graphical attributes into integral, separable or configurable dimensions depending upon how each attribute's reading is affected by the others, taken pairwise. They found that visual elements can help each other when displaying the same information (redundancy or performance facilitation), or inhibit each other when only one element is changing (filtering interference). They also describe a third type called condensation in which opposite variation of each variable occurs simultaneously. Their experimental displays are based on single icons, looked at in isolation. We are extending their experiments to more complex displays and, for now, limiting our analysis to filtering interference analysis (see Section 5.1.)

Our goal is to find the visual characteristics of different visual dimensions when displaying quantitative information, where visual saliency is the property that makes one data value different from the next. The measurement of saliency or texture contrast thresholds is common in texture segregation studies [Bergen, 1991]. Those studies utilize stimulae with regions where the particular visual element differs in some amount with respect to the surrounding region [Landy and Bergen, 1991]. This is similar to our research in that we are also measuring *jnd*'s for visual dimensions, but our stimulae include overlapping textures. Our texture segregation is a more complex problem, since even a single layer of icons can contain two or more textures, e.g. one for spacing and one for icon lightness. We are interested in measuring the segregation between these textures, so we cannot directly apply *jnd* values from those texture segregation experiments to our model. Also, many stimulae are required to explore the full range of a visual dimension, and even more to include interference analysis with secondary elements. Our studies are designed to evaluate, with fewer iterations, a larger portion of the range for each element.

Moving a little closer to experiments directly applicable to scientific visualization synthesis, most of the literature about perceptually effective data representation is based on experience. Authors define sets of guidelines that, in the absence of visual perception theories [Senay and Ignatius, 1994], follow common practice and established knowledge [Eick, 1995]. In general, these approaches rely on a clear definition of the task a visualization must fulfill, making them difficult to apply in our research. Our exploratory visualization methods are geared towards presenting the data as clearly and without bias as possible for scientists to explore.

In our case, instead of evaluating error or speed for a specific task, we qualify the different methods based on design factors present during exploratory analysis. Also, since our participants are expert visual designers with years of experience in design critiques, we are able to simultaneously evaluate multiple visual dimensions from our language. They are not new elements for our subjects, so we can exploit their expertise in a more efficient way. This methodology allows us not only to understand how our visual dimensions are interpreted by our expert designers, but also how the individual visual dimensions are combined by an observer into coherent percepts [Landy and Movshon, 1991].

Dastani [Dastani, 2002] takes a different approach. His goal is to match the structure of the datasets, relational databases in this case, with the perceptual structure of the visual dimensions used in the visualization display. Again, this is difficult to apply to our scientific datasets, but it is interesting to note that he includes in his discussion the choice of values for the visual dimensions not mapped to any data variables but still present in the displays. We also keep track of these when designers evaluate our visualization methods, and their comments on them, so we can build a complete model of utility. Like Dastani, we try to avoid methods with unwanted visual implicatures by non-mapped visual dimensions.

Our evaluation approach comes closest to the work of Healey. He has studied extensively the application of preattentive processing to visualization [Healey et al., 1993]. Preattentive processing allows detection of visual elements in a display without focusing attention on them. Initially, he focused on experiments comparing hue and orientation [Healey et al., 1996]. Participants in his experiments were asked to perform numerical estimation tasks with varying hue and orientation differences, as well as varying display time. Based on this discriminability experiments, he identified guidelines for color selection [Healey, 1996] that we used for our studies.

He also proposed ViA, a visualization system based on perceptual knowledge [Healey et al., 1999]. The goals of this system are very similar to the ones in our research. He builds, by hand, the perceptual knowledge-base used to suggest a visualization method, while we are gathering that knowledge through subjective evaluations. Finally, Healey used perceptually-based visualization displays to visualize datasets with up to 4 data variables [Healey and Enns, 1999; Healey et al., 2004].

Finally, in the case of multivalued visualizations, our layering of icon-based visualizations takes note from Watanabe's texture laciness studies [Watanabe and Cavanagh, 1996]. Texture laciness defines the phenomenon that occurs when two textured surfaces are overlapped and the top one becomes perceptually transparent so the bottom one can be perceived without interference. He identified icon similarity as the main element affecting laciness. In our case, we want some amount of laciness to be present, itself controlled by the saliency required by the visualization. Spatial feature resolution was not studied by Watanabe as a factor of laciness, but it is something we are including in our studies.

It is important to note that this combination of visual dimensions into perceptually relevant entities has been studied for decades, starting with the Gestalt psychologists and their laws of perception and grouping [Ellis, 1939]. These laws are one of the earliest attempts to qualify how the human visual system recognizes relationships among visual dimensions. We are trying to quantify some of those relationships and apply that knowledge to the effective visualization of scientific data.

3.3 Data Visualization

We titled this section Data Visualization to combine both information and scientific visualization literature. Hanrahan [Hanrahan, 2005] recognizes the artificial and somewhat unclear nature of the separation between information and scientific visualization, but acknowledges that most of the research aimed at the definition and characterization of a space of visualization methods has been done in the information visualization field. Our work is very much inspired and guided by the classification models developed for information visualization.

Many researchers in information visualization have followed and applied the previously mentioned work by Bertin on graphic semiology. Cleveland was one of the first in ordering what he called perceptual tasks (our visual dimensions) based on their accuracy when users read visualization displays [Cleveland and McGill, 1984]. [Mackinlay, 1986] augmented his classification by including expressiveness and effectiveness as the two main measures to evaluate how well a certain dimension performed representing data. Mackinlay went as far as to develop a compositional algebra that would describe how dimensions and tasks were matched to choose a certain method. He also acknowledged the existence of situations where visual dimensions would interfere with each other, throwing off the original classifications, but he did not study those cases. This thesis tries to build a similar classification of visual dimensions and complete the quantification of those perceptual interactions.

In our case the task is exploratory, but many researchers have approached the modeling of the space of visualizations based on a taxonomy of tasks [Casner, 1991; Springmeyer et al., 1992; Shneiderman, 1996]. Their results vary and are appropriate for the tasks represented, but they all fall short in the study of perceptual interactions when multiple variables need to be represented. Furthermore, even in the cases where they approached the issue, the type of relational or nominal data they dealt with makes the extrapolation of their findings into our domain difficult.

Other examples of modeling the space of visualizations are [Robertson, 1991; Miceli, 1992; Laper, 1995; Lange et al., 1995; Card and Mackinlay, 1997; Nowell, 1997; Andrienko and Andrienko, 1999; Chi, 2000; Nagappan, 2001; Salisbury, 2001; Jankun-Kelly, 2003]. Most of these works have the commonality of being rule-based: they rely on building a set of rules that will guide the visualization synthesis process. In our case we do not build rules before hand and rely on experimental evidence to build our knowledge of expressiveness and effectiveness for each of our visual dimensions.

Indeed, Johnson, in his list of top scientific visualization problems [Johnson, 2004], recognized the quantification of the effectiveness of visualization methods as one of the major research areas in this field. He also included perceptual issues, multi-field visualization and theory of visualization, all areas that we are addressing in this dissertation. Van Wijk, in two important papers for the field [van Wijk, 2005, 2006], also discusses extensively about the importance of quantifying the value and effectiveness of visualization methods. Close collaboration with the end users to define the goals of the visualizations and determine a basis for effectiveness, is the key to successfully quantify that value. [Tory and Moller, 2004] also talked about some of the new challenges the field of visualization has to tackle and she concentrated on human factors. She included perceptual measurement of effectiveness and an argument for a formal modeling of these to really anchor the field and move forward. We believe we are modestly addressing those issues in this thesis and we will make a valuable contribution to the visualization field.

[Weigle et al., 2000; Taylor II, 2002; Bokinsky, 2003] are works closely related to ours in the sense that they dealt with scalar fields in 2D and tried to develop new techniques to display them effectively. They heavily relied on experimental evaluations to validate their techniques, but they did not explore what are the fundamental expressive characteristics of the visual dimensions that form their visualizations. They built up their techniques based on previous work and their own experience, obtaining valid results and developing a layering technique that let them present effectively multiple variables simultaneously.

Finally, our work has also some similarity to the research on multiple surface visualization [Interrante et al., 1997; House and Ware, 2002; House et al., 2006]. These approaches tried to visualize a 3D object by placing glyphs on its surface. Since the point of view of the object was expected to change, these glyphs did not usually change based on the characteristics of the surface. In our case, the overall textures created by mapping a scalar field to visual dimensions could create the illusion of a surface and might actually represent one, but the icons are used to indicate the very values being explored.

Chapter 4

Using Visual Design Experts to Evaluate Scientific Visualizations

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 to support data validation and decision making.

Visualization is used in any data-intensive domain: data mining, meteorology, geography, transportation sciences, environmental studies, uncertainty analysis, and evolutionary biology are some example application domains. In these fields, a common problem for visualization experts is: given a large set of multivalued data and hypotheses a scientist would like to address, what visualizations best represent the data? And how do we best evaluate these visualizations? Furthermore, does a good method for evaluation provide sufficient information to improve the visualization methods?

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 4.1).

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. They are trained

	Pros	Cons
Non-experts	• Easy access • Unbiased opinion • Help on minor or overlooked issues	• Possible subconscious influence of external factors • No help on fixes/improvements
Visualization Experts	• Easy access • Evaluation of implementation issues • Knowledge of alternative methods	• Too close to development
Domain Experts	• Specific knowledge of tasks and goals • Extrapolate to tasks not being tested	• Infrequent access • Expensive/busy • Biased: too much experience • 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

Table 4.1: Pros and cons analysis for each type of subject in scientific visualization evaluations.

to optimize visual resources in a visual design problem; the ultimate goal of our research is to quantify and model this optimization process.

To achieve this we have performed a series of experiments to evaluate how effective these experts are at evaluating these types of displays.

During the first experiment [Jackson et al., 2003; Acevedo et al., 2007b], artists and visual designers graded the vector visualization methods from a previous study [Laidlaw et al., 2005] on the basis of their subjective estimates of user performance and also verbally critiqued each method's effectiveness.

The results from this experiment encouraged us to develop a methodology to evaluate scientific visualization methods using expert visual designers. This led to our second study where expert illustration educators evaluated multiple 2D scalar visualization methods [Acevedo et al., 2005].

The purpose of these two studies is to learn how visual designers evaluate scientific visualizations against certain design and scientific goals. 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 the design of our visualizations. This, in turn, should speed up and improve its results.

4.1 Critique-based Evaluation of 2D Vector Visualization Methods

Our hypothesis was that designers would rank the methods similarly to the objective task-performance measures in [Laidlaw et al., 2005]. We also hoped that the critiques would help us understand why methods work well by identifying which visual dimensions within each method worked best for the given tasks. Our results are consistent with our hypothesis.

4.1.1 Methodology

In order to evaluate the efficacy of our designer critiques, we modeled our study on a previous quantitative user study [Laidlaw et al., 2005] 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 [Laidlaw et al., 2005] , users were asked to evaluate the merits of the six visualization methods shown in Fig. 4.1:

- GRID: icons on a regular grid.

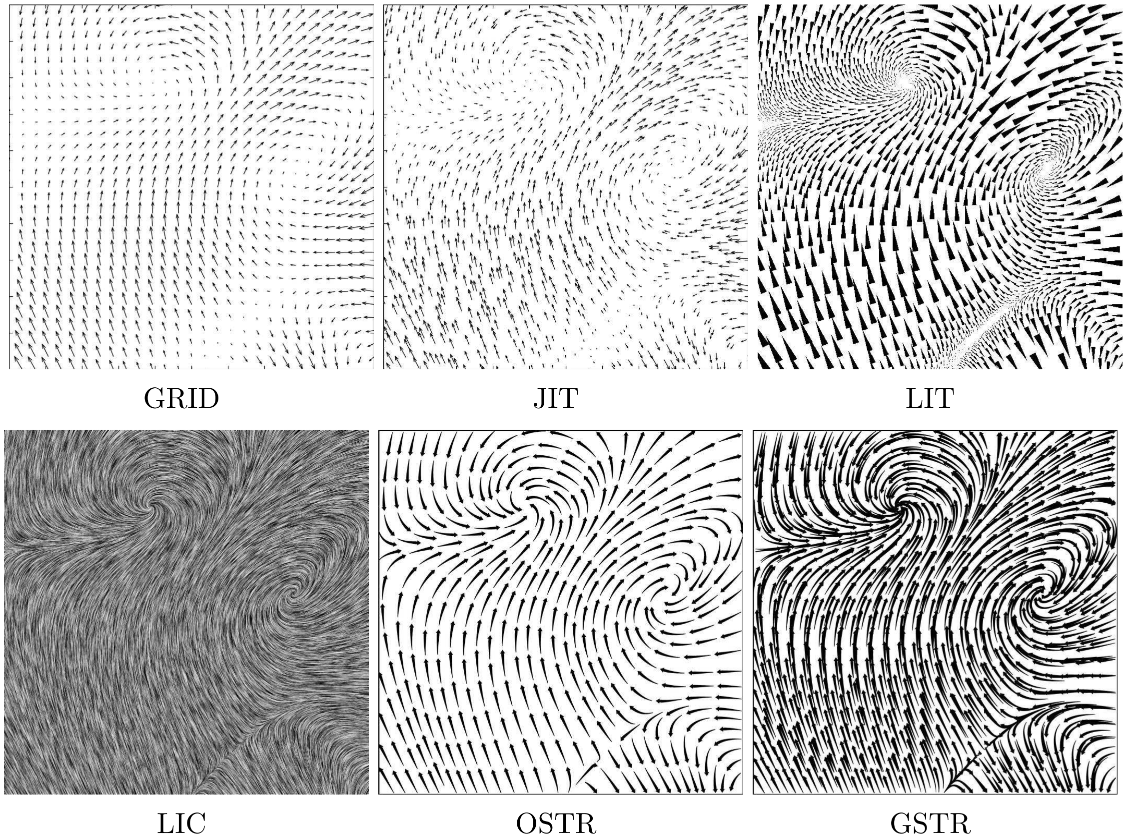


Figure 4.1: The same vector field as visualized by the six visualization methods critiqued by the designers.

- JIT: icons on a jittered grid [Dippé and Wold, 1985].
- LIT: icons using one layer of a visualization method that borrows concepts from oil painting [Kirby et al., 1999b].
- LIC: line-integral convolution [Cabral and Leedom, 1993].
- OSTR: image-guided streamlines (integral curves) [Turk and Banks, 1996].
- GSTR: streamlines seeded on a regular grid [Turk and Banks, 1996].

With these methods, users were asked to perform three tasks designed to mimic generic tasks fluid-flow experts would use to investigate a flow field (Figure 4.2):

- *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.

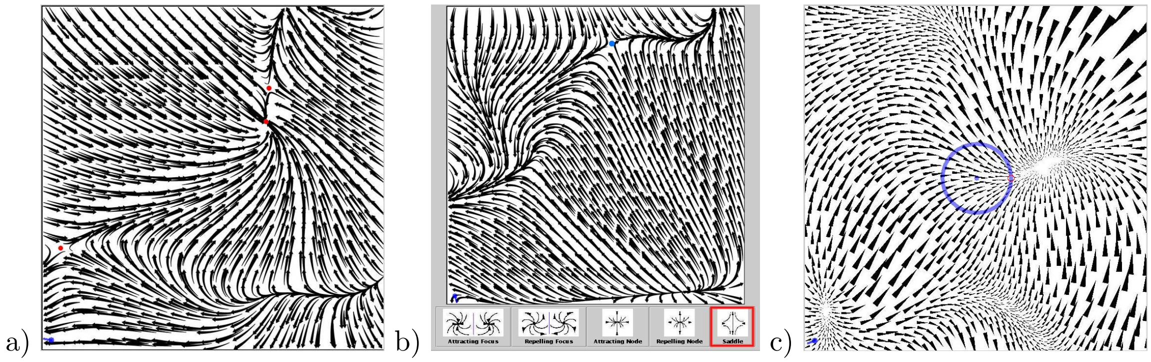


Figure 4.2: Sample stimuli for the three experimental tasks. The solutions for each, marked in red, were provided to participants during the subjective critiques. Their goal was to judge how accurately and quickly a real user of the visualizations would perform these tasks for each method. (a) Counting task with three critical points visible, (b) Type ID task with a saddle-type point (marked in blue), and (c) Advection task with a small red circle indicating the location to where a particle in the center of the large blue circle would advect.

Seventeen users were run through the 90-minute computer-controlled experiment [Laidlaw et al., 2005]: 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 [Laidlaw et al., 2005].

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 4.3 shows one of our visual designers critiquing the 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 users to perform all three tasks quickly and accurately.

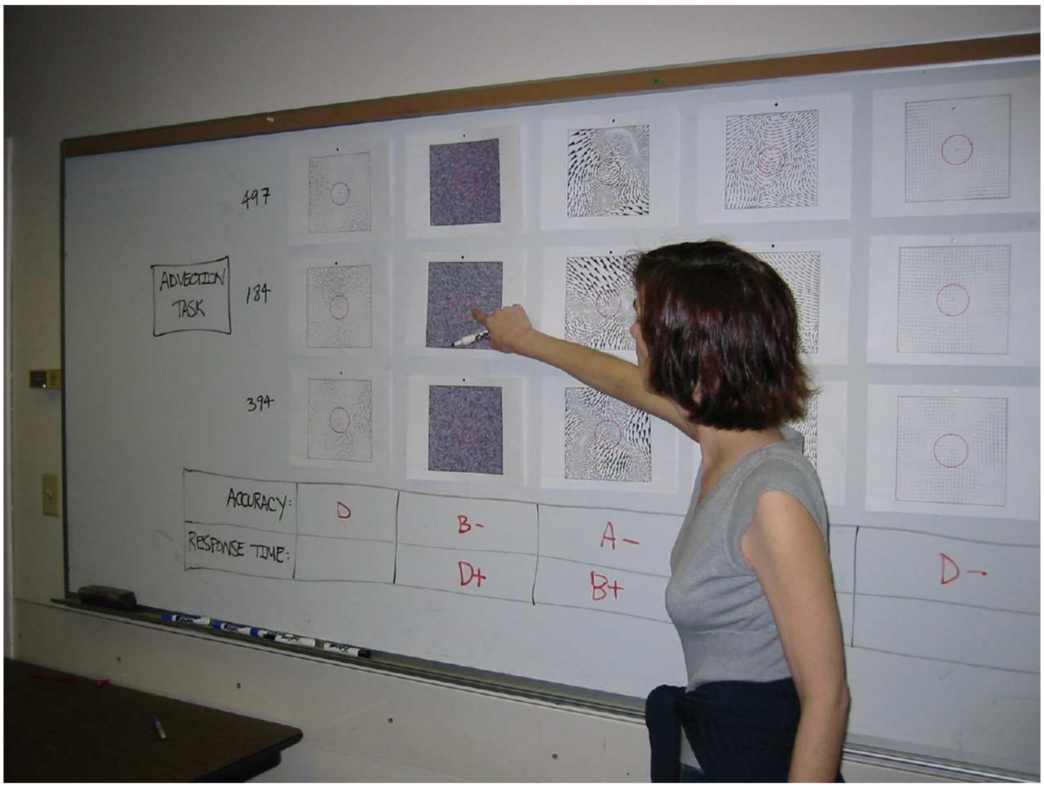


Figure 4.3: During the study, designers rated the different methods subjectively, based on accuracy and time to perform the task. They could also appraise how the visual dimensions used in each method would affect their performance.

	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

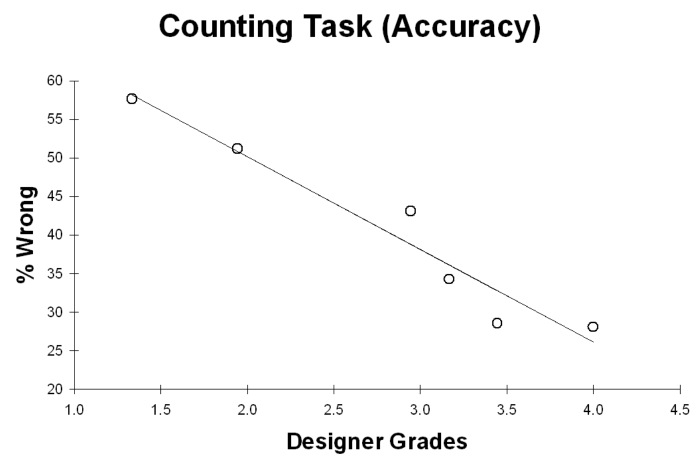
Table 4.2: Linear regression results between designer grades and numerical results from [Laidlaw et al., 2005].

4.1.2 Numerical Results and Discussion

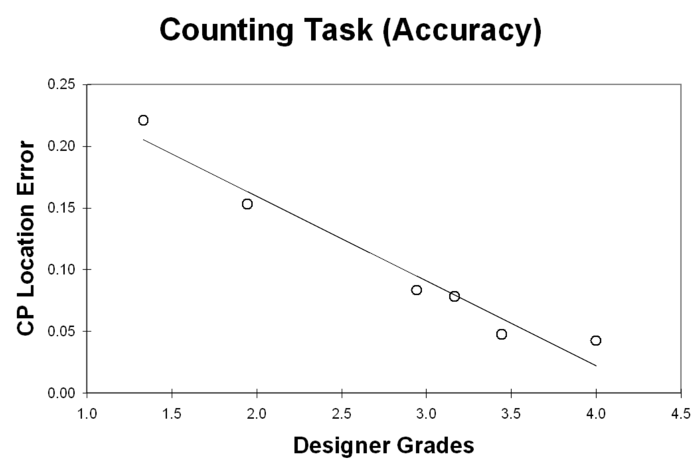
We posed two hypotheses at the onset of this study: first, that designer ratings would be similar to the quantitative performance measures for each task in the previous study [Laidlaw et al., 2005], and second, that the designer critiques would provide additional insight into the merits of each method and how to improve them. Table 4.2 summarizes the linear regression results between the designer grades from the current study and the numerical results from [Laidlaw et al., 2005]

We look first at the critical-point-counting task. This is the most difficult task because no visual cues are present to guide user performance (as there are for the advection and critical point identification tasks [Laidlaw et al., 2005]): during this task, users see only the flow field. Note that, 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.

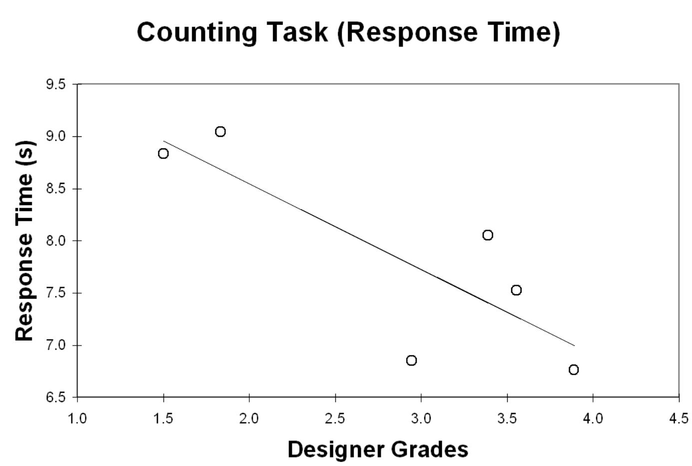
Figure 4.4(a) shows the regression analysis for mean percentage correct in counting the critical points, and also the mean designer grades. It is clear that the designers' pattern of performance matches the quantitatively collected performance measure for this task very well ($R^2 = 0.941, F = 63.9, p = 0.001$). Figure 4.4(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. 4.4(c) shows the regression analysis for the mean time



a)

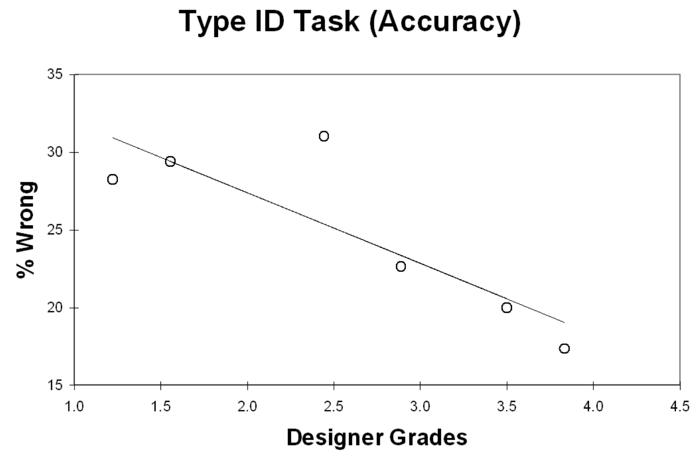


b)

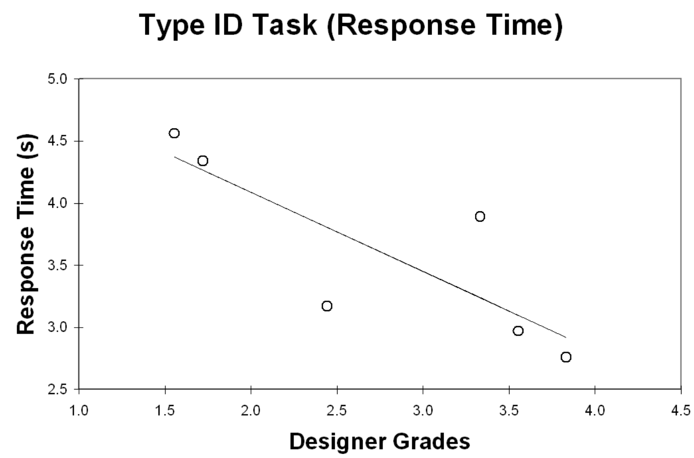


c)

Figure 4.4: Regression analyses for the critical-point-counting task, with plots for counting accuracy (a), location accuracy (b), and response time (c).



a)



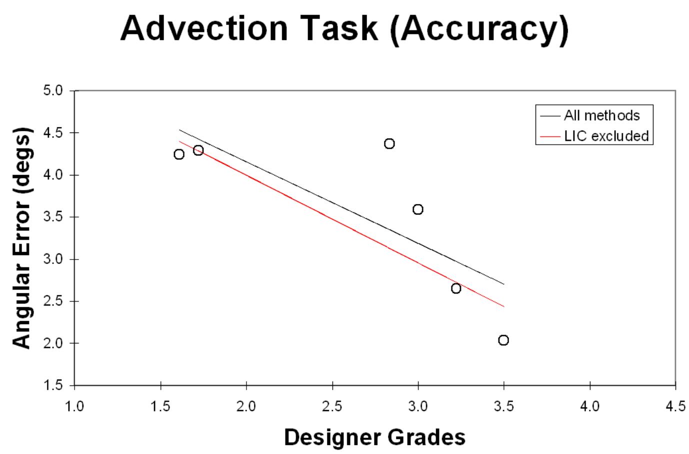
b)

Figure 4.5: Regression analyses for the critical-point-type identification task, with plots for type accuracy (a), and response time (b).

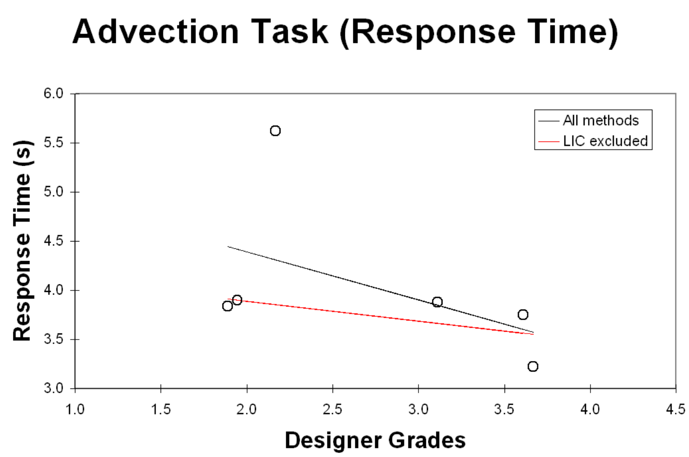
to complete the critical-point-location task and the mean designer grades. Once again, the designers' pattern of performance matches the quantitatively collected performance measure ($R^2 = 0.676, F = 8.3, p = 0.045$).

As can be seen from the graphs in Fig. 4.5, 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 ($R^2 = 0.249, F = 1.3, p = 0.314$).



a)



b)

Figure 4.6: Regression analyses for the advection task, with plots for accuracy (a), and response time (b).

This last discrepancy can be explained by looking at one visualization method in particular: line integral convolution (LIC). As seen in Fig. 4.1, LIC shows no information about the flow 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 flow 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 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 4.6 shows the linear regression plots for this task with and without LIC.

4.1.3 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 flow 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 4.7 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 flow structure and thus aid in identifying particle advection,

Figure 4.7: 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 flow lines help in the advection task, and white marks indicate direction of the flow. Critical points are clearly marked by large white dots, and critical point type is indicated by the surrounding arrows.

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 “understand the overall flow structure and features”.

4.1.4 General Discussion

These results validate our initial hypotheses, but they leave some open questions. Even though the tasks are interesting scientifically, designers seemed to find it very easy to evaluate them. The concepts in their critiques were very basic, even though all six of our methods were state of the art for 2D flow visualization. The specificity of our tasks did not tax our participants’ design expertise . As can be seen from Fig. 4.7, 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 [Laidlaw et al., 2005] 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 [Nielsen, 1992]. In particular, we believe that the participants' visual design expertise is key to providing the types of comments they did during our experiment.

Finally, since the ratings obtained from designers are largely qualitative and do not provide the numeric values necessary to design a visualization method, combining objective and subjective experiments using designers will lead to better, more directly usable results, confirming the hypothesis from Tory and Möller [Tory and Moller, 2005]. 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.

4.2 Evaluation of 2D Scalar Visualization Methods by Illustration Educators

Following the results from the previous study, we set off to quantify how much each individual visual dimension that forms our visualization methods participates in their success or failure. We would have liked to obtain this information from the previous study, but the specificity of the tasks, and the non-uniformity, in terms of the visual dimensions utilized, of the six flow visualization methods used, prevented our participants from consistently commenting about the exploratory use of the methods or that of their intrinsic visual dimensions.

In collaboration with educators from the Illustration Department at the Rhode Island School of Design (RISD), we defined a space of eleven visual dimensions(see Figure 4.8). These eleven dimensions were considered sufficiently expressive and representative for an initial exploration of the space of visualization methods. In order to evaluate the individual expressive power of each dimension, we used them to visualize single-valued continuous scalar datasets in 2D.

We created a framework for evaluating these visual dimensions through feedback from expert visual designers and art educators. Our framework mimics the art education process, in which art educators impart artistic and visual design knowledge to their students through critiques of the students' work. We established a set of design factors that characterize the *exploratory* nature of our visualizations without focusing on any particular *explanatory* task,

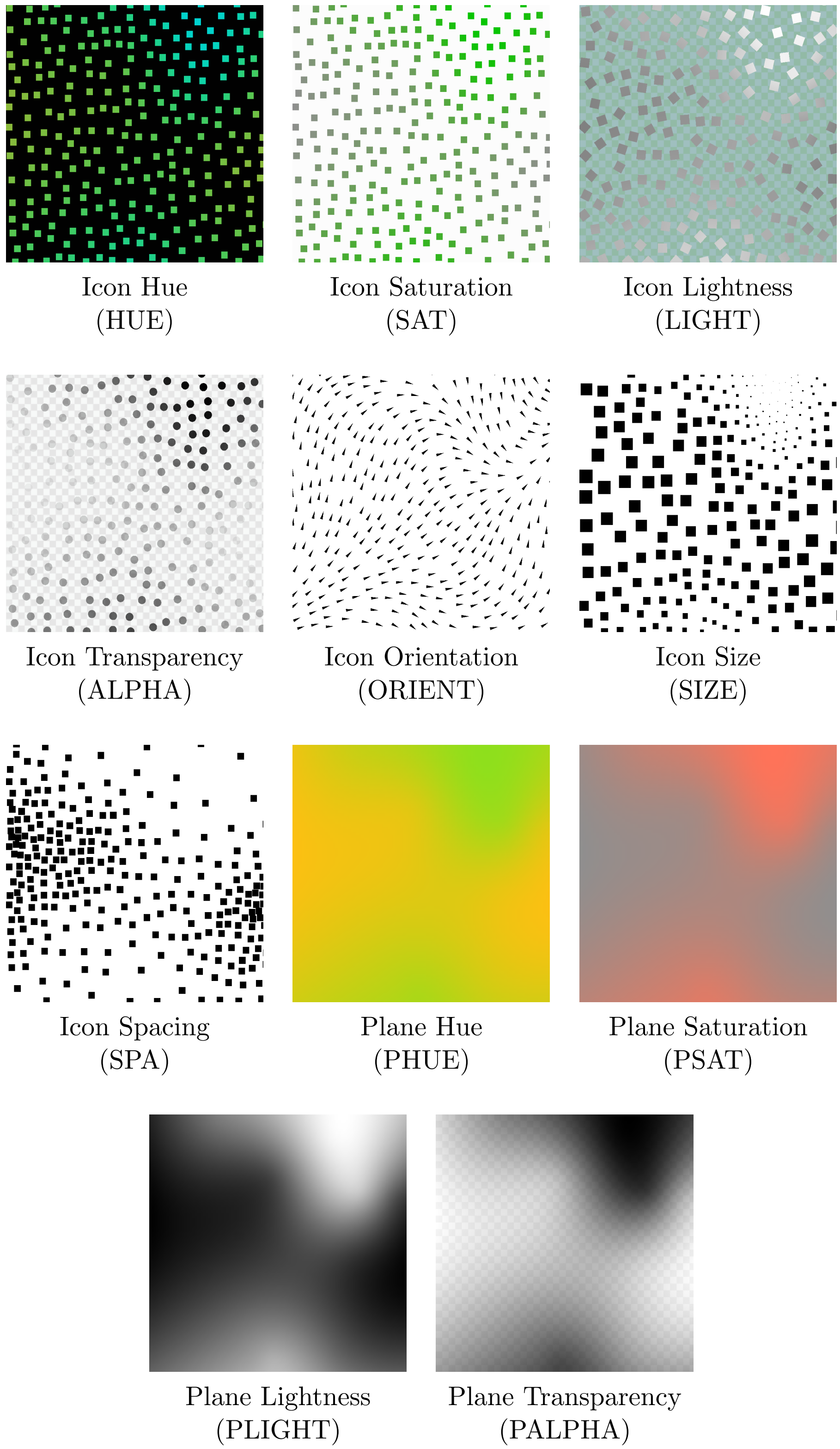


Figure 4.8: Visual dimensions. In this experiment we asked expert visual designers to characterize the utility of each of these dimensions individually.

such as search for extrema or analyzing data gradients.

This study represents our initial step towards our final goal, which is to create a mathematical model of the knowledge collected from design experts and use that model to find a solution for a data visualization problem.

4.2.1 Methodology

In this particular study we had five participants, all expert educators from the Illustration Department at RISD, evaluating 33 different visualization methods. The number of methods comes from 11 visual dimensions and 3 different parameterizations for each one (see Table 4.3 for details on each method's parameterizations).

Following what we did for the previous study, we tried to eliminate the effect of the dataset from the analysis by showing 4 different single-valued datasets, shown in Figure 4.9.

In summary, we created a total of $33 \times 4 = 132$ images that we printed and placed on the wall for our subjects to critique. The setup is shown in Figure 4.10.

The design factors we defined provide information about the quality of the data presented and the capability of a visualization method to work in combination with other methods. For this particular experiment we had a different set of factors than the ones introduced before. These factors are:

- *Data Resolution (DR):* the number of different levels of a data variable that can be distinguished by a viewer.

- *Spatial Feature Resolution (SFR):* the minimum spatial feature size that can be reliably represented with a method, expressed as a percentage of the image width.

- *Visual Linearity (LI):* the perceptual linearity of the mapping from data value to visual dimension; this factor is measured by asking participants to indicate the locations where they see the values of 0, 0.25, 0.5, 0.75, and 1.0 along the image for a linear dataset visualization.

- *Visual Bandwidth (VB):* the percentage of a method that can be occluded when combined with other methods but still remain readable. This design factor is aimed at estimating how different visual dimensions will perform in multilayer methods.

- *Dominance (DO):* the forcefulness or *punchiness* of the data mapping. This indicates how much a method would dominate the composition when combined with other methods, measured as a value from 0 to 10.

- *Time to Read (TR):* the time it takes an average user to comprehend the data, measured in seconds.

VISUALIZATION METHODS		VISUAL DIMENSIONS: HUE	SAT	LIGHT	ALPHA	ORIENT	SIZE	SPA	PHUE	PSAT	PLIGHT	PALPHA	
HUE	1	0, 0.125	1	0.75	0	0	20	17	0	0	0	0	
	2	0.5, 0.25	1	0.85	0	0	20	17	0	0	0	0	←
	3	0.875, 0.625	0.9	0.9	0	0	20	17	0	0	0	0	
SAT	1	0.33	1, 0	0.8	0	0	20	17	0	0	1	0	←
	2	0.75	0.5, 0	0.65	0	0	20	17	0	0	0	0	
	3	0.05	1, 0.5	0.7	0	0	20	17	0	0	0.71	0	
LIGHT	1	0	0	1, 0	0	0	20	17	0.5	0.5	0.7	0.48	
	2	0	0	0.5, 0	0	0	20	17	0.5	0.5	0.7	0.48	
	3	0	0	1, 0.5	0	0	20	17	0.5	0.5	0.7	0.48	←
ALPHA	1	0	0	0	0, 1	0	24	15	0	0	1	0.1	←
	2	0	0	0	0, 0.5	0	24	15	0	0	1	0.1	
	3	0	0	0	0.5, 1	0	24	15	0	0	1	0.1	
ORIENT	1	0	0	0	0	45, 90	20	15	0	0	1	0	
	2	0	0	0	0	0, 90	20	15	0	0	1	0	
	3	0	0	0	0	-90, 90	24	15	0	0	1	0	←
SIZE	1	0	0	0	0	0	1, 40	17	0	0	1	0	
	2	0	0	0	0	0	1, 20	17	0	0	1	0	
	3	0	0	0	0	0	20, 40	17	0	0	1	0	←
SPA	1	0	0	0	0	0	20	50, 0	0	0	1	0	←
	2	0	0	0	0	0	20	25, 0	0	0	1	0	
	3	0	0	0	0	0	20	50, 25	0	0	1	0	
PHUE	1	0	0	0	1	0	0	0	0.25, 0.125	0.9	0.9	0	←
	2	0	0	0	1	0	0	0	1, 0.5	0.9	0.9	0	
	3	0	0	0	1	0	0	0	0.625, 0.875	0.9	0.7	0	
PSAT	1	0	0	0	1	0	0	0	1	1, 0	0.8	0	←
	2	0	0	0	1	0	0	0	0.5	0.5, 0	0.9	0	
	3	0	0	0	1	0	0	0	0.875	1, 0.5	0.7	0	
PLIGHT	1	0	0	0	1	0	0	0	0	0	1, 0	0	←
	2	0	0	0	1	0	0	0	0	0	0.5, 0	0	
	3	0	0	0	1	0	0	0	0	0	1, 0.5	0	
PALPHA	1	0	0	0	1	0	0	0	0	0	0	0, 1	←
	2	0	0	0	1	0	0	0	0	0	0	0, 0.5	
	3	0	0	0	1	0	0	0	0	0	0	0.5, 1	

Table 4.3: The exact parameterizations presented to our subjects for all 11 visual dimensions. Orientation is measured in degrees from the horizontal direction, size in pixels, and spacing also in pixels (for 800×800 images). The pairs of values in the yellow cells indicate the minimum and maximum values of the mapping ranges for each method. Mappings are linear in all cases. The red arrows to the right indicate the particular methods shown in Fig. 2.10.

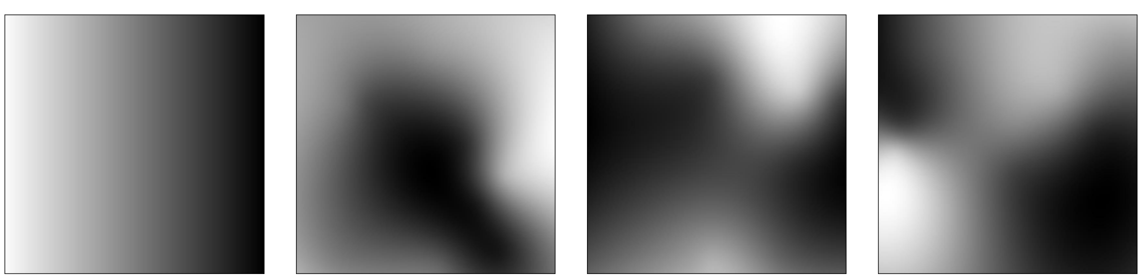


Figure 4.9: During the study, participants were presented with multiple visualization methods representing these four single-valued datasets. The first one is a linear dataset, while the rest are general, continuous and smooth changing height fields.

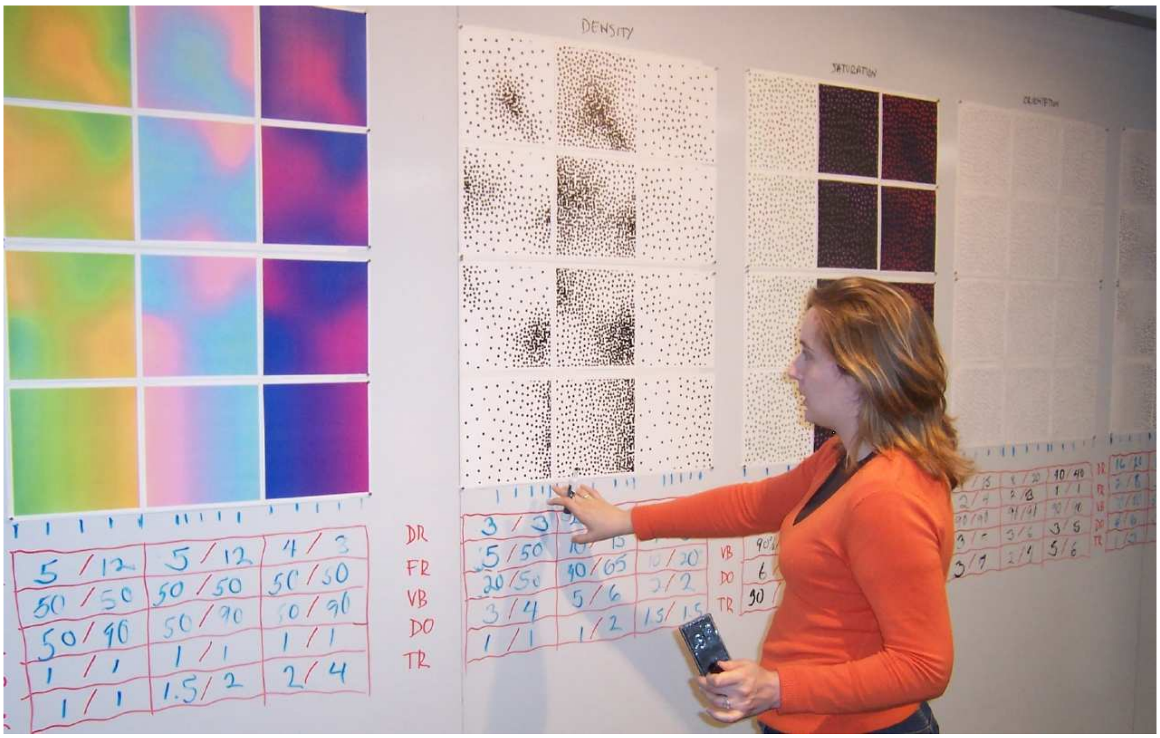


Figure 4.10: A participant in our study critiques visualizations of single-valued 2D datasets. Illustration educators were shown a total of 132 visualizations corresponding to 3 parameterizations of 11 visual dimensions and 4 different datasets. For each parameterization, they evaluate all 6 of our design factors (bottom of the image). Here, one of our participants comments on the reasons for her ratings.

Note that dominance intends to represent what we called saliency, while visual bandwidth and time to read are both factors included in our definition of perceptual interference. This experiment preceded the definition of our perceptual experiments, hence this set of factors is a superset of the ones we finally used for our utility models. In later chapters we will compare the results using these 6 factors with the results using our final set of 4.

As explained before in Chapter 3, Bertin [Bertin, 1983] developed a similar classification for his "retinal properties" (size, value, texture, color, orientation, and shape) according to their level of organization (whether they could be used to represent quantitative, qualitative, or ordered information) and the number of steps they could take (our data resolution factor). Our design factors introduce new measures, like linearity, and also capture some composition characteristics, like visual bandwidth and dominance. Our data resolution and spatial feature resolution factors capture the fact that we are targeting quantitative datasets.

For this experiment we created a novel experimental methodology for capturing quantitative knowledge from visual design experts. This is a clear improvement with respect to the previous study, since we are trying to provide designers a way to convey their critiques through the use of our design factors. We videotaped the sessions, which last approximately 3 hours, and we asked participants to provide in-depth explanations of the numerical ratings and their thought process.

For their training, participants were introduced to all methods and design factors the day before their critique, when they were given the instructions for the experiment. A webpage [1] was available for them to review all visual dimensions we were interested in, as well as the goal of the experiment and the introduction to the different factors. They were encouraged to familiarize themselves with all parts of the experiment and write down any questions they had. They were instructed not to actually perform the design factor estimations. Before and during the actual evaluation sessions, participants were allowed to ask any questions or make any comments about the study.

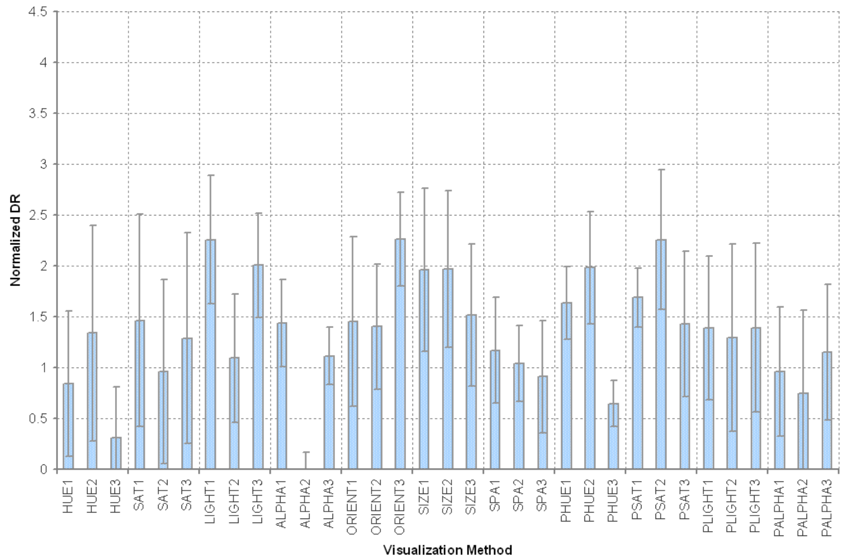
4.2.2 Results

The results of this study allowed us to characterize the expressiveness of individual visual dimensions when visualizing single-variable scalar datasets for exploratory visualization.

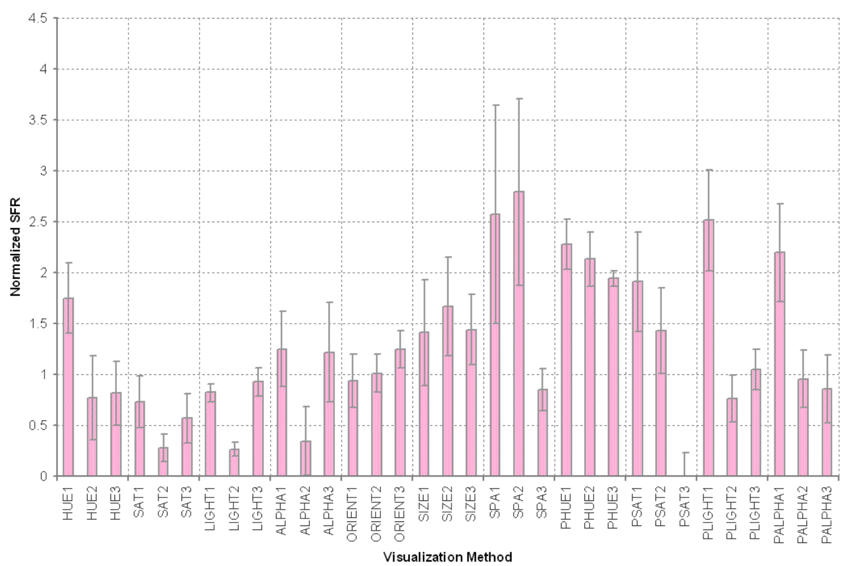
In general, the between-subject scores had a very high variance. While we believe the relative balance between methods was maintained due to having all stimuli presented at once and allowing score modification, each participant had his or her own range of values for each design factor. Hence, we normalized all the scores for each participant individually.

We obtained, then, a relative measure for each design factor. Figures 4.11, 4.12, and 4.13

[1] http://www.cs.brown.edu/people/daf/study

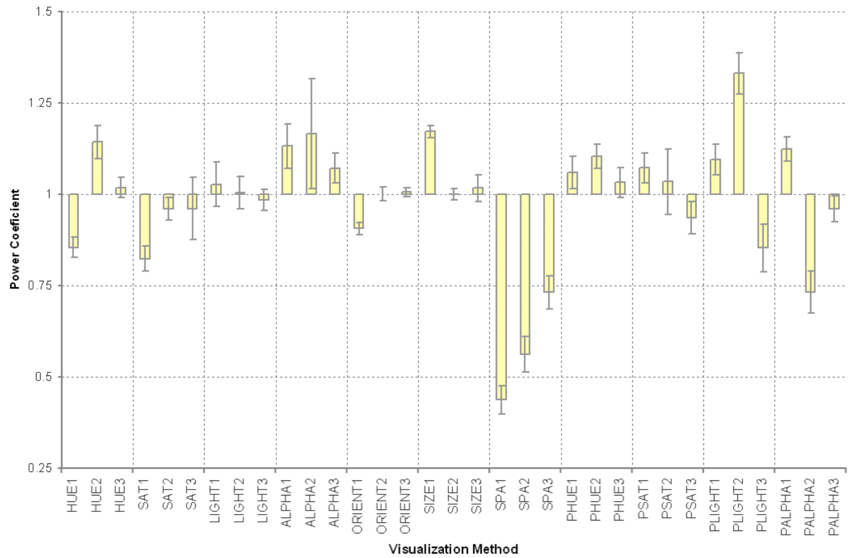


(a) (DR) Data Resolution: higher = more levels

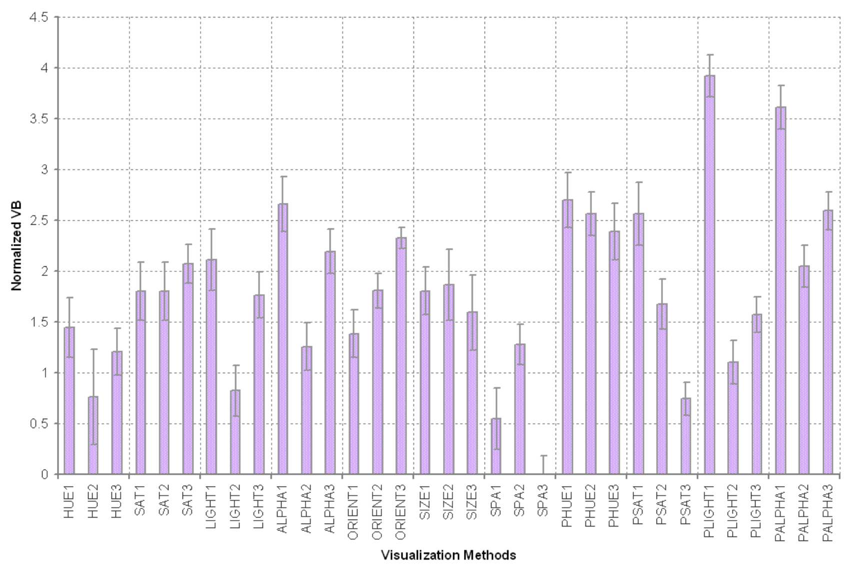


(b) (SFR) Spatial Feature Resolution: higher = higher resolution

Figure 4.11: Normalized mean results for all methods (translated so $min = 0$). Error bars indicate standard error.

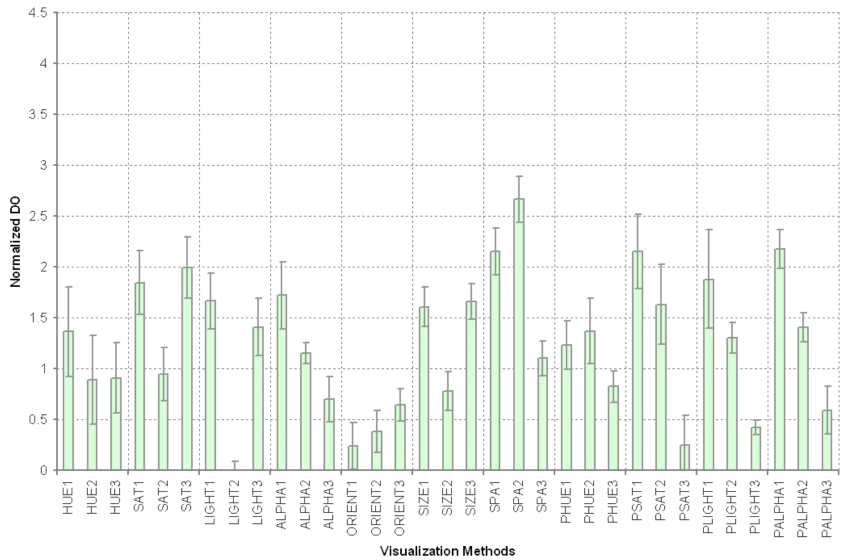


(a) (LI) Linearity results: Fitted power coefficient, α, for $y = x^\alpha$

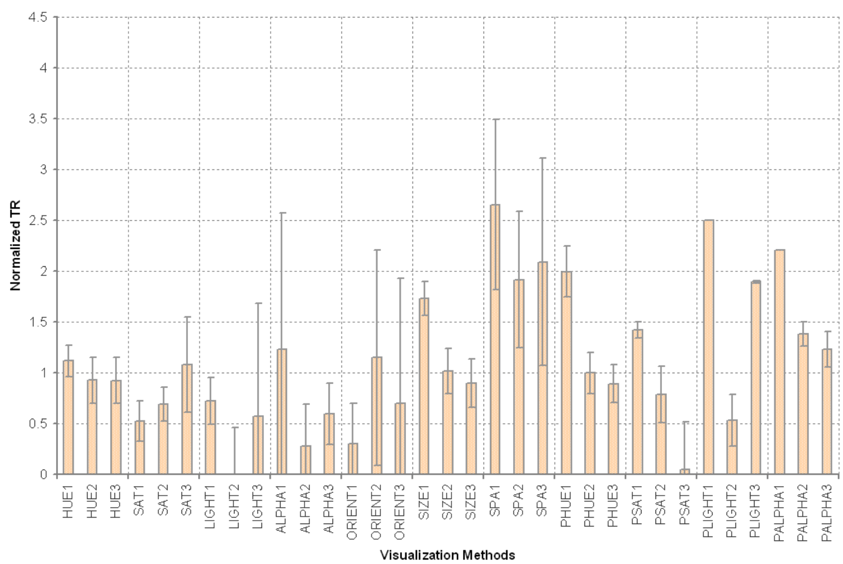


(b) (VB) Visual Bandwidth: higher = can be occluded more

Figure 4.12: Normalized mean results for all methods (translated so $min = 0$). Linearity results are not normalized nor translated. Error bars indicate standard error.

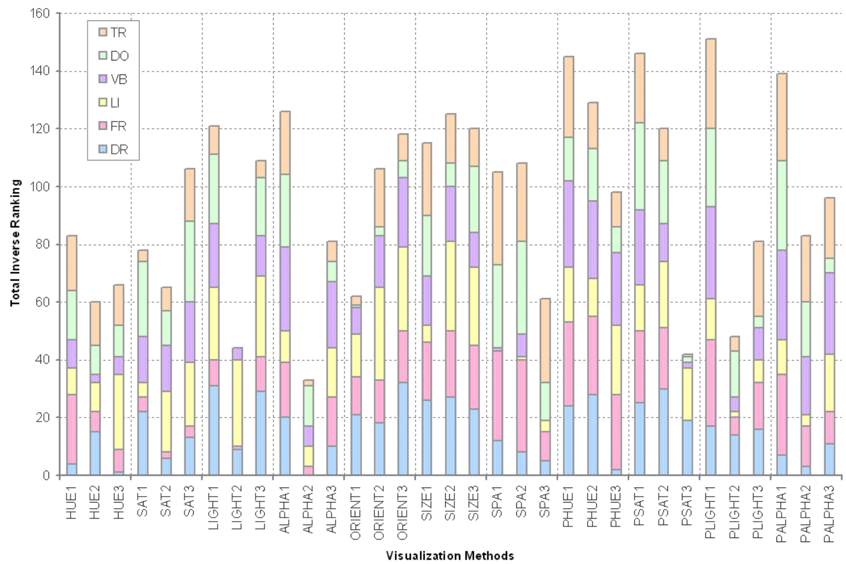


(a) (DO) Dominance: higher = more dominant

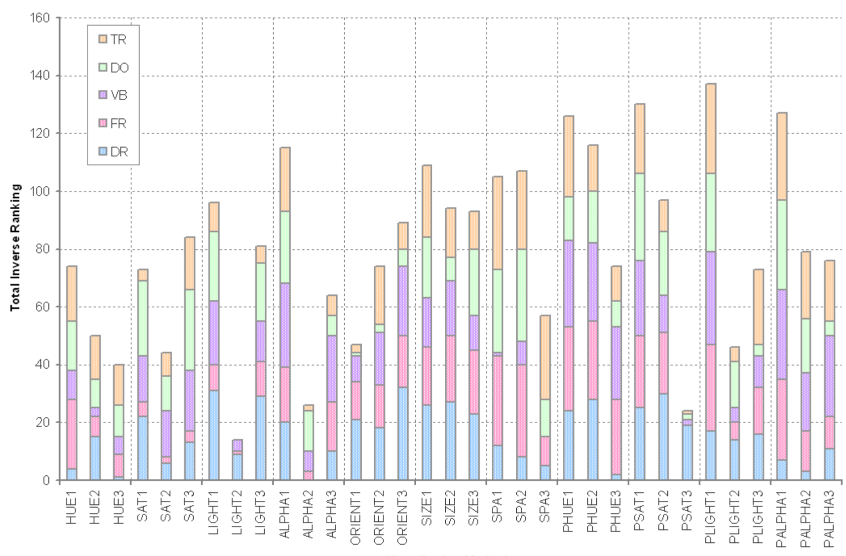


(b) (TR) Time to Read: higher = faster to read

Figure 4.13: Normalized mean results for all dimensions (translated so $min = 0$). Error bars indicate standard error.



(a) Overall inverse rank results. $Max.Possible = 33 \times 6 = 198$



(b) Overall inverse rank results without Linearity. $Max.Possible = 33 \times 5 = 165$

Figure 4.14: Stacked-columns summary visualization of inverse rankings for all design factors (a) (best is 33, worst is 1, for each factor). Since we can *fix* linearity by remapping the data to the visual dimensions using the power function, we can eliminate that factor from the overall ranking score. We show this in (b).

show the normalized results for all factors. It is easy to spot how methods that work very well for some factors really struggle in others. These graphs show the relative strength of each method and their relative differences, since absolute values cannot be obtained back from these normalized scores.

With these results, we obtained an inverse ranking (33 is best, 1 is worst) based on the mean normalized scores. Figure 4.14 shows an overall summary of rankings. We used a stacked-columns graph where the larger the colored area for each factor the higher value that dimension obtained. We can see how, if the goal is to obtain the best possible visualization method with respect to all factors (with equal weights), we can just pick the top method (Plane Lightness 1 in this case, i.e. full grayscale representation, not surprisingly.) If weights are required, then an optimization process can be easily set up to solve for a set of methods that fulfill the goals best.

4.2.3 Discussion

In general, all our participants felt this is an important and very interesting line of research. None of them were used to making numeric judgments about tasks they perform from experience. They understood our goal of trying to extract that expert knowledge but we felt that, in this case, we over-taxed their expertise by making them concentrate on numbers. They enjoyed the freedom of asking questions and explaining their decisions, but the ultimate need for a numerical estimation created problems.

The study setup was also well received. They are used to comparative critiquing in class and in art and design critiques, where comparative critiquing is an established technique [Feldman, 1994]. Being able to do these comparisons helped self-balance the evaluations within each participant's results and made them much more confident of their own evaluations. For this reason we had very good intra-rater reliability for each participant, but it was difficult to get good inter-rater reliability.

This comparative critique required printing the stimuli, which meant all hue, saturation, lightness, and transparency images were very much influenced by the quality of the printing. We believe, as did our expert subjects, that extra care must be taken in creating this images in the future, but the advantage of having them all visible at once outweighed the use of a high resolution monitor (potentially the tool used by end-users) to view images one by one.

The main problem with our results was the high variance we obtained among the five experts. Our hypothesis is that this could be due to their slightly different interpretations of some of the factors. Also, although our use of expert educators, as opposed to students, was based on the expectation that they would be able to distance themselves from their personal taste, this is very difficult to achieve. As you can see by the standard error bars displayed, even with the normalized scores our participants showed a relatively high degree of variance in their responses. We decided that a ranking of the means would be the appropriate way to report our results. We believe the high standard errors point more to a methodological problem (correct explanation of design factors, calibrated printing, etc.) rather than true significant differences in the methods' performances.

The rankings shown in Fig. 4.14 (b) provide a good overall impression about the utility of the different methods. In general, full range methods (the #1 parameterization for most dimensions, as shown is Table 4.3) obtained the best scores and, in particular, color plane methods performed better than icon-based methods. Among the icon-based cases, size and spacing have a very consistent set of high rankings for most design factors. This was expected, given that these methods do not depend on good printer calibration or illumination conditions in the room. Orientation would be the other method in this category, but it performed quite poorly due to the overwhelming sense of flow it conveys, one that interferes with the scalar reading of the individually oriented icons.

As mentioned before, one advantage of using design experts to do the evaluations is that they can pinpoint reasons why a method does not work. They commented about the neutrality of the shapes used; circular icons would be better than square ones, since square ones create very visible orthogonal lines that mislead the viewer. Also, they mentioned the very prominent effects of the negative space; we must take extra care in doing a very even random placement of the icons, since holes are very quickly detected and incorrectly interpreted as data features.

Commenting on the appropriateness of our design factors, one of our participants noted that the choice of an effective visualization method will be affected by what the data actually is, e.g., visualizing temperatures is not the same as looking at wind speed or altitude data. In our case, we want to apply our resulting design knowledge to any type of scalar data, so we are considering the use of a seventh factor called *intuitive association*. This would measure whether there are any associative readings of a method that might interfere with the desired numerical reading and should be avoided.

It is important to note that some of the factors we chose, in particular visual bandwidth and dominance, are aimed at multivalued visualization situations. We decided to include these factors to introduce our participants further to what real cases would be. We reminded them that single-valued scalar field visualizations were used for the purposes of this particular study only, but our final goal was to understand how these dimensions work in combination for complex multivalued datasets. The same way we did for the previous vector field visualization study, in the next chapter we will correlate these subjective scores with more quantitative perceptually-based experimental results from our next set of experiments.

Finally the length of the study was deemed excessive by all participants. Some of them took as long as 6 hours to complete all evaluations. Even when we moved the training session to the day before the critiques, and provided them with an online resource for reference, forcing them to do a continuous critique session probably impacted the results due to fatigue: participants had to provide a total of 165 numerical estimates, the visual linearity ratings, and explain out-loud their thought process throughout the experiment.

4.3 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 cost of training designers in the scientific goals of the visualization methods is more than recovered by their ultimate contributions.

The experiments presented here lay out ways of using expert visual designers as evaluators of mappings between the data and the visual dimensions that form our visualization methods.

The main result of these two studies 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.

This thesis continues this line of research by combining perceptual experiments and subjective critiques. This strategy should yield the best of both worlds while allowing a complete, if always hypothetical, analysis of the high-dimensional space of exploratory scientific visualization methods.

Chapter 5

Experimental Evaluation of Perceptual Interactions

To solve some of the problems revealed by the previous studies, we turned to a more psychophysically oriented design for our experiments. Relying on simpler, more perceptual tasks would make it easier to get reliable quantitative data, but it would lead us away from the benefits of using subjective critiques from expert visual designers.

Our goal with these studies is to analyze the quality of the data we can obtain and how we can build upon it to accomplish our overall goal of modeling the utility of visualization methods. To that end we reduced the number of dimensions to a manageable size so we could explore the subspace they form as exhaustively as possible. Accomplishing our goal with this few visual dimensions would clarify the methodology to incorporate other dimensions.

In this chapter we describe two experiments. In the first experiment we quantified how perceptual interactions among visual dimensions affect effective data exploration. During the experiment, participants quantified three different design factors for each method: the spatial resolution it could represent, the number of data values it could display at each point, and the degree to which it is visually linear. Icon saturation, icon lightness, icon size, and icon spacing are the four dimensions we evaluated. We measured filtering interference for all three design factors, which characterizes how different levels of a visual dimension held constant affect the evaluation of a data-coupled dimension.

The second experiment goes a little further in exploring multivalued datasets. In this study we perform an experimental quantification of how factors such as icon size, spacing, layer order and color affect the relative saliency and interference among five different dimensions: saturation, lightness, orientation, size, and spacing. We included orientation in this experiment to compare our results with existing visualization literature [Healey et al., 2004] using orientation as a method to visualize continuous scalar fields. These two design factors serve to represent what dominance, visual bandwidth and time to read represented in our second study with expert visual designers.

Our novel experimental methodology in both studies allows us to generalize this perceptual information, gathered using ad-hoc artificial datasets, onto quantitative rules for visualizing real scientific datasets. From both experiments we were able to fit mathematical models that describe the relationships among dimensions and their expressiveness characteristics.

5.1 Subjective Quantification of Perceptual Interactions

During this experiment, participants quantified three different design factors for each method: the spatial resolution it could represent, the number of data values it could display at each point, and the degree to which it is visually linear. These form a subset of the design factors we evaluated before, but represent the basic factors for single-variable visualization methods.

We devised three different tasks that participants would have to perform in order to provide us with their evaluations. In the previous experiments we asked participants to judge how easy or hard it would be for a real user of a visualization to perform a certain task. This is meaningful from the point of view of an illustration expert, but we wanted to measure the actual perceptual capabilities of our visualization methods. To accomplish this we must test the participant's perceptual system and extract our characterization based on their results on those tests. These indirect perceptual tasks should make the experiment easier on the participants but still powerful and generalizable from our perspective.

The main novelty of our approach is the quantification and modeling of how the different visual dimensions interact with each other. This interaction can be explored at many levels [Carswell and Wickens, 1990] but the present study is limited to filtering interference among the various elements. This type of interaction is based on the visual dimensions being mapped to data one at a time, while the rest remain constant across the visualization. We chose to limit our experiment to four visual dimensions: icon saturation, icon lightness, icon size and icon spacing. We realize that these choices greatly constrain our otherwise exponentially large exploration space but, with just these elements involved, we are able to generate an extensive set of examples for our experimental participants to evaluate.

5.1.1 Methodology

Table 5.1 shows the values we chose for each of the four visual dimensions involved. Size indicates the diameter of the circular icons, while spacing indicates the distance between two icons. We utilize a Poisson disk distribution to randomly place icons across the image.

	values			
saturation	0	0.33	0.66	1
lightness	0	0.33	0.66	1
size (pixels)	2	5	7	10
spacing (pixels)	0	3	6	10

Table 5.1: Values used for each of our visual dimensions.

We experimentally chose the upper limits for size and spacing so we could explore methods with reasonable spatial feature capabilities.

With these parameters we defined six possible value ranges, pairs (b_i, e_i), for each visual dimension: $(0.00, 0.33)$, $(0.00, 0.66)$, $(0.00, 1.00)$, $(0.33, 0.66)$, $(0.33, 1.00)$, and $(0.66, 1.00)$. Here, and in the rest of this chapter, we will use these normalized values to refer to the ranges of all our dimensions. That way, for size, a range of $(0.00, 1.00)$ will correspond to a range of $(2, 10)$ in pixels. Similarly for spacing ranges. For icon saturation and lightness methods we combined these six ranges with all possible combinations for the other two dimensions, creating a total of 96 visualization methods. For icon size and spacing methods we kept icon saturation and lightness constant at 1.00, so 24 combinations (6×4) were defined for each of those two dimensions. Note that even constraining our experiment to a small number of elements, and only four possible values per element, the number of combinations is quite large: 240 different visualization methods.

Data Resolution Identification Task

For this task we asked participants to evaluate how many different levels of the data variable a method is able to represent. Figure 5.1 explains how we created the stimulae for this task. The task participants were asked to perform was to define in what region of the image they perceived a sine-wave pattern. Since they were told the pattern would be more pronounced at the top left corner of the image, they just needed to place 3 marks to approximately bound the region where they perceived the pattern.

Using a vertical sine-wave pattern with constant spatial frequency, λ, across the image (Figure 5.1(a)), we linearly decrease the amplitude a from left to right (Figure 5.1(b)). While the amplitude values remain constant vertically across the image, we linearly move the zero value of the sine-wave from $a/2$ at the top to $1 - a/2$ at the bottom. Figure 5.1(c) shows the final appearance of such a dataset using grayscale.

To create these datasets we had two extra variables to fix, the initial amplitude a and the frequency λ. We tested several values for these variables and decided to evaluate

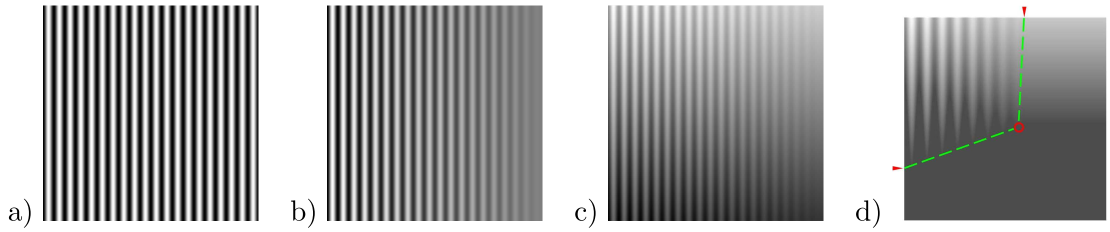


Figure 5.1: Process for creating the stimulae for the data resolution identification task. (a) Shows a vertical sine-wave dataset. (b) Shows the same dataset with amplitude values a linearly decreasing from left to right. (c) Shows the final appearance of the datasets used for this task, where we also linearly move the zero value of the sine-wave from $a/2$ at the top of the image to $1 - a/2$ at the bottom. (d) Shows how participants would mark the area where they perceive the sine-wave pattern.

eight different combinations using two amplitude values ($a = \{0.2, 0.6\}$) and four spatial frequency values ($\lambda = \{160, 80, 40, 20\}$ measured in cycles across the image width, $c/width$). To avoid multiplying by 8 the full set of 240 methods, we decided to use only combinations that utilize the full range of the data-mapped visual element, i.e. $b_i = 0.00$ and $e_i = 1.00$. During the analysis of the results we can still describe the data resolution capabilities for any subrange. Figure 5.2 shows examples of images used for this task for each of the four visual dimensions.

To obtain actual data resolution values we developed the following process. The marks placed by a participant delimit a region on the image where the pattern is visible (see Figure 5.1(d)) . The right and bottom boundaries indicate lines where the difference between the extremes values of the sine-wave are last perceived by the subject, i.e. the *just noticeable differences* (*jnd*) boundary. The basic idea to obtain data resolution values is to follow these boundary lines, starting from the top mark, jumping from one level to the next a distance equal to the amplitude at each point. With this process we will also obtain actual values, in the range $(0.00, 1.00)$, for each level identified.

Since there are two different initial amplitude values used, the results will overlap after a certain distance. The total data resolution of a visualization method will be given by the number of levels obtained for the $a = 0.2$ dataset, plus the number of levels for the $a = 0.6$ datasets with values greater than the maximum level obtained from the $a = 0.2$ dataset.

Spatial Feature Resolution Identification Task

For this task we asked participants to evaluate the size of the smallest spatial feature a method can represent. Our approach for this task was to indirectly ask the question by exploring the limits of each participant's visual perception. In this case our datasets were vertical sine-wave patterns that maintain constant amplitude a but linearly change their

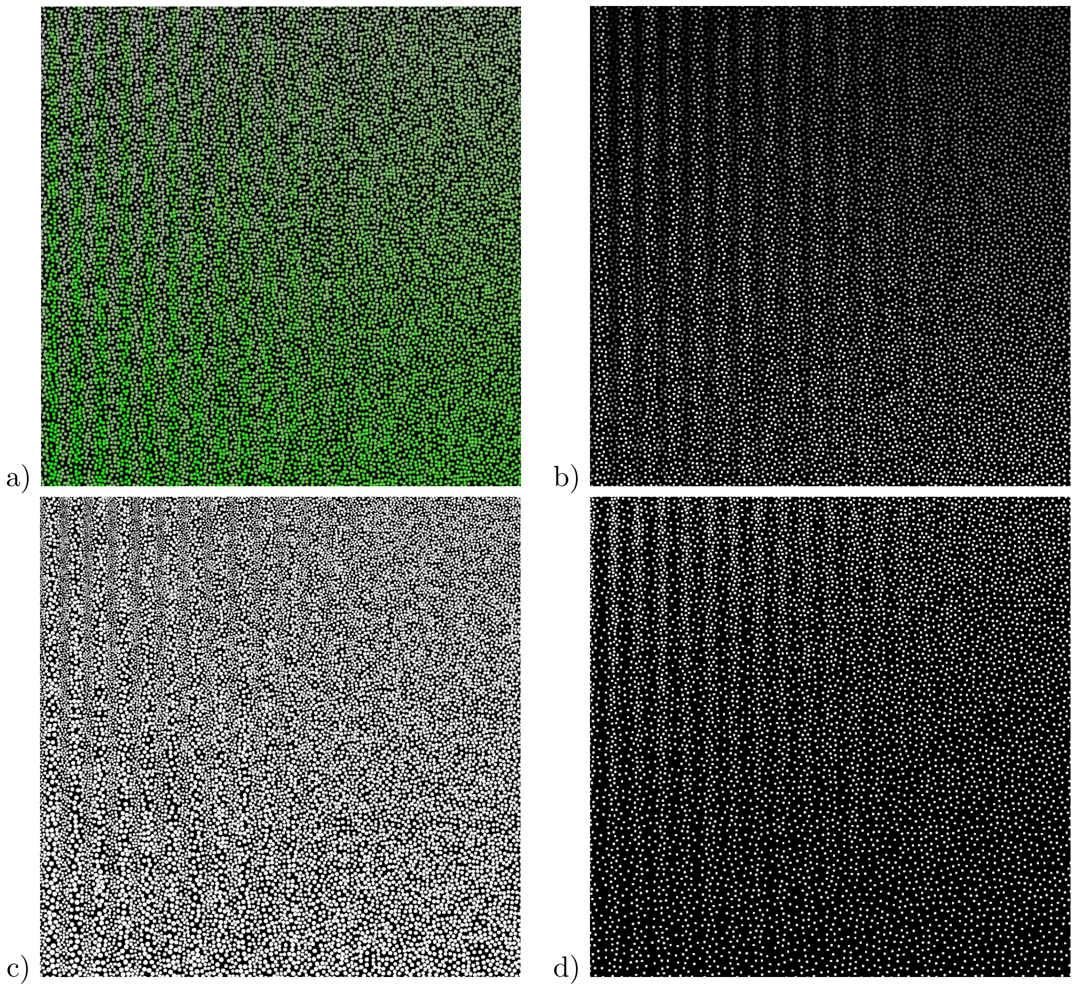


Figure 5.2: Examples of various stimulae used for the data resolution task using saturation (a), lightness (b), size (c), and spacing (d). All of them with $\lambda = 20$ $c/width$ and $a = 0.6$.

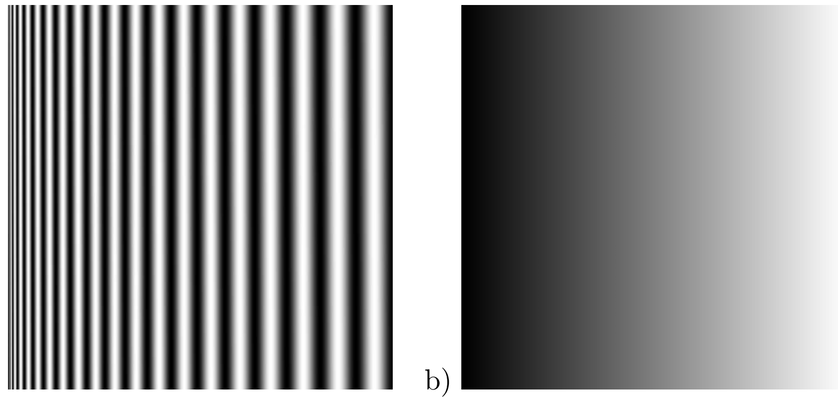
a) b)


Figure 5.3: (a) Shows an example stimuli for the spatial feature resolution identification task dataset, with wavelength λ linearly decreasing from right to left. (b) Shows the stimuli for the visual linearity perception task.

frequency, λ, from left to right across the image. Figure 5.3 (a) shows an example of this dataset using lightness values from 0.00 to 1.00 (i.e. $a = 1.0$).

By asking participants to place a mark when they stop perceiving the sine-wave pattern, we obtained our minimum feature size measurement. $\lambda/2$ at that point would be the minimum spatial feature a method can represent. For this task we use all 240 visualization methods mentioned before. The amplitude for each display is indicated by the range (b_i, e_i). Figure 5.4 shows examples of images used for this task for each of the four visual dimensions.

Visual Linearity Perception Task

In this task participants were shown visualizations of a linear dataset that progressed from a value of 0 on the left of the image to a value of 1 on the right edge (see Figure 5.3 (b)). They were told that 0 and 1 were at the very edge of the images and were asked to place five marks for the values of $0.0, 0.25, 0.50, 0.75, 1.0$. The two extremes would indicate regions where they do not perceive a change in the visualization's border regions. A visually linear method would maintain a constant ratio between data changes and visualization changes.

Experimental Setup

We ran a fully randomized within-subjects study where 6 computer science graduate students performed all three tasks on the computer screen for icon lightness, size and spacing. The data resolution task for icon saturation was ran separately, since we decided to include saturation as one of the modeled elements after the experiment with the first three was already done. Also, the results obtained for the other two tasks for the rest of the visual dimensions indicated we only required this task to be performed.

The full study (for icon lightness, size and spacing) consisted of 9 separate sections (3

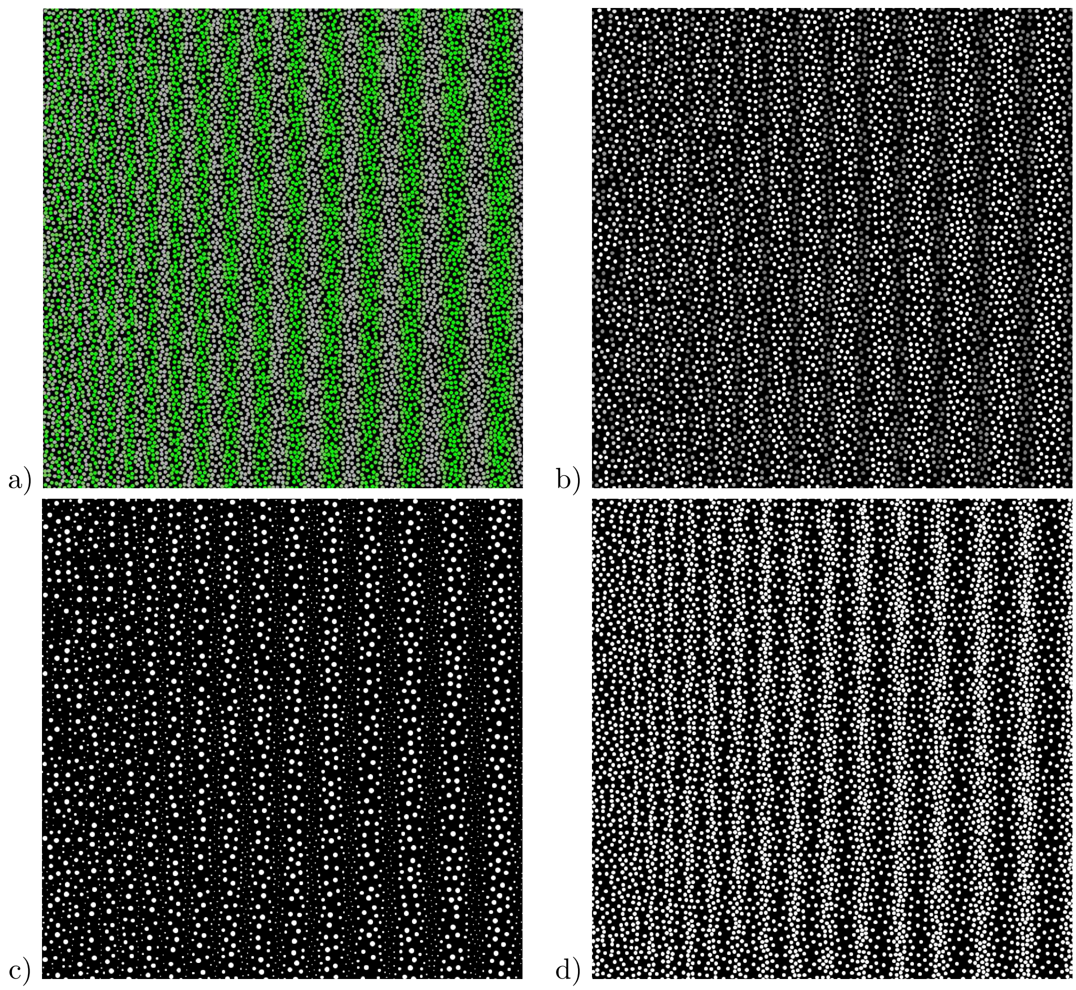


Figure 5.4: Examples of various stimulae used for the spatial feature resolution task using saturation (a), lightness (b), size (c), and spacing (d). All of them with $(0.00, 1.00)$ range.

tasks × 3 elements) with a training subsection and a real trial subsection within each one. Response time was recorded for the real trials. There was no time limit during any part of the study, although participants were instructed to proceed as quickly and accurately as they could. Participants took an average of one hour and forty minutes to complete the whole study and were paid for their participation. They were given written instructions before each task. Stimulae for all tasks consisted of images of size 900×900 pixels displayed one by one on an LCD display.

5.1.2 Results and Discussion

Our results successfully characterize the capabilities of each visual dimension, using a variety of value ranges, in combination with potentially interfering dimensions.

Our spatial feature resolution results confirmed our expectations. The minimum size feature perceived by our participants can be modeled as:

$$w_0(v_i) = s_i + p_i,$$

where s_i is the size of the icons and p_i their spacing, all of them in the same units (e.g. pixels or % of the image width.) Note that this factor can be evaluated at any point across the image. To make this model independent of the dimension used to represent the data or its range, we must consider the cases when v_i corresponds to size or spacing. To include those cases, the model we will use will be:

$$w_0(v_i) = \frac{min(s_i) + max(s_i) + min(p_i) + max(p_i)}{2}$$

This model represents the size of a half cycle from the original sine-wave dataset. Intuitively, using our icon-based representation, any changes in the data occurring in a space smaller than the size of the icons plus the spacing around them will not be captured. In portions of the dataset where this is the case (the left side of our experimental stimuli), the data seems random and loses all structure.

During the visual linearity task, all participants reported difficulty completing the task. They easily placed the marks for the extreme values but they could not judge, in general, the 25% intermediate differences we were asking them to indicate, especially for icon lightness methods. Participants also complained about possible inaccurate gamma calibration of the monitors used. It is still worth noting that practically all methods, for all three visual dimensions, exhibited clear constant-value areas for the extreme values, sometimes as large as 30% of the image width.

The data resolution task yielded the more interesting results of all tasks. Given the

$$a = n_0\hat{\lambda} + n_1 e^{\frac{1}{\hat{\lambda}+1}} + n_2 s_i + n_3 s_i^2 + n_4 e^{\frac{1}{s_i+1}} + n_5 p_i + n_6 p_i^2 + n_7 e^{\frac{1}{p_i+1}}$$
$$b = m_0\hat{\lambda} + m_1 e^{\frac{1}{\hat{\lambda}+1}} + m_2 s_i + m_3 s_i^2 + m_4 e^{\frac{1}{s_i+1}} + m_5 p_i + m_6 p_i^2 + m_7 e^{\frac{1}{p_i+1}}$$

		λ	$e^{1/(\lambda+1)}$	s	s^2	$e^{1/(s+1)}$	p	p^2	$e^{1/(p+1)}$	R^2	F	p
SAT	a	0.252	2.215	-0.448	0.024	-1.919	--	--	0.189	R^2=0.65	F(6,58)= 18.1	p<0.0001
	b	0.765	7.231	-1.335	0.072	-4.889	--	--	--	R^2=0.83	F(5,59)= 57.1	p<0.0001
LIGHT	a	--	4.521	-1.015	0.048	--	-0.505	0.026	--	R^2=0.83	F(5,59)= 54.7	p<0.0001
	b	--	11.594	-2.340	0.104	--	-0.608	--	--	R^2=0.90	F(4,60)=128.7	p<0.0001
SIZE	a	--	1.281				-0.289	0.016	-0.406	R^2=0.92	F(4,12)= 35.5	p<0.0001
	b	--	4.437				-0.684	0.032	-0.738	R^2=0.99	F(4,12)=201.1	p<0.0001
SPA	a	--	1.996	-0.702	0.040	--				R^2=0.79	F(3,13)= 16.3	p=0.0001
	b	--	4.925	-0.674	--	--				R^2=0.88	F(2,14)= 52.7	p<0.0001

Table 5.2: Regression results for all four methods tested. Each model corresponds to one of the coefficients for the logarithmic model $w_1(v_i) = a_i Ln(x_i) + b_i$. $\hat{\lambda}$ is the spatial frequency of the data measured in $c/degree$: $\hat{\lambda} = \lambda/45$. The grayed out areas mark the size and spacing parameters that do not apply for SPA and SIZE methods respectively.

results obtained we fit a logarithmic model to represent the total number of *jnd's* users would perceive at each point in the range of the visual dimension.

$$w_1(v_i) = a_i Ln(x_i) + b_i$$

This follows Weber's Law of perception which states that the relationship between a stimulus and its perception is logarithmic. In other words, the threshold necessary to detect a change in a particular dimension increases logarithmically as the dimension's value increases.

In order to fit this model to our data, we modeled both coefficients as a function of our independent variables (size, spacing, and the frequency of the sine-wave λ.) We perform a linear regression to obtain the coefficients of a model as follows:

$$a_i = n_0\lambda + n_1 e^{\frac{1}{\lambda+1}} + n_2 s_i + n_3 s_i^2 + n_4 e^{\frac{1}{s_i+1}} + n_5 p_i + n_6 p_i^2 + n_7 e^{\frac{1}{p_i+1}},$$

where we include both quadratic terms, which would capture a maximum or minimum in the ranges studied, and inverse terms, which we transformed to an exponential factor due to $p_i = 0$ being present in the data. These inverse terms use an exponential function to limit the effect of low values in the models. Through participant comments and our own observations, we noticed the vertical sine-wave patterns produce very strong linear cues that induce subjects to continue perceiving the pattern when, locally, there is no clear evidence of it. Limiting this effect will yield more realistic results applicable in real cases, where such strong linear structures are not present.

The results of the regression are detailed in Table 5.2. Note that only coefficients with a significant contribution capturing the variation in the data are used in the final models.

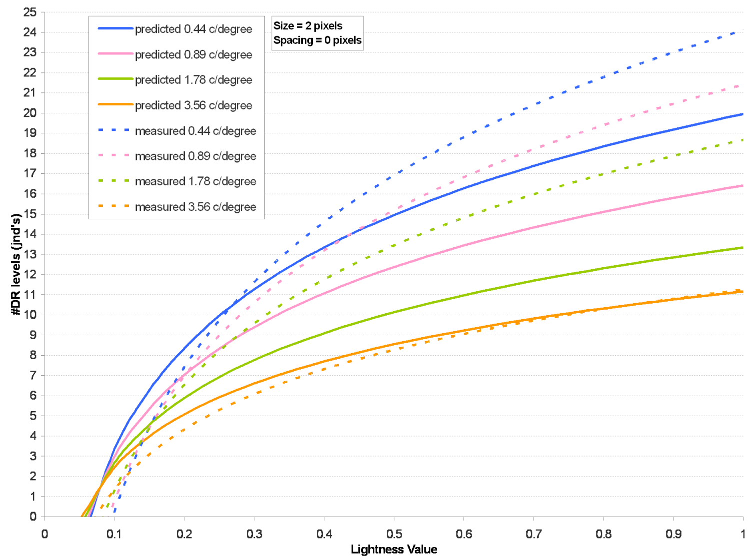


Figure 5.5: Data resolution results for the lightness method. This figure shows the initial fitted functions (dotted lines) and the results using the overall model (solid lines) for different data frequencies.

It is important to note that the logarithmic model for w_1 can only be applied when

$$\lambda < \frac{900}{2 * w_0(v_i)}$$

with w_0 measured in pixels and λ in $c/width$. Any values obtained beyond this limit would come from extrapolation but they are false, since the size and spacing values limit the precision of the method in terms of spatial feature resolution.

For the purposes of facilitating the comparison of our results with previously published psychophysical results, we can transform our λ values from $c/width$ units to $c/degree$ (denoted as $\hat{\lambda}$). Using an approximate value of 45 degrees for the total field of view occupied by our stimuli during the experiment, we obtained values of $\hat{\lambda} \in \{3.56, 1.78, 0.89, 0.44\}$ for our four experimental conditions.

The coefficients shown on the table were obtained through regression on the mean results from the experiment. The resulting standard error for all models is approximately ± 2 levels. Figure 5.5 shows the comparison between the experiment data and the model predictions. Despite the lack of fit for this particular instance ($s_i = 2$ and $p_i = 0$), the model fits the overall space of parameters very well. This combination of parameters shows the highest DR values of all.

The power of this model comes from the fact that it provides an understanding of the distribution of $jnd's$ throughout the range of our visual dimensions. Given a dataset to be represented, we can chose the best method to represent it depending upon the distribution of its values across the data range. For example, although data resolution values for size are the lowest of all three methods, as the spacing and size of the icons for the other methods increase and their expressiveness capabilities deteriorate, size becomes the best method to be used.

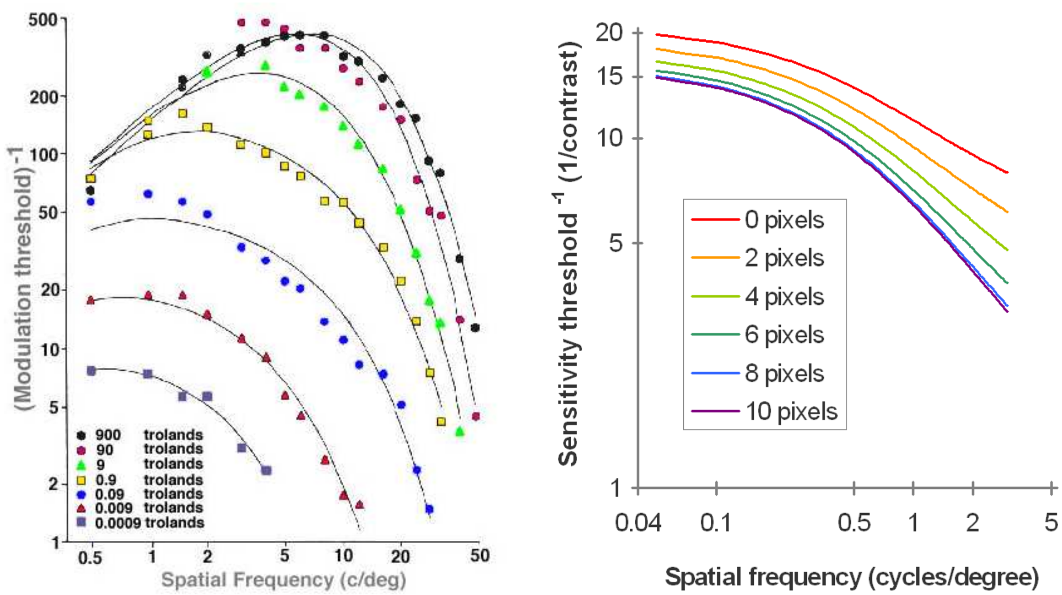
These are exactly the type of data we wanted to gather from our experiments. It is clear that the different methods have different optimal conditions to be applied, and these models allow us to determine which method is best suited to represent the data in different situations.

In comparing our results with the existing literature, Bertin [Bertin, 1983] provides one of the few examples of quantitative results for data resolution values. Although he does not explicitly measure them for icon lightness, he recognizes that the smaller the icons, the fewer levels our perception should be able to differentiate. Our results contradict that for all spacing levels. For size, he proposes an average of 20 distinguishable levels when the ratio between the extremes of the range is 1 to 10. Our range is only 2 to 10 and our results are lower than those 20 levels (approximately 7 levels for the lowest spatial frequency data tested), as expected. The more surprising result is that, for icon spacing, our participants can differentiate a maximum of 9 levels (for $s_i = 2$ pixels and $\lambda = 20\ c/width$), while Bertin does not expect more than 5.

In general, the results obtained follow our expectations. Our perceptual system contains specialized cells to detect lightness and saturation changes, whereas size and spacing changes seem to get processed differently. Our results validate this trend of better results for icon lightness. They also generate some surprising evidence for the perceptual ordering of icon size and spacing.

To further validate our model, at least for the lightness case, we compare it now to existing psychophysical models for contrast sensitivity. Note that this is the closest experimental results we can compare to. Contrast sensitivity functions (CSF) describe how our threshold for detecting a change in lightness decreases, up to a point, as the spatial frequency of the signal (a sine-wave pattern similar to ours) increases. Figure 5.6(a) shows several CSFs for different overall luminance of the stimuli. Note that, the lower the overall brightness, the less sensitive we become to lightness changes.

Even though our stimuli are based on discrete icons, and we are interested in measuring the effect of the size and spacing of those icons on the utility of our methods for scientific visualization, we can try to obtain a similar set of curves than the ones depicted in Fig. 5.6(a).



(a) (b)

Figure 5.6: (a) Contrast sensitivity functions (CSFs) for different overall illumination levels [Lamming, 1991]. (b) Resulting CSFs from our model for lightness. We simulate variable overall illumination (measured in trolands) by modifying the icon spacing parameter (measured in pixels). The same trend can be observed in both graphs: i.e. the sensitivity threshold increases as illumination decreases.

Using our model, the contrast[1] threshold necessary to detect a change (an increase or decrease of 1 jnd) is constant for a particular data spatial frequency at any lightness level. If we then modify the spatial frequency values, we can plot the resulting threshold values as frequency increases (see Figure 5.6(b)).

Furthermore, we can simulate the effect of modifying the overall display luminance by plotting how these curves change as the spacing of the icons increases, effectively lowering the overall luminance. Those curves are also shown in Fig. 5.6(b). Following the same trend as in Fig. 5.6(a), the threshold for change detection increases as the overall display luminance decreases. It is important to point out that the threshold values obtained and the shape of the curves is not indicative. Our measure of lightness is based on the average value of the red, green, and blue components sent to the graphics engine. We have not measured the actual brightness output of the monitor, hence we cannot directly compare our values. Also, our discrete representation of the data (the sine-wave) using icons creates a high frequency signal at the borders of the icons, which interferes with the spatial frequency of the data at low values of the latter. This could be the reason we do not observe the loss

[1] Contrast is a ratio between the maximum and minimum lightness of an image $(maxL-minL)/(maxL+minL)$. In our case, this ratio changes horizontally across the image as the amplitude of the sine-wave decreases.

of sensitivity at low spatial frequencies. However, the trend of the curves, as the overall luminance changes, shows the expected behavior.

5.1.3 Methodology Discussion

Even though we dramatically reduced the number of combinations of visual dimensions we explored, the experiment posed a big design challenge. Participants commented on its apparent extraordinary length due to the similarity of all the images. As we saw from the results, the actual values obtained establish clear differences. With these 6 participants we were able to fully randomize the order of the three tasks to eliminate any possible learning effects. Nevertheless, fatigue was a big factor that, although it did not explicitly show up in the data collected, will require moving to shorter between-subjects designs, or even multiple sessions, for subsequent experiments.

With this experiment we solved the high variance issue we had during our previous study with expert visual designers. The cost for this was two-fold. First, participants evaluated methods one at a time, impeding direct comparisons among displays that were possible in the first study. Comparative critique is a very useful tool design educators utilize constantly, but one that we had to sacrifice to improve the quality and quantity of data obtained. Secondly, we did not use expert visual designers as subjects, so we could not expect feedback on why a method performs as it does for a given task. Our tasks now are more perceptual than conceptual and the low variance of the data, along with consistent trends, validates our choice of non-expert participants.

We have also performed a small set of pilot trials to assess the possibility of including data resolution trials where size and spacing varied with the data while lightness and saturation were modified constantly across the display, completing this way a full permutation of the four methods. Our results showed no effect of lightness or saturation in the DR results for size and spacing. Subjects were able to detect the sine-wave pattern approximately the same no matter what lightness or saturation the icons were. While saturation levels did not even affect the time subjects took to perform the task, lightness levels did, in fact, affect the speed of the responses (low lightness required longer time), but not the accuracy.

5.1.4 Conclusion

In summary, we have obtained two sets of predictive models for spatial feature resolution and data resolution. They can be used to quantify the utility of each of the methods we studied as conditions change for our independent variables. Furthermore, the limits resulting from the spatial feature resolution model help constrain the applicability of the data resolution models to useful ranges. Explicitly exposing these limitations to the users will help them

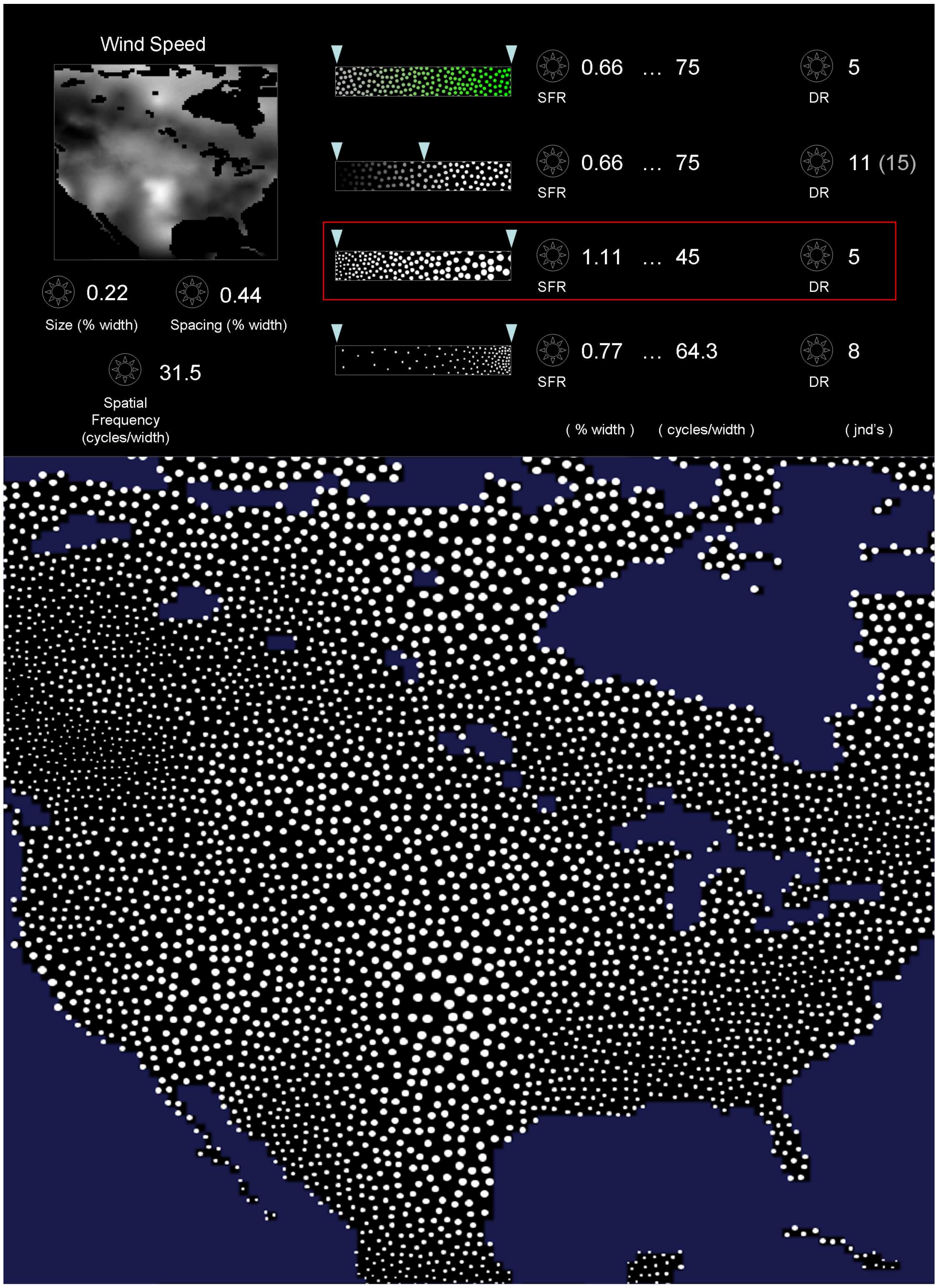


Figure 5.7: The models obtained up to this point allow us to control the individual capabilities of each method. Given a dataset, its spatial frequency, and a choice of size and spacing parameters for the icons, we can present users detailed information about the spatial feature resolution allowed by each method, and its data resolution capabilities. Note that users can also decide whether the full range of the visual dimension is used or only a portion of it. In this example, using the bottom half of the lightness range would yield 11 out of 15 perceivable levels of lightness for the full range. The current choice, shown in the large display, is highlighted in red.

make effective decisions when choosing a visual representation for their data.

Going back to our vision for this thesis shown in Fig. 1.1, we have now advanced towards that goal in that we have individual control of the capabilities of each visualization method. Figure 5.7 shows a mockup display that would use the currently proposed models. Size, spacing and, data spatial frequency can be set and SFR and DR values can be obtained for all methods. The spatial frequency can be obtained directly from the data or it can be chosen by the user to investigate the capabilities of the different methods. Also, the figure includes knobs for the SFR and DR design factors. With the models we have obtained it is possible to modify those knobs and find the size, spacing, and frequency required to achieve the values requested.

5.2 Modeling Perceptual Dominance Among Visual Cues in Multilayered Icon-based Scientific Visualizations

In this next experiment we quantify how factors such as icon size, spacing, layer order and color affect the relative saliency and interference among five different 2D scalar visualization methods: saturation (SAT), lightness (LIGHT), orientation (ORIENT), size (SIZE), and spacing (SPA). This experiment should get us one step closer to our goal: we are moving from single-valued to two-valued datasets, and we are exploring the perceptual interactions among the different methods.

We define saliency as the perceived dominance of some visualization method over another when representing scientific data. This means that perception and correct understanding of the data must be assessed, not just the realization that some property of the icons is changing across the display (which a preattentiveness analysis would assess.) For example, orientation changes are very preattentive. Yet, as we will see, reading a scalar field from changes in icon orientation is very difficult, making it, in our definition, not very salient with respect to other methods. We measure saliency as the difference in time that participants take to recognize each of the datasets in our stimuli (see Figure 5.8).

Saliency can be used to visualize the importance of some variables over others: users may want some variables to dominate the composition while others should recede to the background to serve as context. They may, however, want all variables to have similar dominance of the final display, so not to highlight any particular one and bias the exploratory analysis of the dataset. These relationships among data variables get translated to the visualization methods as the saliency of those methods.

Our experiment also recovers the perceptual interference among methods, which we define as the amount of distraction a method creates when users are trying to read another method present in the same display. We define these interferences as the time participants

a) b) c) d)

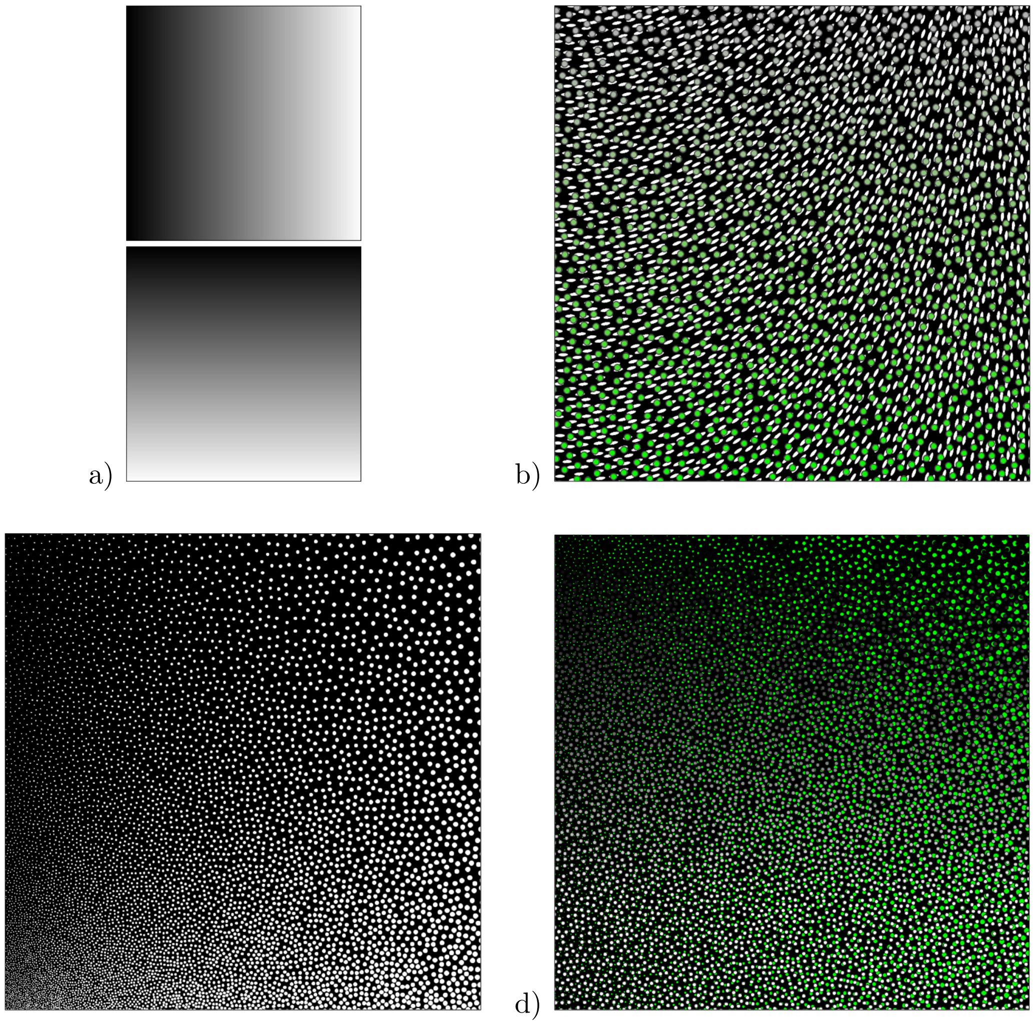


Figure 5.8: Sample stimuli for the experiment. (a) The two linear scalar fields used in the experiment. Images (b)-(d) are examples of the stimuli presented in the study: they all represent both linear fields simultaneously, and participants were asked to judge which one they perceived first, i.e. which one is the more salient of the two. (b) Icon orientation on the bottom layer and saturation on the top, (c) a single-layer example with size and spacing, (d) another two-layer display with size on the bottom and lightness on top. Top-layer icons have a gray-valued border at half the lightness value of the inside circle, so as to minimize simultaneous contrast issues. Circles are used for all methods except orientation, which uses ellipses.

take to recognize each method while the distractor method is simultaneously changed and all other factors of the final display are controlled.

Note that these two factors correspond to the dominance, visual bandwidth, and time to read factors we used in Section 4.2. We will correlate the results from this experiment to the subjective ratings designers gave us in that previous study.

Our main contribution is a set of predictive models that, given a particular combination of methods, approximates the expected perceptual interference among them and the saliency level of the combination. This is a useful tool in generating effective visualizations based on the perceptual characteristics of the methods involved. Furthermore, with the derivatives of these models, we can confidently guide the user towards higher or lower saliency and interference by changing some or all of the factors involved.

5.2.1 Methodology

Our experimental methodology is inspired by psychophysical studies on visual search and cue interaction [Callaghan, 1984; Bergen and Landy, 1991]. We developed an experiment in which the stimuli resemble real visualization displays, which are notably difficult to evaluate perceptually. While still effectively controlling the experimental factors, this methodology allows us to generalize our results, and our predictive models, to real applications with complex multivalued datasets.

Experimental Factors

In order to control the saliency of a method we use a set of *knobs* that control some of our visual dimensions. Here, we analyze and model how the independent variables icon size, spacing, color, and layer order affect the saliency of five scalar visualization methods: icon saturation, lightness, orientation, size, and spacing. The independent variables are not tied to data and remain constant across the display, while data variables are mapped to methods. We decided to include orientation so we could compare our results with scientific visualization literature that identifies orientation as an effective dimension for scalar data visualization.

We measure saliency through a visual-attention experiment. Using displays that show a two-valued scalar dataset (see Figure 5.8) and measuring the time participants take to recognize each of the values, we obtain a model of saliency in terms of how much the two times differ.

We presented the experimental task as a question to the participants: "Which of the two linear gradients do you perceive and understand first? Once you understand one of the gradients, hit a key (H or V) to indicate whether it is the horizontal or the vertical gradient.

After that, continue exploring the image until you either understand the second gradient, in which case you hit the other key, or the image times out after 10 seconds". A one-second distractor image was placed between stimuli so as to minimize carry-over effects from the previous choice.

The dependent measure is the time participants take to hit H (t_H) and V (t_V). The independent variables are:

- Number of layers: (*2 levels*) Either 1 or 2 layers are possible.
- Order of the layers: (*2 levels*) This indicates which method is on the top layer.
- Size and spacing of the icons on each layer: (*3 levels each*) For the 1 layer case these are *2 factors* when size and spacing themselves are not the methods involved, *1 factor* when one of them is involved, or *no factor* when both are involved. For the 2 layer case, these will be *4, 3,* and *2 factors* respectively.
- Color: (*2 levels*) For cases where neither of the two visual dimensions involve color (i.e. orientation, size, and spacing), this variable indicates which of the two will be colored and which will be white.
- Directional mapping: (*2 levels*) This indicates which method is used horizontally and which vertically.

Directional mapping is part of the experiment to avoid biasing the results due to lack of control for this variable. A horizontal orientation of the dataset might be easier to recognize than a vertical, due to a known natural human preference to more easily perceive horizontal things. We wanted to investigate its effect, if any, on saliency. Ranges for the five methods and levels for the size, spacing and color independent variables are explained in Table 5.3.

Stimuli

To facilitate the direct application of our results to real visualization cases, with continuous scalar fields, we would like to measure saliency utilizing those types of datasets. Unfortunately, asking subjects to recognize such datasets requires that we are confident they understand all parts of them such as extrema, gradient variations and other details. This would require subjects to be extremely familiar with the data, or us to provide ground truth examples for each dataset separately to compare against. In either case, it would take a long time for subjects to go through each visualization display making sure they perceive and understand the correct dataset.

With this in mind, we designed our study to capture some of the continuity of real datasets, while maintaining a low level of required knowledge about the data for the subjects.

		VISUAL DIMENSIONS					
		HUE	SATURATION	LIGHTNESS	ORIENTATION (degrees)	SIZE (pixels)	SPACING (pixels)
VISUALIZATION METHODS	SAT	green	[0,1]	0.6	--	2, 6, 10	0, 5, 10
	LIGHT	--	0	[0,1]	--	2, 6, 10	0, 5, 10
	ORIENT	--, green	0, 1	1, 0.6	[0,90]	2, 6, 10	0, 5, 10
	SIZE	--, green	0, 1	1, 0.6	--	[2,10]	0, 5, 10
	SPA	--, green	0, 1	1, 0.6	--	2, 6, 10	[0,10]

Table 5.3: Parameterization for the five visualization methods utilized in the experiment. Each row indicates the parameters for each method in terms of the relevant dimensions defined. Gray cells indicate the range of the mappings to the data variable. Since color is one of our binary independent variables, the color (in red) or no-color (in black) settings are indicated for orientation, size and spacing methods. Size and spacing, as independent variables, have three levels each, as indicated in the two right-most columns of the table.

We decided that two perpendicular linear datasets would provide such a stimulus (see Fig. 5.8). To our knowledge, even this simple combination of datasets has not been studied before from the point of view of visual cue interference in the cognitive or visualization literature.

We presented our stimuli on a 1280x1024 CRT monitor. Visualization displays were images of 900x900 pixels on a black background. The illumination of the room was kept low to avoid distraction when the changing images flashed on the screen, and we gamma-corrected both lightness and saturation ranges for approximate visual linearity. Subjects sat at approximately 30” from the monitor with no chin rest.

For the two layer cases, icons on the top layer have an extra border around them to facilitate differentiation of the layers and to try to minimize simultaneous contrast issues with background or other icons. The border width is approximately 20% of the internal diameter of the icons, with spacing values also measured from this internal part. For orientation cases, we use ellipses where the small axis has the size characteristics of regular circular icons, and the big axis is three times the length of the small one.

Experimental Logistics

We performed a full factorial design for all factors in the one-layer cases and, for the two-layer cases, we use a blocked randomized fractional factorial design using an orthogonal array [Heydayat et al., 1999] for the size and spacing factors of both layers. Fractional factorial designs have fewer trials than full factorial ones, but some effects become confounded. Using an orthogonal array to choose what trials to run from the full factorial set, we assure that at least all main effects are estimable free of interactions. In particular, we used an orthogonal array $L_9(3^4)$, which gives us 9 combinations to test as opposed to the full factorial of 81 that two factors per layer and three levels each would yield. This is still a balanced design, since each level of each of the variables occurs equally often (3 times in this case). With this particular design, interaction effects cannot be estimated. We weighted the possibility of including the interaction effects among size and spacing values for both layers, but this would mean going to an array with 27 combinations, dramatically increasing the time participants needed to complete the study. Based on our previous experiments, we decided to minimize the possibility of subject fatigue. Being the first model of its kind, to the best of our knowledge, we believe a main effects model will provide important clues towards the inclusion or not of interaction effects in later experiments. For all other independent variables we used a full factorial design.

We created a blocked design for our experiment. Table 5.4 shows all blocks present in the study based on pairs of visualization methods. The table shows what a full factorial design would be. Subjects were introduced to the experiment and taught the task at hand. Each subject ran through 5 full blocks of method pairs. We divided each block into two sections, one per directional mapping setting. That made 10 sections of approximately 30 stimuli each. Each section contained all combinations for number of layers, layer order, color, size and spacing, following the mixed fractional and full factorial design explained above. Training was provided before each section so subjects could familiarize themselves with the pair of methods being studied in that section. We used eight images for training, a number that we reached after piloting the experiment and concluding subjects understood both methods and the task well enough. The order of the blocks was randomized and stimuli within each block were also randomized.

A total of six paid participants ran through the experiment, taking approximately one hour to complete the study with short breaks between sections. All of them were graduate students from Brown University with various levels of computer expertise, no previous scientific visualization experience, and normal or corrected to normal vision. Since we are measuring perceptual saliency, we did not require special knowledge of visualization or computers from our participants.

5.2.2 Results

Before going further with the analysis of the results, we normalized the times per subject to eliminate the variability in perceptual skills among subjects. This is validated by the fact that standard deviations for all subjects were comparable (between $0.9s$ and $1.7s$), while mean times per subject for the first key stroke for the full experiment ranged from $1.9s$ up

SINGLE LAYER		Horizontal Variable				
		SAT	BRI	ORI	SIZ	SPA
Vertical Variable	SAT			(3x3)	(3x1)	(1x3)
	BRI			(3x3)	(3x1)	(1x3)
	ORI	(3x3)	(3x3)		(3x1)	(1x3)
	SIZ	(1x3)	(1x3)	(1x3)		(1x1)
	SPA	(3x1)	(3x1)	(3x1)	(1x1)	

TWO LAYERS		Horizontal Variable				
		SAT	BRI	ORI	SIZ	SPA
Vertical Variable	SAT		(3x3)x (3x3)x2	(3x3)x (3x3)x2	(3x3)x (1x3)x2	(3x3)x (3x1)x2
	BRI	(3x3)x (3x3)x2		(3x3)x (3x3)x2	(3x3)x (1x3)x2	(3x3)x (3x1)x2
	ORI	(3x3)x (3x3)x2	(3x3)x (3x3)x2		(3x3)x (1x3)x2x2	(3x3)x (3x1)x2x2
	SIZ	(1x3)x (3x3)x2	(1x3)x (3x3)x2	(1x3)x (3x3)x2x2		(1x3)x (3x1)x2x2
	SPA	(3x1)x (3x3)x2	(3x1)x (3x3)x2	(3x1)x (3x3)x2x2	(3x1)x (1x3)x2x2	

(a)

	layer 1		layer 2	
trial	size	spacing	size	spacing
1	6	0	6	0
2	6	5	2	10
3	6	10	10	5
4	2	10	6	10
5	10	10	2	0
6	10	0	10	10
7	10	5	6	5
8	2	5	10	0
9	2	0	2	5

(b)

Table 5.4: Experimental organization tables. (a) Each cell indicates the number of stimuli for each combination of visual dimensions. In the *single layer* case, the numbers indicate (*size* x *spacing*) levels. In the *two layer* cases the numbers indicate ($size_v$ x $spacing_v$) x ($size_h$ x $spacing_h$) x *layerorder*. Note that for the SIZE, SPA, and ORIENT combinations we also controlled for the layer color (final x2). For simplicity, the table shows the full factorial combinations. Size and spacing, as independent variables, were subject to the orthogonal array design shown in (b), where all levels are indicated in pixels.

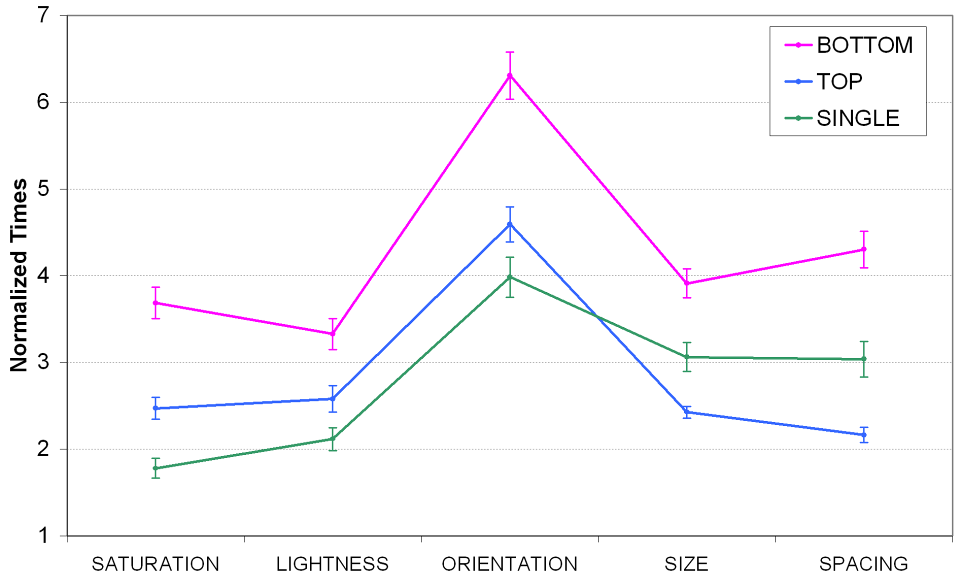
to 4.0s. Normalizing the results maintains the relative order of the measures while losing the units, seconds. This is acceptable in our case, since we will still obtain a numerical, yet relative, quantification of saliency and interference. Furthermore, if we were able to maintain the units of time, they would not be applicable to other more general data visualization cases that did not match our experimental setup with two linear datasets. After normalization we translated all values so no normalized times would be negative. This step does not affect the results, and facilitates the explanations and conclusions.

We must also take into account the cases where subjects reached the 10s timeout limit. When neither key had been pressed we eliminated that instance. This occurred 22 times out of 1,680 total stimuli presented. The more interesting case occurred when the first key was pressed within the 10s limit and the second was not. Those cases indicate a clear dominance of the first display detected over the second one. This timeout case occurred a total of 300 times over the course of the experiment. Given our fractional factorial design, all trials are key to obtaining meaningful information about the interaction among the independent variables. Simply discounting these trials would effectively make the analysis infeasible. Our approach to solve this was to consider the timeout trials as censored observations.

We used the LIFEREG procedure in the SAS statistical program [Cody and Smith, 2006] to perform a maximum likelihood estimation of the time responses for all 300 timeout observations. This procedure fits a Weibull distribution to what is essentially unknown "failure" times for the timeout cases, modeling them as a function of the subject ID (since we are dealing with normalized times and 10 seconds translates to a different value for each subject), the size and spacing values for each method, their layer order, their color, and their directional mapping. Although inputing the estimated censored data resulting from this procedure at its conditional mean is common practice, we decided to utilize the 95% quantile results to be on the safe side. Indeed, once participants decide not to hit that second key in less than 10 seconds, we have no way of knowing whether 20 or 60 seconds would be the time they would require. This procedure allowed us to estimate those values with confidence, based on the distribution of the non-censored observations.

Overall Times

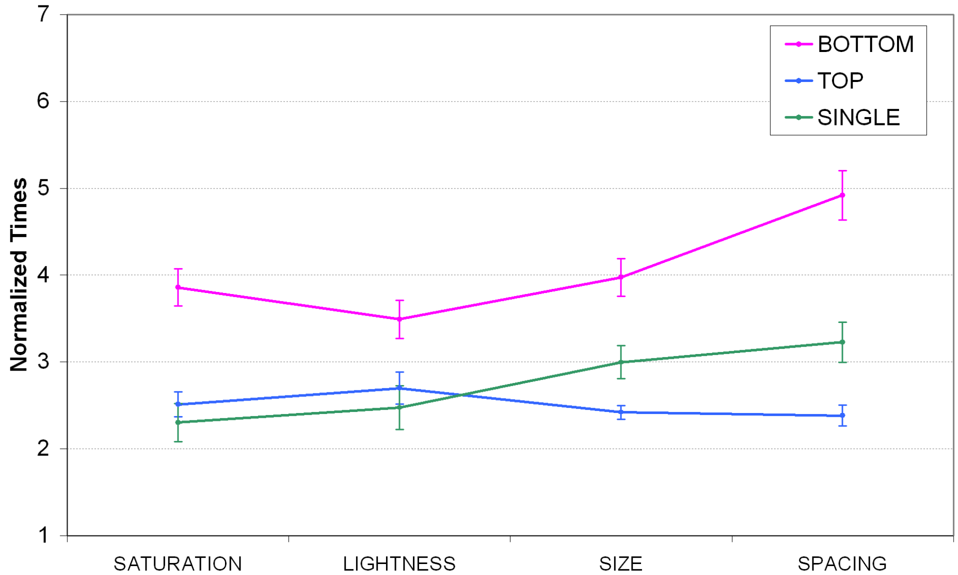
Figure 5.9 shows the normalized times of all five methods in all conditions. The mean times to recognize orientation as a scalar field are significantly higher than the rest. All participants declared difficulty understanding orientation as a scalar value. The pseudo-flow effect was so distracting as to prevent them from understanding the linear scalar datasets. For this reason, all further analyses of the experimental data exclude orientation cases. Figure 5.10 shows the mean times excluding all those from orientation stimuli. Eliminating



α=0.0125		SINGLE		BOTTOM		TOP	
		F	p	F	p	F	p
SAT	BRI	1.390	0.239	1.170	0.279	0.270	0.602
SAT	ORI	73.080	<.0001	71.830	<.0001	118.360	<.0001
SAT	SIZ	15.940	<.0001	0.530	0.469	0.050	0.822
SAT	SPA	15.480	<.0001	3.990	0.046	2.510	0.113
BRI	ORI	52.030	<.0001	93.440	<.0001	105.260	<.0001
BRI	SIZ	8.610	0.004	3.560	0.060	0.610	0.435
BRI	SPA	8.240	0.004	10.020	0.002	4.550	0.033
ORI	SIZ	9.640	0.002	70.060	<.0001	143.820	<.0001
ORI	SPA	10.290	0.001	49.170	<.0001	181.060	<.0001
SIZ	SPA	0.010	0.944	1.890	0.170	2.160	0.142

α=0.025		SAT		BRI		ORI		SIZ		SPA	
		F	p	F	p	F	p	F	p	F	p
SINGLE	BOTTOM	46.700	<.0001	16.840	<.0001	32.460	<.0001	7.310	0.007	10.380	0.001
SINGLE	TOP	6.180	0.013	2.450	0.118	2.210	0.137	4.110	0.043	4.950	0.026
BOTTOM	TOP	34.730	<.0001	11.770	0.001	28.340	<.0001	71.770	<.0001	94.230	<.0001

Figure 5.9: Mean normalized times and standard errors for one- (SINGLE) and two-layer (BOTTOM and TOP) cases. Significant differences are indicated in red on the tables. The top table contains between-method comparissons, while the bottom table shows the within-method test results. Note that the α values are adjusted based on the number of significance tests performed using the Bonferroni adjustement over an original value of $\alpha = 0.05$.



α=0.017		SINGLE		BOTTOM		TOP	
		F	p	F	p	F	p
SAT	**BRI**	0.270	0.601	1.100	0.294	0.930	0.335
SAT	**SIZ**	4.820	0.030	0.120	0.733	0.230	0.632
SAT	**SPA**	8.680	0.004	9.780	0.002	0.460	0.497
BRI	**SIZ**	2.680	0.104	2.050	0.152	2.170	0.141
BRI	**SPA**	5.670	0.019	17.890	<.0001	2.800	0.095
SIZ	**SPA**	0.580	0.446	8.410	0.004	0.040	0.836

α=0.025		SAT		BRI		SIZ		SPA	
		F	p	F	p	F	p	F	p
SINGLE	**BOTTOM**	29.330	<.0001	4.210	0.041	6.140	0.014	10.620	0.001
SINGLE	**TOP**	0.230	0.632	0.210	0.650	2.120	0.146	2.660	0.104
BOTTOM	**TOP**	12.710	0.0004	7.990	0.005	48.320	<.0001	73.040	<.0001

Figure 5.10: Mean normalized times, after eliminating all observations containing ORIENT, and standard errors for one- (SINGLE) and two-layer (BOTTOM and TOP) cases. Significant differences are indicated in red on the tables. The top table contains between-method comparissons, while the bottom table shows the within-method test results. Note that the α values are adjusted based on the number of significance tests performed using the Bonferroni adjustement over an original value of $\alpha = 0.05$.

these observations from the analysis also eliminated 12 out of the 22 no-response cases and 158 of the the 300 timeout cases. We recalculated the right-censored data estimates after removing the orientation trials.

Interesting to note from these graphs is how size and spacing methods are recognized faster when they are on the top layer of two-layer cases than for single-layer cases. Although this is only a trend in the data (these differences are not significant at the chosen α levels), this seems to confirm the known preattentive precedence of the other three methods over these two for the single-layer cases.

The graphs in Figs. 5.9 and 5.10 show the overall experiment trends. However, we want to explore the relative saliency and interference between each pair of methods for all conditions. Also, the non-significance of some differences from Fig. 5.10 comes from the fact that the overall times combine cases when each method was recognized first and second, hence we cannot assume a normal distribution. A more appropriate way of looking at these data is shown in Fig.5.11 and explained later in this section.

Directional Mapping

Before going into a detailed analysis of each pair of methods and the performance of each one, we evaluated whether the orientation of the dataset in the display had any effect in the choices participants made.

We performed a correlation analysis between the number of times each method was chosen first and their orientation when that was the case. Table 5.5 shows the results of the analysis. Even though the LIGHT method shows a significant tendency to be chosen first when it is presented vertically, an analysis of the times for this case (LIGHT chosen first) indicates that the times when LIGHT is vertical are not significantly different than the times when it is horizontal ($F(1, 223) = 0.16, p = 0.687$). From this analysis we concluded the orientation of the stimuli did not have a significant effect in the experiment. Therefore we utilized both directional mappings as a repeated measure of the same stimuli.

Relative Saliency and Perceptual Interference

Figure 5.11 shows the mean time differences for each pair of methods. These values are averaged over all size and spacing levels. Values significantly different from zero indicate a significant dominance of one method over another, and it is clear from the graph that the strength of the dominance varies greatly. For example, the main trend is that the method displayed on top for the two-layer cases is usually dominant, although that is not the case for LIGHT and SPA when combined with SAT. In those cases, with SAT on the bottom layer, there is no clear salient method. SIZE, on the other hand, is more salient than SAT in

	horiz./first	p
SAT	102/210	0.365
LIGHT	98/225	0.031
SIZE	96/204	0.221
SPA	109/238	0.109

Table 5.5: This is a summary of the directional mapping analysis of the experimental data. The left column shows the number of times a method was horizontal (numerator) when chosen first (denominator). Given that the choice is a binary variable, the p value is calculated from a binomial distribution with probability 0.5, since each method was shown the same number of times horizontally and vertically.

		BOTTOM			
		SAT	LIGHT	SIZ	SPA
TOP	SAT		SAT	SAT	SAT
	LIGHT	--		LIGHT	LIGHT
	SIZ	SIZ	SIZ		SIZ
	SPA	--	SPA	SPA	

Two-layer combinations

	SAT	LIGHT	SIZ	SPA
SAT			--	SAT
LIGHT			LIGHT	--
SIZ				--
SPA				

One-layer combinations

Table 5.6: These tables indicate the method that is expected to dominate in each pair. The strength of the saliency is not indicated here and can be obtained from Fig. 5.11. Equally salient methods are indicated with –.

that situation. We summarize all these trends in Table 5.6, where the method expected to be salient is indicated for one- and two-layer cases. It is important to note that the key to this experiment is not only to find these general trends, but to identify in what conditions the less obvious solutions provided the desired effect, i.e., when and how do we get the bottom layer to be more salient?

While the time differences offer information about the relative saliency of the methods, the actual times to recognize each method can shed light on the perceptual interference one method causes another. If a method that performs well in general, with relatively low recognition times (e.g., SAT), has high recognition times in one particular situation (e.g,. SAT on the bottom layer combined with SIZ), we can conclude that the other method in that situation interferes significantly in the process of understanding the first. Figure 5.12 shows the summary of recognition times for all methods in all pairs tested. Observe the similar patterns SAT and LIGHT have for all cases. Also, the similar response that SIZ and SPA have against SAT and LIGHT.

From that figure we can predict what methods will be more easily interfered with.

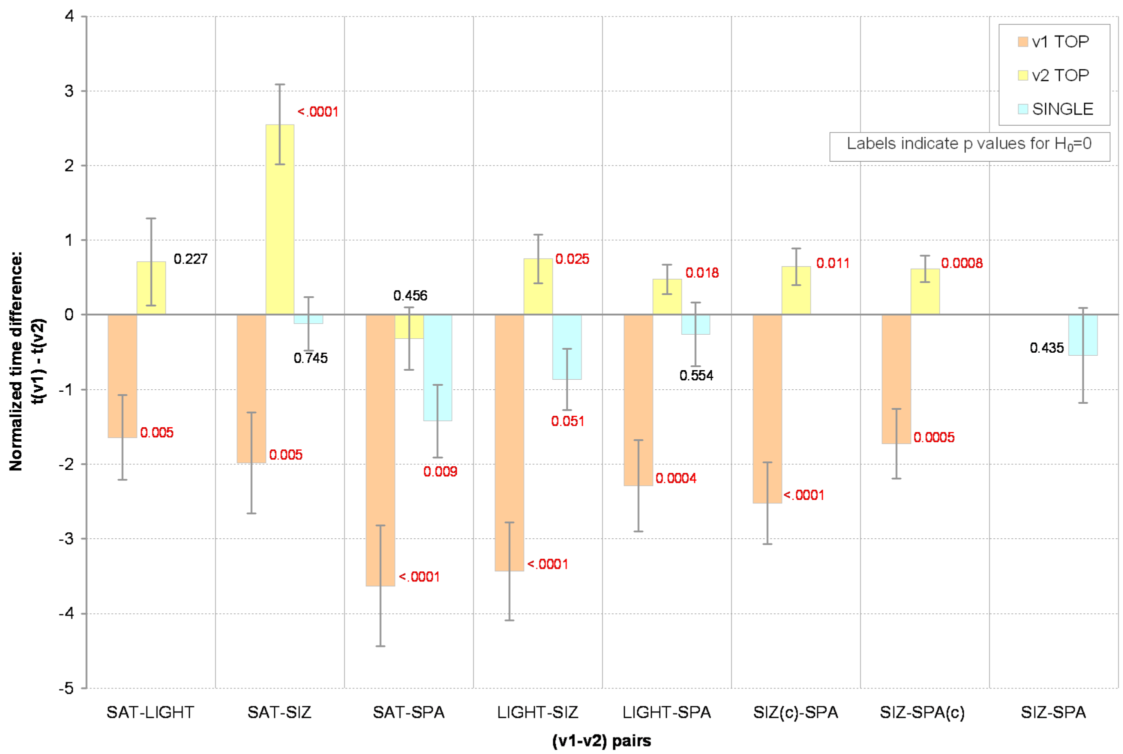
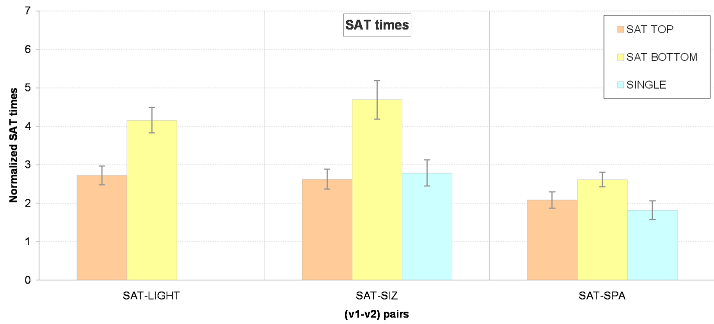
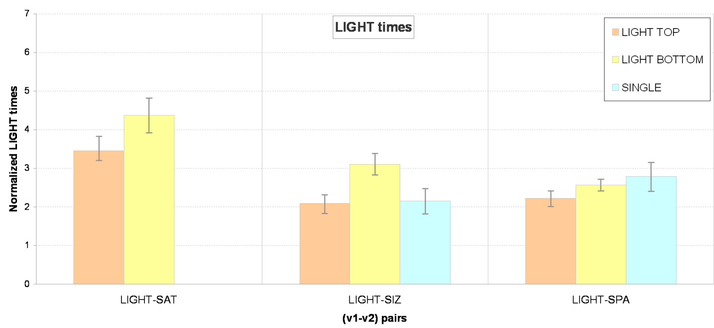


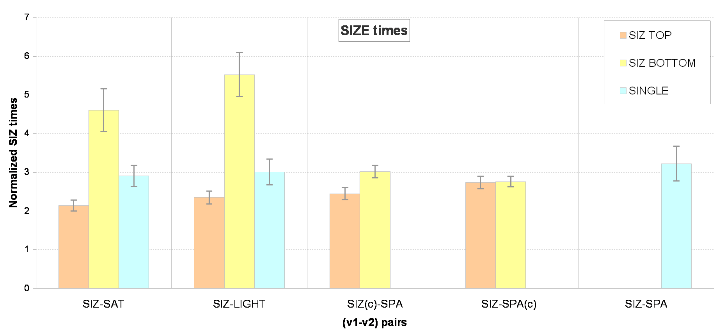
Figure 5.11: Mean normalized time differences and standard errors for each pair of methods for one- and two-layer cases. v_1 is always the first method of the pair's name, so negative difference values mean the time to recognize v_2 is longer than the time for v_1. For SIZ and SPA methods, (c) indicates the colored method.



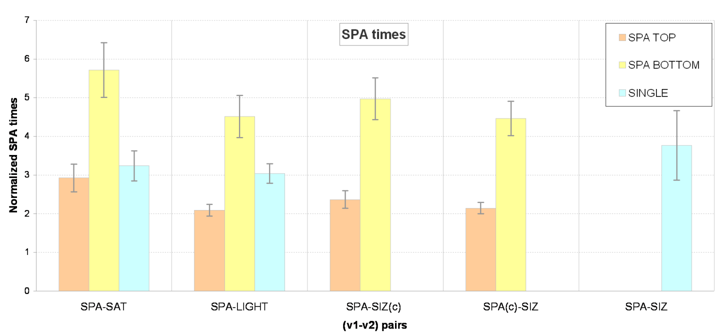
	TOP-BOTTOM		TOP-SINGLE		BOTTOM-SINGLE	
	F	p	F	p	F	p
SAT-LIGHT	12.4100	0.0006	--	--	--	--
SAT-SIZ	12.4100	0.0006	0.0500	0.8296	6.4100	0.0126
SAT-SPA	3.8500	0.0521	0.4800	0.4907	4.3200	0.0398



	TOP-BOTTOM		TOP-SINGLE		BOTTOM-SINGLE	
	F	p	F	p	F	p
LIGHT-SAT	2.4100	0.1226	--	--	--	--
LIGHT-SIZ	8.6200	0.0040	0.0100	0.9058	3.6600	0.0581
LIGHT-SPA	1.8100	0.1812	2.4700	0.1189	0.3700	0.5415



	TOP-BOTTOM		TOP-SINGLE		BOTTOM-SINGLE	
	F	p	F	p	F	p
SIZ-SAT	21.5000	<.0001	1.0500	0.3078	5.1900	0.0245
SIZ-LIGHT	32.0700	<.0001	0.6700	0.4161	9.6300	0.0024
SIZ(c)-SPA	1.6200	0.2050	--	--	--	--
SIZ-SPA(c)	2.0300	0.1558	--	--	--	--
SIZ-SPA	--	--	1.7600	0.1862	0.5000	0.4792



	TOP-BOTTOM		TOP-SINGLE		BOTTOM-SINGLE	
	F	p	F	p	F	p
SPA-SAT	13.9400	0.0003	0.0900	0.7698	5.5100	0.0205
SPA-LIGHT	21.0500	<.0001	1.6600	0.2002	3.9600	0.0489
SPA-SIZ(c)	15.8500	<.0001	--	--	--	--
SPA(c)-SIZ	28.8500	<.0001	--	--	--	--
SPA-SIZ	--	--	1.7500	0.1867	0.7000	0.4051

Figure 5.12: Mean normalized times and standard errors for each method for all combinations tested. The tables show the results of the significance tests performed in each case. Significant values (at $\alpha = 0.017$ after applying the Bonferroni adjustement for multiple tests) are indicated in red. Some values, in green, are not significant under Bonferroni's conservative method but would appear to be so on visual inspection of the graphs.

LIGHT maintains its times quite consistently across the board, followed by SAT, then SIZE, and finally SPA, which fluctuates a lot depending on what layer it is on.

5.2.3 Predictive Models

Given the distributions of timing data for the two key presses for each pair, we obtain the following measures based on the time shown before:

- *Relative Saliency,* $w_2(v_i, v_j) \in (-1, 1)$: Here v_i and v_j are two of our visualization methods. $w_2 = -1$ indicates that v_i is much more salient than v_j, and $w_2 = 1$ indicates the opposite. Differences are normalized with respect to the maximum time difference throughout the experiment, i.e. 10 seconds.

- *Perceptual Interference,* $w_3(v_i|v_j) \in (0, 1)$: This measures how much v_j interferes with the reading of v_i. To measure this, we set $w_3(v_i|v_j) = \frac{t(v_i|v_j) - min(t(v_i))}{max(t(v_i)) - min(t(v_i))}$, where $t(x)$ is the time participants took to recognize method x. To obtain the extreme values we must look across blocks for all instances where v_i was presented with the same parameters (size, spacing, color). We assume that the minimum time is how long a participant would take to recognize a dataset using v_i when presented by itself.

To obtain these measures we have fit a set of models to each one of the recognition times shown in Fig. 5.12. To the best of our knowledge, there are no models that explain this type saliency or interference responses so, following the success of the models utilized for the data resolution factor in the previous experiment, we used the following base equation:

$$
\begin{aligned}
t(v_i) = a_0 s_1 + a_1 p_1 + a_2 s_2 + a_3 p_2 + \\
+ a_4 s_1^2 + a_5 p_1^2 + a_6 s_2^2 + a_7 p_2^2 + \\
+ a_8 e^{\frac{1}{s_1+1}} + a_9 e^{\frac{1}{p_1+1}} + a_1 0 e^{\frac{1}{s_2+1}} + a_1 1 e^{\frac{1}{p_2+1}}
\end{aligned}
$$

It contains linear, quadratic and inverse terms to account for the observed behavior of fluctuation of saliency as size and spacing values are modified. The inverse term uses an exponential term to limit the effect of low values. In this equation, s_1 and p_1 correspond to the size and spacing parameters for the first method of the pair (v_1, v_2), while s_2 and p_2 correspond to the second method's parameters. The same as before, while the full model contains a maximum of 12 degrees of freedom, only coefficients that contribute significantly to explain the variance in the data are included in the models. After performing linear regression for all times in the experiment, Table 5.7 shows the coefficients of the models and the excellent fit results obtained given the high number of error degrees of freedom.

	TWO-LAYER COMBINATIONS														
	s_1	p_1	s_2	p_2	s_1^2	p_1^2	s_2^2	p_2^2	$e^{(1/(s1+1))}$	$e^{(1/(p1+1))}$	$e^{(1/(s2+1))}$	$e^{(1/(p2+1))}$	R^2	F	p
	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}			
SAT		-0.686	0.201	-0.409		0.047		0.041	5.962	-2.344			0.761	F(7,73)= 33.2	<.0001
LIGHT	-0.321	-1.208	-0.137	0.905		0.065		-0.099			8.251		0.851	F(7,73)= 59.6	<.0001
SAT				-0.880	0.028			0.066	12.069	-0.821	-7.253		0.809	F(6,71)= 50.1	<.0001
LIGHT		-0.336	-1.502	1.191			0.089	-0.104	6.824				0.828	F(6,71)= 56.9	<.0001
SAT	-0.120	0.219							1.874				0.778	F(3,50)= 58.3	<.0001
SIZ										2.025		0.855	0.705	F(2,51)= 60.9	<.0001
SAT				-0.341					5.254				0.702	F(2,51)= 60.2	<.0001
SIZ		0.727				-0.044		0.006	-3.129	2.349			0.896	F(5,48)= 82.7	<.0001
SAT		0.165							3.017		-1.982		0.742	F(3,51)= 48.4	<.0001
SPA		-1.999				0.132				1.969	5.707		0.926	F(4,50)= 155.5	<.0001
SAT	-0.513	0.116	0.545		0.033		-0.039		1.723				0.89	F(6,48)= 64.5	<.0001
SPA		-0.238	-0.238						4.532				0.667	F(3,51)= 34.0	<.0001
LIGHT		0.131		-0.231					3.874			-1.261	0.739	F(4,50)= 35.5	<.0001
SIZ		-1.629		-0.813		0.097		0.049	9.586				0.889	F(5,49)= 78.7	<.0001
LIGHT	-0.195	0.153						-0.013	3.330				0.838	F(4,50)= 64.6	<.0001
SIZ		0.128								1.018			0.803	F(2,52)= 106.1	<.0001
LIGHT		0.147							3.081		-1.875		0.817	F(3,49)= 73.0	<.0001
SPA										2.819			0.732	F(1,51)= 139.3	<.0001
LIGHT	0.447		0.451		-0.035	0.006	-0.031						0.874	F(5,48)= 66.4	<.0001
SPA		-0.075									2.007		0.8	F(2,51)= 101.7	<.0001
SIZ(c)		0.182								0.906			0.815	F(2,52)= 114.9	<.0001
SPA		-0.314	0.288							2.894			0.921	F(3,51)= 198.8	<.0001
SIZ(c)		0.421				-0.028				1.259			0.878	F(3,51)= 122.1	<.0001
SPA											1.943		0.681	F(1,53)= 113.5	<.0001
SIZ		0.576				-0.039				0.895			0.847	F(3,51)= 94.1	<.0001
SPA(c)			0.117							3.459	-1.647		0.891	F(3,51)= 139.2	<.0001
SIZ		0.339				-0.021				1.169			0.898	F(3,51)= 149.6	<.0001
SPA(c)		-0.085									2.102		0.829	F(2,52)= 125.7	<.0001

	SINGLE-LAYER COMBINATIONS														
	s_1	p_1	s_2	p_2	s_1^2	p_1^2	s_2^2	p_2^2	$e^{(1/(s1+1))}$	$e^{(1/(p1+1))}$	$e^{(1/(s2+1))}$	$e^{(1/(p2+1))}$	R^2	F	p
	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}			
SAT		0.227								0.995			0.823	F(2,16)= 37.2	<.0001
SIZ		0.202								1.118			0.866	F(2,16)= 51.9	<.0001
SAT									1.517				0.798	F(1,17)= 67.2	<.0001
SPA	1.301				-0.099								0.836	F(2,16)= 40.7	<.0001
LIGHT		0.152								0.820			0.715	F(2,15)= 18.8	<.0001
SIZ		0.156								1.323			0.845	F(2,15)= 40.8	<.0001
LIGHT									2.272				0.763	F(1,17)= 54.7	<.0001
SPA									2.470				0.884	F(1,17)= 129.3	<.0001

Table 5.7: Results from the linear regressions performed on each method's recognition times. Each model coefficient is shown along with the model factor it corresponds to. Only coefficients that significantly contribute to explaining the data variance are included in the models. For the two-layer combinations, the green cells indicate the method on top.

siz1	spa1	siz2	spa2
7	3	2	4

$S(v_1, v_2)$

TWO-LAYER COMBINATIONS

TOP \ BOTTOM	SAT	BRI	SIZ	SPA
SAT		(-0.646)	(-0.171)	(-0.454)
BRI	(-0.579)		(-0.205)	(-0.230)
SIZ	0.324	0.098		
SPA	(-0.267)	(-0.031)		

TOP \ BOTTOM	SIZ	SIZ C	SPA	SPA C
SIZ				0.015
SIZ C			(-0.164)	
SPA		(-0.009)		
SPA C	(-0.035)			

TWO-LAYER COMBINATIONS...DERIVATIVES

SIZ1 — TOP \ BOTTOM	SAT	BRI	SIZ	SPA
SAT		0.021	(-0.015)	(-0.005)
BRI	0.029		0.010	(-0.005)
SIZ	(-0.015)	(-0.025)		
SPA	0.000	(-0.004)		

SPA — TOP \ BOTTOM	SIZ	SIZ C	SPA	SPA C
SIZ				0.055
SIZ C			0.066	
SPA		0.015		
SPA C	0.020			

SPA1 — TOP \ BOTTOM	SAT	BRI	SIZ	SPA
SAT		0.060	0.038	0.153
BRI	0.040		0.118	0.037
SIZ	(-0.027)	0.011		
SPA	0.035	0.011		

SIZ — TOP \ BOTTOM	SIZ	SIZ C	SPA	SPA C
SIZ				(-0.037)
SIZ C			(-0.029)	
SPA		0.030		
SPA C	0.033			

SIZ2 — TOP \ BOTTOM	SAT	BRI	SIZ	SPA
SAT		0.162		0.119
BRI	0.227			0.029
SIZ				
SPA	0.063	0.064		

SPA2 — TOP \ BOTTOM	SAT	BRI	SIZ	SPA
SAT		(-0.020)	0.004	
BRI	(-0.071)		0.025	
SIZ	(-0.039)	(-0.010)		
SPA				

ONE-LAYER COMBINATIONS

	SAT	BRI
SIZ	(-0.008)	(-0.066)
SPA	(-0.254)	(-0.022)

ONE-LAYER COMBINATIONS...DERIVATIVES

SIZ1	SAT	BRI
SIZ		
SPA	0.006	0.000

SPA1	SAT	BRI
SIZ	0.003	0.004
SPA		

Figure 5.13: Using our new predictive models, by simply modifying the size and spacing values for our methods we obtain the predicted saliency sign and its strength and, through the derivatives, we also get information about what the effect of modifying the size and spacing parameters is. Green cells indicate non significant differences (from Fig. 5.11). The method precedence to define v_1 and v_2 is given by the order {SAT,LIGHT,SIZ,SPA}. A negative value indicates v_1 is more salient, and vice-versa.

In general, the quadratic terms contribute very little to the models. Even though they are significant, the coefficient values are almost zero. The inverse terms are consistently the more important factors in explaining recognition times. In particular the ones modifying the icon sizes. This is a logical result given that, in general, increasing the size of the icons makes them more prominent and hence easier to recognize: more salient. This effect is modulated by the inverse spacing factors and the linear factors. This combination allows for the observed effect that saliency between methods sometimes reaches an inflection point along the size or spacing ranges.

As explained before, having this time models at hand we can now generate the relative saliency and perceptual interference metrics.

Relative Saliency

For each pair of methods we can define:

$$w_2(v_i, v_j) = \frac{t(v_i) - t(v_j)}{10}, \quad with \quad w_2(v_i, v_j) \in [-1, 1]$$

The limit in the saliency value is important. We used the calculated estimates for the right-censored data from the experiment to calculate the model parameters. For that reason, some predicted time values will be greater than 10^2, so we placed a cutoff for differences smaller than -10 or larger than 10 with loosing any generality in the predicted saliency estimates.

Implementing this function, we can now predict and control relative saliency between pairs of methods (see Fig. 5.13). Not only that, using the derivative of the model we can provide guidance as to what dimension to modify to increase or decrease saliency in a display (also shown demonstrated in Fig. 5.13).

Perceptual Interference

We define this measure based on how much the recognition time of a method changes when presented in combination with another method. For this we define the following function:

$$w_3(v_i|v_j) = \frac{t(v_i|v_j) - min(t(v_i))}{max(t(v_i)) - min(t(v_i))}, \quad with \quad w_3(v_i|v_j) \in [0, 1]$$

To obtain the extreme values for $t(v_i)$we look at the minimum and maximum predicted times for v_i with the same parameterization but when combined with any other method and in any order of the layers. This values give us an exact measure of how v_j really affect the

[2] Since we are dealing with per-subject normalized times, the maximum normalized censoring point was 10.76. Since we translated the normalized times to make the minimum correspond to zero, the maximum time difference possible during the experiment is approximately 10.

reading of v_i. If the predicted value is already greater than 10 we assume full interference ($w_3 = 1$).

With a similar implementation of this equation as we did for saliency, we can create two interactive tools (one for $w_3(v_i|v_j)$ and another for $w_3(v_j|v_i)$) to explore and control the perceptual interference between the methods. Figures 5.14 and 5.15 show these tools.

5.2.4 Discussion

The complete set of results from this experiment allows us to control and predict the relative saliency and perceptual interference between pairs of methods.

By simply analyzing the time difference plots (Fig. 5.11) we obtain a very much expected result. That is, the method on the top layer will be the most salient of the two. Some exceptions to this rule, and the inclusion of the single-layer cases, make the summary in Table 5.6 a very useful design aid. But the time difference plots show very different strengths for the saliency of the methods.

We generated a set of statistically sound predictive models that we can use to fine tune the saliency of each method. Furthermore, based on the performance of the methods across different combinations, we can also control the amount of perceptual interference they will receive depending upon their companion method.

It is important to note that, given the very simple dataset we used for our experiment and the task participants performed, we do not have information about how accurately the linear dataset is represented or how its integrity is preserved. In fact, for a method such as SPA, it is clear that in the area where the spacing is smaller, it is not possible to see through the other method behind it. Participants declared that they obviously realized this, but they could answer the question anyway by looking through the area with more sparse icons. This local interference was not captured by our results (see how the time plots for SAT-SPA, LIGHT-SPA, and SIZ-SPA in Fig. 5.12 show almost no increase in time, hence no interference caused by this fact).

Comparing our results with existing literature is difficult. The closest experiments are those related to preattentive processes. As mentioned in Chapter 3, these studies tend to focus on visual search type of tasks with a stimulus present/not-present type of question. Yet, they describe a precedence of preattentive perception for different visual dimensions. Looking at our interference measures, our overall precedence is LIGHT, SAT, SIZ, and SPA. This seems to match the findings from [Healey et al., 2004] although our characterization of both interference and saliency goes beyond their discussion and provides a way of controlling those factors as the parameters of the visualization change. Furthermore, the findings from [Healey et al., 2004] and other visual search type tasks were based on single-layer

siz1	spa1	siz2	spa2
7	3	2	4

$I(v_1|v_2)$

TWO-LAYER COMBINATIONS

TOP \ BOTTOM	SAT	BRI	SIZ	SPA
SAT		0.086	0.169	0.008
BRI	0.058		0.192	0.081
SIZ	0.707	0.231		
SPA	0.035	0.186		

TOP \ BOTTOM	SIZ	SIZ C	SPA	SPA C
SIZ				0.171
SIZ C			0.080	
SPA		0.181		
SPA C	0.149			

TWO-LAYER COMBINATIONS...DERIVATIVES

SIZ1 — TOP \ BOTTOM	SAT	BRI	SIZ	SPA
SAT		(-0.011)	(-0.015)	(-0.005)
BRI	0.017		(-0.007)	(-0.005)
SIZ	(-0.009)	(-0.025)		
SPA	(-0.008)	(-0.004)		

SPA — TOP \ BOTTOM	SIZ	SIZ C	SPA	SPA C
SIZ				0.027
SIZ C			0.011	
SPA		0.015		
SPA C	0.012			

SPA1 — TOP \ BOTTOM	SAT	BRI	SIZ	SPA
SAT		(-0.022)	0.022	0.016
BRI	0.007		0.013	0.015
SIZ	0.000	0.015		
SPA	0.012	0.004		

SIZ — TOP \ BOTTOM	SIZ	SIZ C	SPA	SPA C
SIZ				0.000
SIZ C			0.000	
SPA		0.000		
SPA C	0.000			

SIZ2 — TOP \ BOTTOM	SAT	BRI	SIZ	SPA
SAT		0.020		0.031
BRI	0.112			0.029
SIZ				
SPA	0.039	0.033		

SPA2 — TOP \ BOTTOM	SAT	BRI	SIZ	SPA
SAT		(-0.008)	0.000	
BRI	(-0.035)		(-0.017)	
SIZ	(-0.034)	(-0.010)		
SPA				

ONE-LAYER COMBINATIONS

	SAT	BRI
SIZ	0.175	0.000
SPA	0.034	0.641

ONE-LAYER COMBINATIONS...DERIVATIVES

SIZ1	SAT	BRI
SIZ		
SPA	(-0.003)	(-0.004)

SPA1	SAT	BRI
SIZ	0.015	0.009
SPA		

Figure 5.14: Interactive tool for interference prediction for v_1. Using our new predictive models, by simply modifying the size and spacing values for our methods we obtain the predicted interference strength and, through the derivatives, we also get information about what the effect of modifying the size and spacing parameters is. The method precedence to define v_1 and v_2 is given by the order {SAT,LIGHT,SIZ,SPA}. The higher the value the more interference v_2 causes to v_1.

siz1	spa1	siz2	spa2
7	3	2	4

$I(v_2|V_1)$

TWO-LAYER COMBINATIONS

		BOTTOM			
		SAT	BRI	SIZ	SPA
TOP	SAT		0.760	0.292	0.541
	BRI	0.663		0.373	0.321
	SIZ	0.035	0.074		
	SPA	0.356	0.210		

		BOTTOM			
		SIZ	SIZ C	SPA	SPA C
TOP	SIZ				0.189
	SIZ C			0.292	
	SPA		0.225		
	SPA C	0.221			

TWO-LAYER COMBINATIONS...DERIVATIVES

SIZ1		BOTTOM			
		SAT	BRI	SIZ	SPA
TOP	SAT		(-0.032)	0.000	0.000
	BRI	(-0.012)		(-0.017)	0.000
	SIZ	0.006	0.000		
	SPA	(-0.008)	0.000		

SPA1		BOTTOM			
		SAT	BRI	SIZ	SPA
TOP	SAT		(-0.082)	(-0.016)	(-0.137)
	BRI	(-0.034)		(-0.105)	(-0.023)
	SIZ	0.027	0.005		
	SPA	(-0.024)	(-0.007)		

SIZ2		BOTTOM			
		SAT	BRI	SIZ	SPA
TOP	SAT		(-0.142)		(-0.089)
	BRI	(-0.115)			0.000
	SIZ				
	SPA	(-0.024)	(-0.031)		

SPA2		BOTTOM			
		SAT	BRI	SIZ	SPA
TOP	SAT		0.011	(-0.004)	
	BRI	0.036		(-0.042)	
	SIZ	0.005	0.000		
	SPA				

SPA		BOTTOM			
		SIZ	SIZ C	SPA	SPA C
TOP	SIZ				(-0.028)
	SIZ C			(-0.055)	
	SPA		0.000		
	SPA C	(-0.008)			

SIZ		BOTTOM			
		SIZ	SIZ C	SPA	SPA C
TOP	SIZ				0.037
	SIZ C			0.029	
	SPA		(-0.030)		
	SPA C	(-0.033)			

ONE-LAYER COMBINATIONS

	SAT	BRI
SIZ	0.000	0.079
SPA	1.000	0.289

ONE-LAYER COMBINATIONS...DERIVATIVES

SIZ1	SAT	BRI
SIZ		
SPA	(-0.008)	(-0.004)

SPA1	SAT	BRI
SIZ	0.011	0.005
SPA		

Figure 5.15: Interactive tool for interference prediction for v_2. Using our new predictive models, by simply modifying the size and spacing values for our methods we obtain the predicted interference strength and, through the derivatives, we also get information about what the effect of modifying the size and spacing parameters is. The method precedence to define v_1 and v_2 is given by the order {SAT,LIGHT,SIZ,SPA}. The higher the value the more interference v_1 causes to v_2.

displays, while we include two-layer displays as part of our models.

It is interesting that they present orientation as an effective method to represent scalar fields. This is not surprising, since their evaluation was based on preattentive processing. We believe our initial results in this current experiment (see Fig. 5.9) clearly show the difficulty in reading orientation as a scalar magnitude in continuous 2D datasets. Expert designers alerted us of this issue before, and the recognition times further confirm it. Also, participants commented they relied on learned strategies to decide whether the orientation-mapped dataset was horizontal or vertical. They declared it was usually the first to be detected (preattentive), but in order to answer the experiment's question they needed to look at the borders of the display and figure out the direction the gradient was going. In this process they sometimes perceived the other method's gradient and sometimes did not, but their confidence about their performance was very low. The use of a strategy like the one explained defeats the purpose of the experiment, since non-linear dataset do not present any type of boundary indicators that would help understand the data variable in the center of the display. For all these we eliminated the orientation observations and conclude that orientation would create too much interference in the reading of other visual dimensions to make the overall display effective.

In utilizing our models interactively to predict the utility of different methods, we observed a lack of fit between expected results and the predictions for some extreme cases. In particular the predicted saliency values for the SAT-LIGHT combination when LIGHT is on top indicate a dominance of SAT even for low spacing values of the top layer. In general, the non-significance the normalized time differences show for this case in Fig. 5.11, would indicate a model for this case would not be effective. Indeed, that is the reason we indicate the non-significant cases in the interactive tool in Fig. 5.13.

The tools presented are a simple example to show the utility of the models we have developed. They should be used in combination and with a visualization goal in mind, i.e. having a clear idea of what data variables we want to highlight and be able to quickly understand in our final display. Once an acceptable visualization is reached, the derivative information helps steer the data exploration process through the changing requirements as new discoveries are made.

5.3 Chapter Summary

In this chapter we have demonstrated the use of perceptual experiments to evaluate the utility of several visualization methods, both in isolation and in pairs. We have matched the results from more limited studies that concentrated on target identification tasks, while augmenting those findings with a set of quantitative predictive models for a total of four different design factors and four different visualization methods.

The limited scope of our investigation allowed us to fully evaluate the multiple combinations of independent variables in both experiments, leading to a successful quantification and modeling of the experimental results. Our hope is that the methodologies utilized and lessons learned will facilitate the study of other visual dimensions and their perceptual capabilities, always oriented to their effective use in scientific visualization displays.

These two experiments followed another set of two that utilized expert visual designers to evaluate similar design factors for scientific visualization methods. Figure 5.16 shows ranking comparisons for all design factors evaluated during the second expert visual designer experiment. The visual bandwidth factor is compared to the perceptual interference factor here, dominance corresponds to the saliency factor, and time to read corresponds to an overall estimation of the recognition time in this experiment. Data resolution and spatial feature resolution are directly obtained from the models developed from the previous perceptual experiment.

We utilize relative rankings due to the high variance of the actual values obtained from the designers. They still provide a good visual intuition for the validity of our perceptual experiments' results and the confirmation that they more or less follow the subjective rankings obtained from the designers.

After this last experiment on saliency and perceptual interference, we have once again taken one more step towards accomplishing our vision. Figure 5.17 shows the new factors we are able to control thanks to our predictive models.

Note that the results from the first perceptual experiment allowed us to control the data resolution and spatial feature resolution of the individual methods, and the second experiment controls between-methods saliency and interference. Except for single-layer cases, we have no information on how the data legibility characteristics of the methods (DR, and SFR) are maintained when a second method is shown on the two-layer cases.

We can argue that this information is available for the single-layer cases, since we can obtain the legibility limits for a combination LIGHT-SIZ, for example, by measuring the DR and SFR values for LIGHT with low and high size values. These limits will be conservative worst-case predictions for the combination, since we cannot be sure were the different lightness values will lie with respect to the size values.

To try to constrain these limits to a more accurate range and define them for the two-layer cases, we have performed the next experiment. In it, we brought back expert visual designers to evaluate real two-valued datasets. Our hope was that their expert knowledge of the visual dimensions at hand would help us define how the legibility characteristics of the methods are maintained when combined in multivalued visualization displays. Obtaining that information would put us in a position to hypothesize how combinations of more than

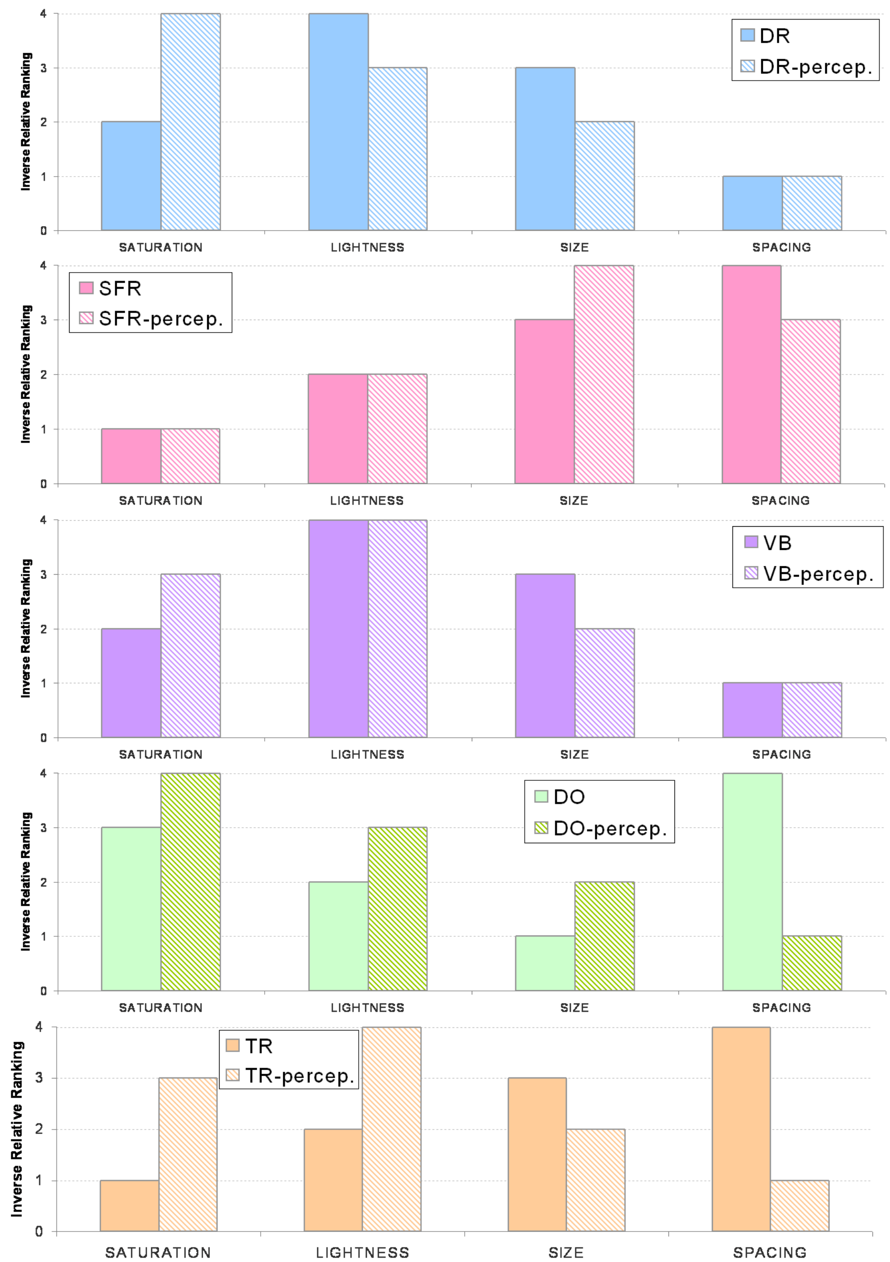


Figure 5.16: These graphs show a comparison of the relative rankings obtained from the expert designers experiment from Section 4.2 and the two perceptual experiments presented in this chapter. Only the 4 methods we evaluated in the latter experiments are shown here. Observe how, with few exceptions, the relative rank order for all design factors is approximately the same.

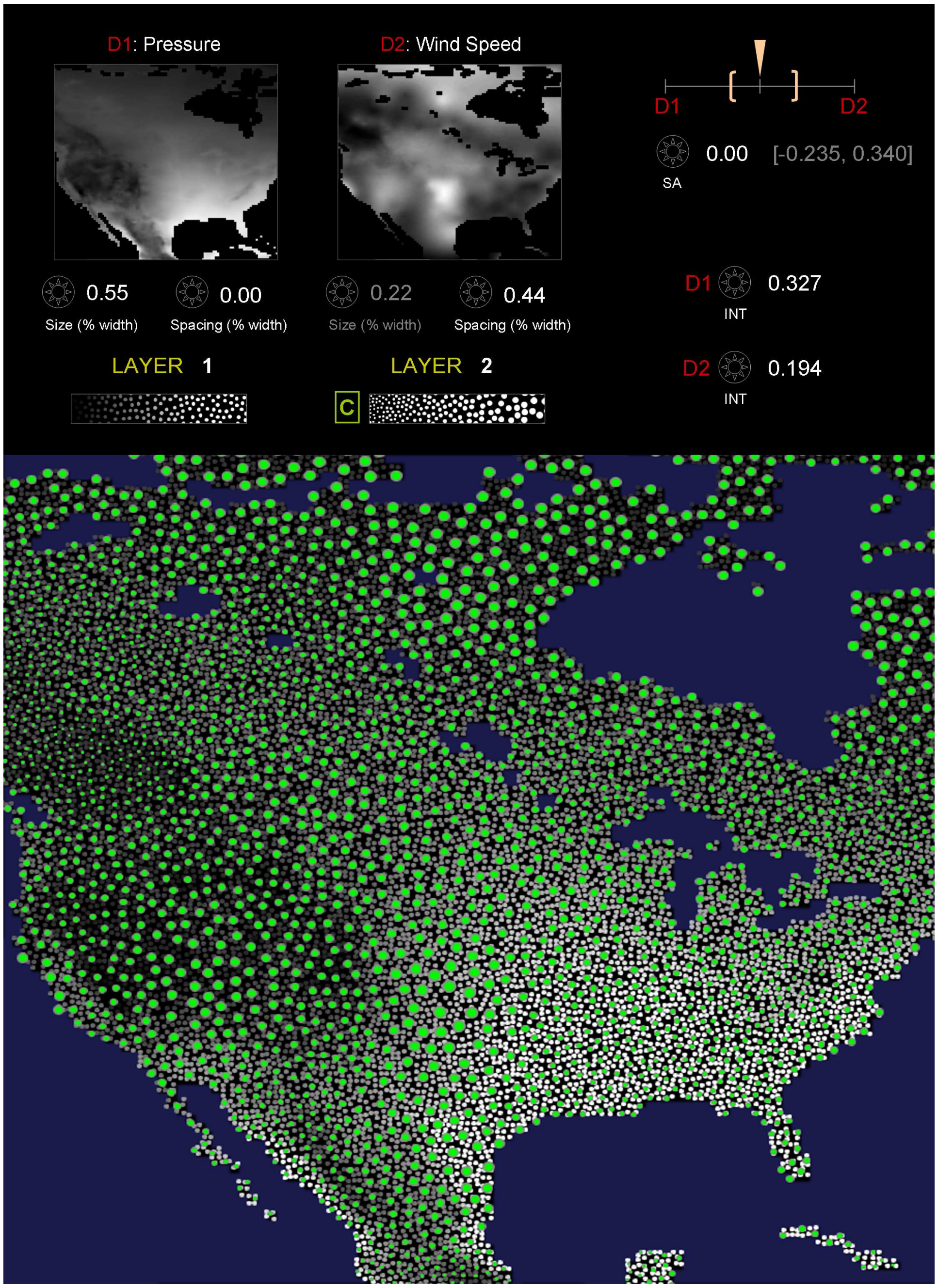


Figure 5.17: The new models obtained allow us to control the saliency and perceptual interference characteristics between two methods. Given two values from a dataset, a choice of method for each value, a choice of size and spacing parameters for the icons of each method, and the number and order of layers to be used, we can present users detailed information about both relative saliency between both methods and the interference that each method suffers from the other. Saliency limits are displayed based on the parameters utilized in our study. Interference values go from 0 to 1. The current choice is shown in the large display.

two methods would perform. Our initial vision from Fig. 1.1, so far limited to individual utilities and some pairwise perceptual interactions, would become a reality for high order combinations of methods.

Chapter 6

Evaluation of Multivalued Visualization Methods

In this chapter we describe the experiment we designed to measure how the legibility of our visual dimensions was improved or impeded when utilized for two-valued dataset visualizations.

In order to complete the utility model for our visualization methods we require one more piece of data: the legibility variations when two methods are combined. The legibility of a method describes whether the data it represents is perceived correctly or not. As we mentioned before, the previous study did not evaluate how well the characteristics of the datasets themselves were preserved when two of them were combined in the same display. In that study we quantified saliency, which describes the perceptual dominance of a method in the final composition, and interference, which pertains to the time it takes to perceive and understand the data.

To evaluate this legibility factor we brought expert visual designers back. Our hypothesis for this study was that, using their expert knowledge of perceptual interactions among our methods, they would be able to effectively explore the space and indicate how the different combinations of methods affected their individual legibility characteristics. Once more, we hoped to engage their experience-based intuition to quickly trim off portions of our search space that we would need to exhaustively test through a perceptual experiment.

This chapter describes first our study to evaluate this hypothesis. As we will explain, we could not disprove the null hypothesis in this case, leaving us in a position where we could not advance any more towards our grand vision for the research. We also report on an informal study to utilize the models we have already developed in a real situation.

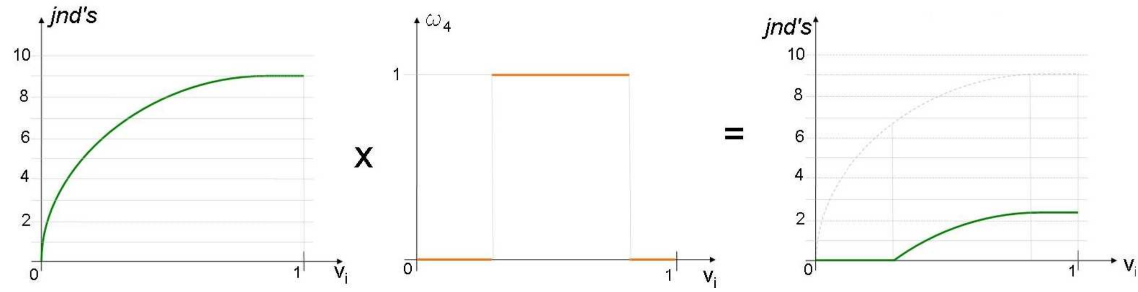


Figure 6.1: We expect our legibility models, $w_4(v_i|v_j)$, to be step functions that define the portions of the range of method v_i that are correctly readable ($w_4 = 1$) or not ($w_4 = 0$). Multiplying this function with, in this case, our data resolution model for the same method, $w_1(v_i)$, would result in the real data resolution model for method v_i when combined with method v_j.

6.1 Subjective Critiques of Two-valued Datasets Using Expert Visual Designers

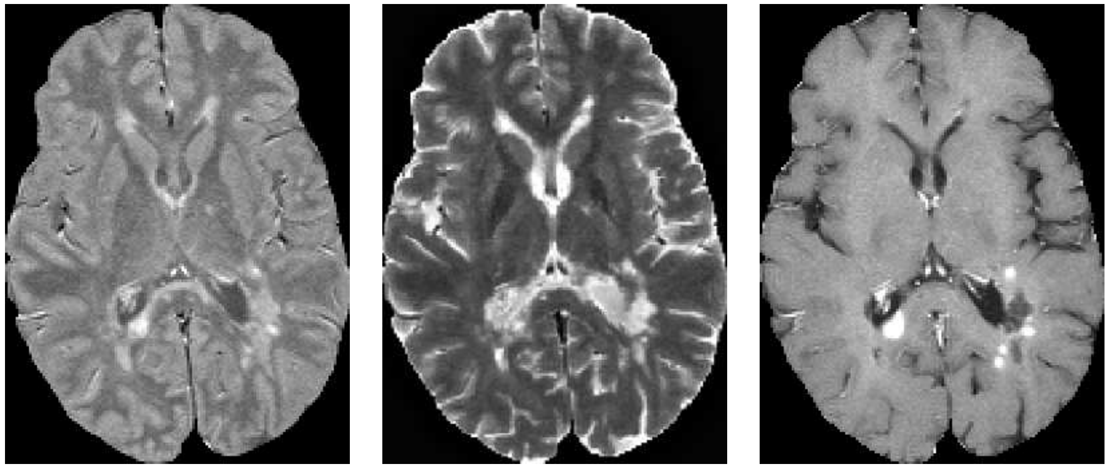
The goal of this experiment was to measure how the legibility of our visual dimensions was improved or impeded when utilized for two-valued dataset visualizations. We wanted to obtain the information necessary to model data legibility, thus becoming our 5th design factor, w_4, after spatial feature resolution (w_0), data resolution (w_1), saliency (w_2), and perceptual interference (w_3). In fact, our feature legibility definition utilizes two of those factors as the benchmarks. When we talk about preservation of data features we are interested in:

- How the different levels of data (related to data resolution, DR) are perceived in displays combining two methods.
- How different sized features (related to spatial feature resolution, SFR) are perceived in those situations.

Hence, legibility will be a factor that will multiply our DR and SFR models to provide guidance as to the loss of information when different methods are combined. As a simple example, Fig. 6.1 shows a case where the beginning and end portions of the range for method v_i are lost when this method is combined with v_j. The resulting DR model for v_i is then obtained by multiplying both functions. Note that the model for DR obtained in Section 5.1.2 is cumulative, so a loss at any subrange means a non-increase in the number of perceivable jnd's.

6.1.1 Methodology

The study presented our subjects with an interactive tool that allowed them to modify the same independent variables used in the previous experiment: icon size, spacing, number of



(1) (2) (3)

Figure 6.2: Three scalar values obtained from an MRI dataset of a human brain. We chose these three for the variety of spatial frequencies and different value spreads over the data range, as can be seen in the grayscale variations. All values are normalized to the full range of lightness. Data value (1) is dominated by a relatively narrow range of values (5 or 6 jnd's) from the top third of the full range (15 to 20 jnd's). It also has medium-to-high frequency information (80-100 cycles/width). Data value (2) has even higher frequency data (100-150 cycles/width) and utilizes the full gray level range of values more evenly (10 to 15 jnd's). Finally, data value (3) from the MRI dataset has a low number of value levels (4 or 5 jnd's) spread over the full range. It also has relatively low frequency information (10-20 cycles/width) mixed with points of high frequency highlights.

layers and their order, and icon color. The data used were three scalar magnetic resonance imaging (MRI) values from a human brain dataset, which were presented to them in pairs. The characteristics of the different MRI data variables covered a wide spectrum of high and low values, high and low frequency features, as well as partial and total value spread over the full ranges (see Fig. 6.2).

The setup, shown in Fig. 6.3, consisted of two monitors showing the original pair of data variables, the individual mappings onto one of the four visualization methods (saturation, lightness, size or spacing), and the final display combining both variables in one image. It also provided the basic interface controls to explore the space of possible combinations by modifying the independent variables explained before.

The images utilized the same parameter space as our last experiment, and the combination stimuli were the same size and presented on the same monitors as before. Our gamma correction functions were still valid for this experiment. Users had the ability to zoom the images on the left side monitor (see Fig. 6.3) to make them the same size as the combination display for easy comparison. At the zoomed-in state the perceptual characteristics of the visualization methods were the same as they were in our previous studies, covering approximately the same visual angle.

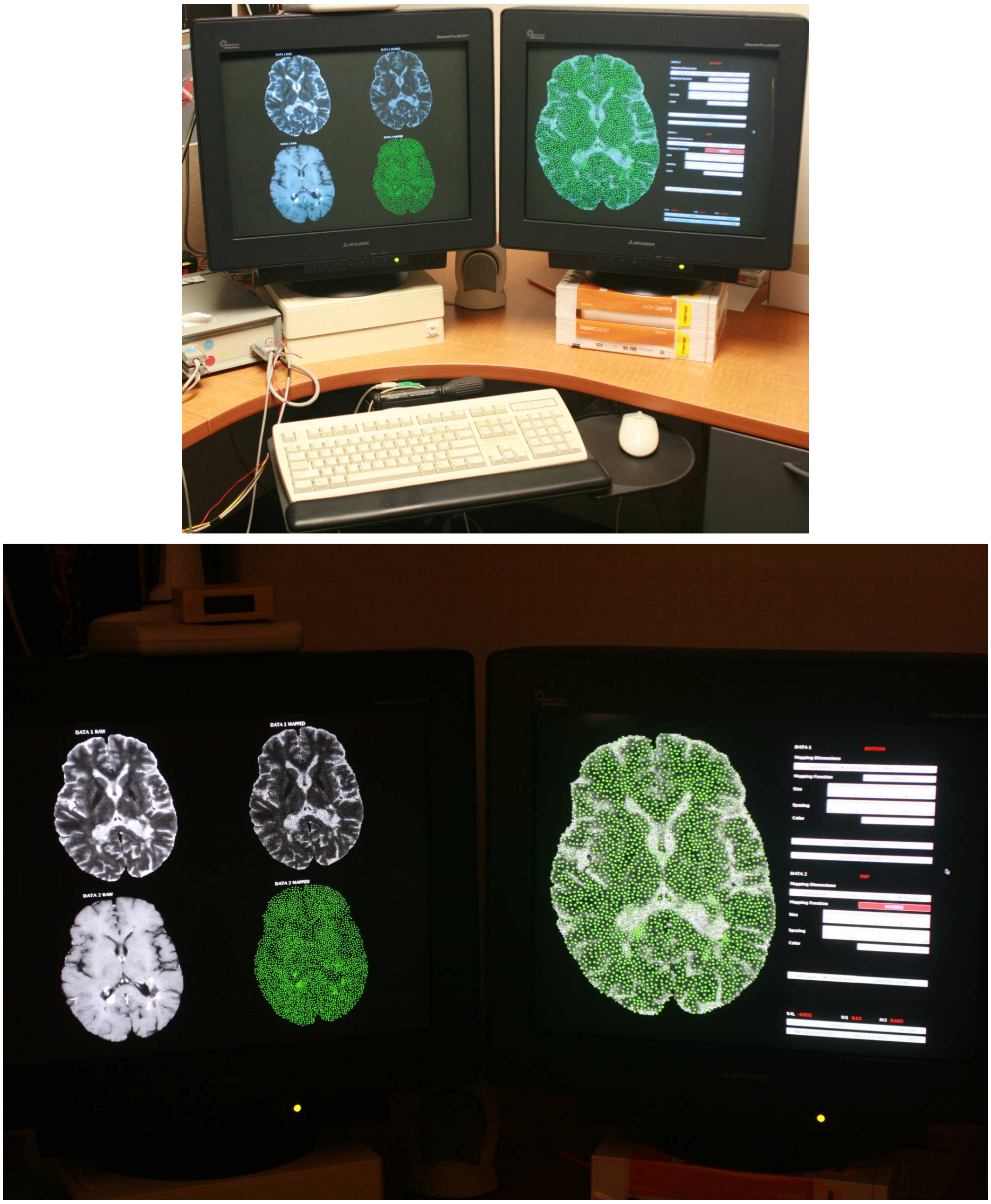


Figure 6.3: Experimental setup (top), and detail of the displays at the illumination level maintained during the experiment (bottom). The experiment took place in an illumination-controlled room where two side by side monitors displayed, from left to right, two of the original data values from the dataset shown in Fig. 6.2, the two individual mappings from each data value to a different visualization method, the combination display, and the user interface to control the visualization parameters.

Experimental Tasks and Protocol

We designed an experimental protocol that would allow us to obtain the legibility information we needed while providing expert visual designers with enough freedom to comment on the validity and effectiveness of the visualization methods presented. The basic protocol was:

1. We introduced to the visual problem and interface to the participants.
2. We choose a pair of methods to represent the data among the 6 possible pairs.
3. Task: Which data variable do you think users will perceive first and why?
4. Task: Comment on the types of features visible and invisible within the combination.
5. Task: How much legibility is gained or lost from the single-value to the two-value visualization?
6. Task: By only changing these factors: #layers, order of layers, color, size, and spacing, how would you modify this combination so:
   (a) Variable 1 dominates the composition.
   (b) Variable 2 dominates the composition.
   (c) Both variables are equally dominant.

   During the process you must try to maintain the legibility of as many features of the dataset as possible, or comment on how changing some of those factors affects the data legibility.
7. Task: For each solution provided on the previous point, switch the data variable involved and comment whether the same or a different solution applies. Modify your parameterization until you reach an appropriate solution.
8. Task: Pick a different pair of methods and go back to point 2 until all six pairs of visualization methods are explored.
9. Task: Freely design what you think would be the most appropriate visualization method for any combination of data variables and using any of the 4 visualization methods available.

We explained to the participants that our interest in data features was two-fold. First, we wanted them to comment how well low, medium and high values in the dataset were visible and understandable in the combination display. Second, we would also like them to comment how well low, medium and high frequency elements were preserved in the final combination display. This division of the full range of data values and frequencies was meant to facilitate the participant's task and our own data analysis.

Indeed, this protocol was designed so the level of freedom these experts are used to was preserved as much as possible, while we asked them to do something they are not used to do: think aloud as they try to solve a visual communication problem. In two pilot sessions we explored the possibility of not subdividing the ranges. This did not help participants, who were only partially identifying data feature issues without exhaustively exploring all combination options.

Throughout the study, we reminded participants about the exploratory goal of the combined display, i.e. that the design goal was to show as many features as possible from the original datasets, with no particular areas that were more important that others. We explained that the reason we asked them to help was that we realized the conversion from the original grayscale to the icon-based representation was lossy, hence we needed to explore the best ways to loose as little information as possible.

The questions on points 3 and 6 were meant to check the accuracy of the predictions from our previously generated statistical models for saliency and interference. Also, by asking them to design combinations for the three different situations from point 6, we hoped to gather derivative information that, although available from our models, would provide a glimpse into their design process. We could potentially analyze their paths through the design space and try to obtain some guidelines.

Finally, the order of the visualization methods shown and the pair of data values represented was randomized among the four participants to avoid order effects.

Participants and Data Gathering

A total of four expert visual designers ran through our experiment, with a maximum allowed time of 2 hours to avoid fatigue and lack of concentration.

For data gathering we videotaped the full sessions, paying close attention to how participants explored the different displays presented and what tools and techniques they used to answer our questions. As shown in Fig. 6.4, we developed a simple scorecard that once completed for each pair of visualization methods, would provide the necessary data to generate our legibility model. Each cell would contain a zero or a one depending on whether that particular portion of the data range or frequency range was correctly represented in the combined display. It would also contain any comments regarding how size, spacing, layer order, or color affected that value. We decided against providing them with the scorecards to fill out during the experiment since, from previous experience, we realized this would put

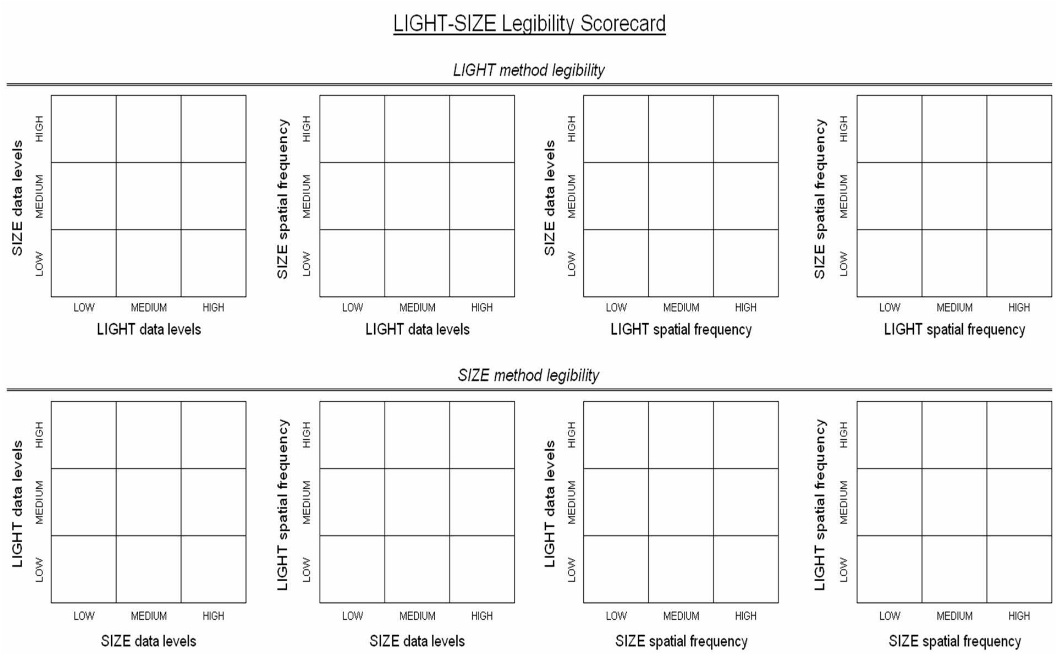


Figure 6.4: We create these scorecards for each pair of visualization methods. Studying the videotapes and looking at the resulting displays would allow us to fill out how DR information and SFR characteristics were preserved when methods were combined. Note that we have a full set of interactions represented: how data levels of one method affect the reading of the levels of another, how spatial frequencies affected the reading of data levels, and the corresponding opposite combinations.

experts in the uncomfortable position of assigning numbers to their knowledge and design process.

In these scorecards we hoped to capture some of the intuition expert designers bring with them to solve our particular visual problem. Note that the interactivity of the application would allow experts to tell us how modifying each parameter would affect legibility. This interactivity and verbal data gathering would substitute an exhaustive perceptual experiment and, using expert designers, allow for a much faster exploration of the vast number of possible combinations of parameters.

Study Motivation

Comments from our participants in the previous study and our own observations motivated the current experiment. The linear datasets used before were chosen because participants would intuitively understand what the real data was, hence, we assumed they always perceived the "real" data fully: i.e. a continuous and straight gradient across the display. This assumption allows us to apply the interference and saliency models to real datasets. However, we cannot assume this total legibility for datasets other than the linear one we used. In other words, when two visual dimensions interfere each other, the time to read those dimensions increases, but their legibility might not be affected at all (the opposite could also be true). Even though it might take time for users to shift and focus their attention on an interfered dimension, they will eventually get the same number of jnd's or perceive the same size features as in the single-valued case.

In fact, participants declared our assumption of full legibility already broke down for some cases in the previous experiment. Not only did methods using spacing cover part of the other methods, making it literally impossible to really see them fully, they declared that, for most combinations, they could probably have provided a different "time-to-perceive" value for different parts of the display. This signals a latent discontinuity of our models across the ranges of the methods they qualify, one that we have not captured so far.

Our first attempted solution was to adapt the previous experiment to allow for multiple time responses. It seemed logical to try and approach this legibility modeling through a perceptual experiment, given the success of the two last studies. It quickly became clear that the time commitment required for participants would be too high. Either we lifted the 10s time limit and let them explore all parts of the display freely, or we divided the range of our visual dimensions in pieces and evaluate each separately. Even dividing the ranges of our methods in three portions would increase the number of stimuli required to an unmanageable amount, requiring many more participants and longer experiment times, with the subsequent fatigue issues.

Similarly, lifting the timeout limit to avoid increasing the number of stimuli would not fare much better. The free exploration of each display would still require subjects to report times for all nine portions of the display. It is clear that the experimental interface and interaction technique would need to be carefully planned in order to really capture the participant's perception times. Otherwise, we ran the risk of obtaining subjective estimates of the order in which participants perceived each portion of the display.

To solve these problems we relied on the expertise of our visual designers.

6.1.2 Results and Discussion

All our participants completed different amounts of the full experimental plan, but none completed all tasks. The different situations encountered during the experiments, explained in the next subsections, limited our results but provided valuable clues to identify the reasons for this outcome.

All participants declared the interface was simple to learn and did not affect their decisions. As an initial step in the experiment, they spent around 10 minutes exploring the visualization space our parameters represented and getting used to modifying layer order, data mapping methods, and the constraints each one had.

The main results obtained from the experiment were a positive evaluation of our saliency and interference models, and the realization that the solution to a multivalued visualization design problem depends on the particular spatial distribution of the datasets involved and might not be safely generalized to other situations: i.e. with our experiment, we were not able to find a general model of legibility that could by applied to any pair of data values.

During the study, we had a hard time maintaining participants on-course through our protocol. The main reason for this was our inability to provide them with other visual dimensions to explore possible solutions. They constantly asked for options unavailable to them such as the use of icon hue, the use of different shape icons, a broader range of sizes and spacings, or non-linear mappings to emphasize low or high values (although they realized the scientific nature of the data does not permit this type of scaling, which would visually misrepresent data relationships and potentially lead to the wrong conclusions). Our experiment was constrained by the number of methods we had been exploring in other experiments, along with the limited parameterizations we utilized. Although this created a manageable space to explore during our previous studies, visual designers are not used to this kind of constraints.

Variable Saliency Tasks

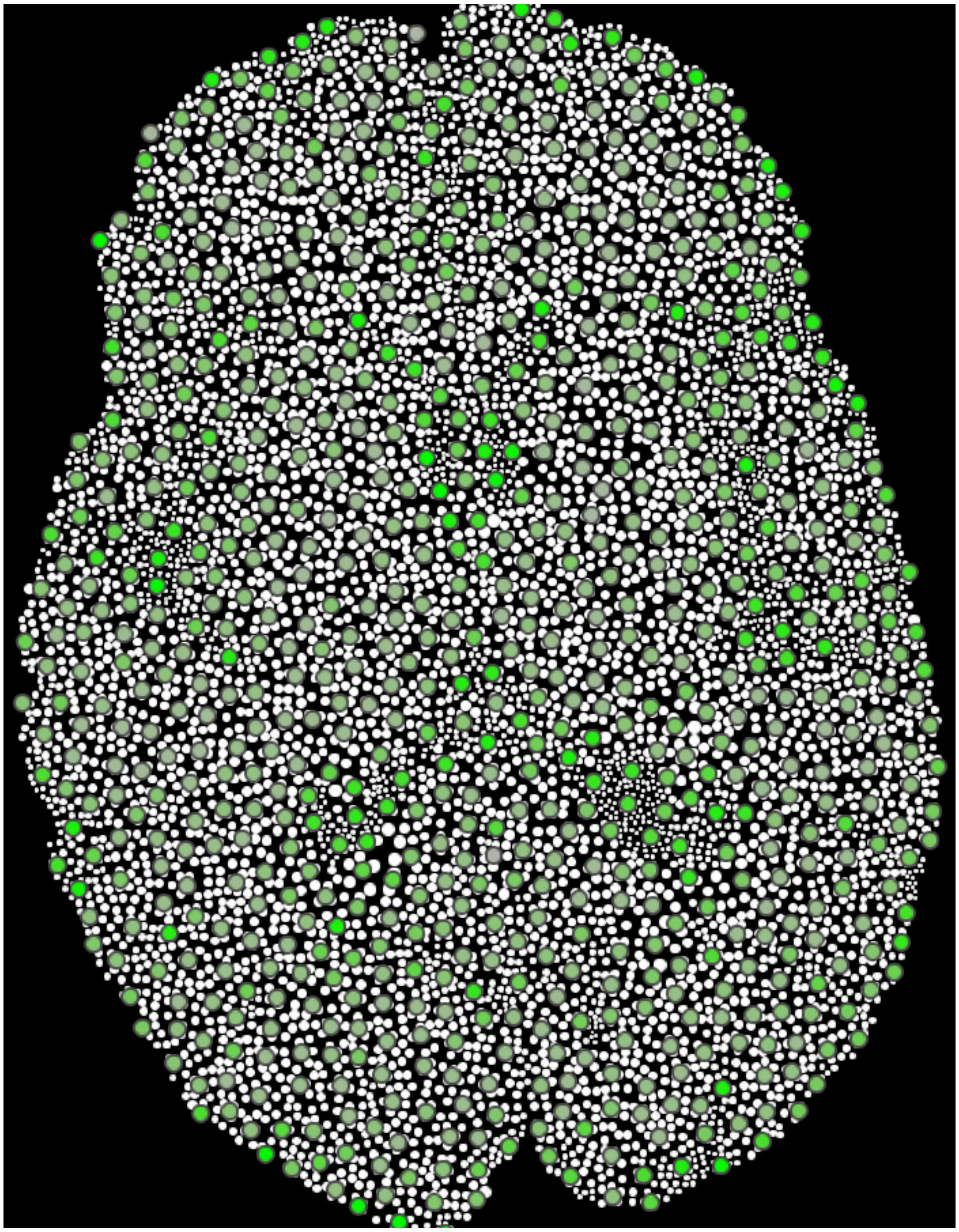
In general all experts validated the predictions from our models for saliency and interference. The user interface included in it the values predicted for those two factors, and although participants were not informed what the figures indicated, we were able to check them against their comments as they were exploring the space of visualizations. Figures 6.5 through 6.8 show examples of some of the solutions participants provided to this task, including our models' predictions.

It is important to note that this task, as opposed to the previous experiment, includes an element of legibility. Participants were asked to keep both variables legible as much as possible while highlighting one or the other, which accounts for the small deviations observed in the values for saliency and interference shown in the figures. The relative values within a given combination were correctly predicted. For example, Fig. 6.7 is an equal saliency solution and, while the actual value predicted is not zero, the legibility constraint made designers increase the size from 2 to 4 pixels to make saturation readable. Our model does not take into account legibility due to the simple dataset used for the previous experiment and it predicts size 2 to produce a saliency value of -0.009, the closest to zero in the range tested. During the previous experiment, participants were still able to recognize saturation changes with that small size, but the continuity and smoothness of the data might have helped in that regard. These legibility limitations are what we are trying to capture with this experiment.

In particular, notice how the bottom layer is usually defined to be as solid as possible. Indeed, this solution solves the visual problem we presented to participants: to display "two" data values simultaneously. House [House et al., 2006] reached a similar set of solutions for the visualization of two overlaid textured surfaces. Although not using the same stimuli (they used stereo animated images and three dimensional surfaces) or methodology (they reached these solutions using a genetic algorithm approach to explore large areas of the visualization space), our experts confirmed their results, while our predictive models also reached a similar outcome.

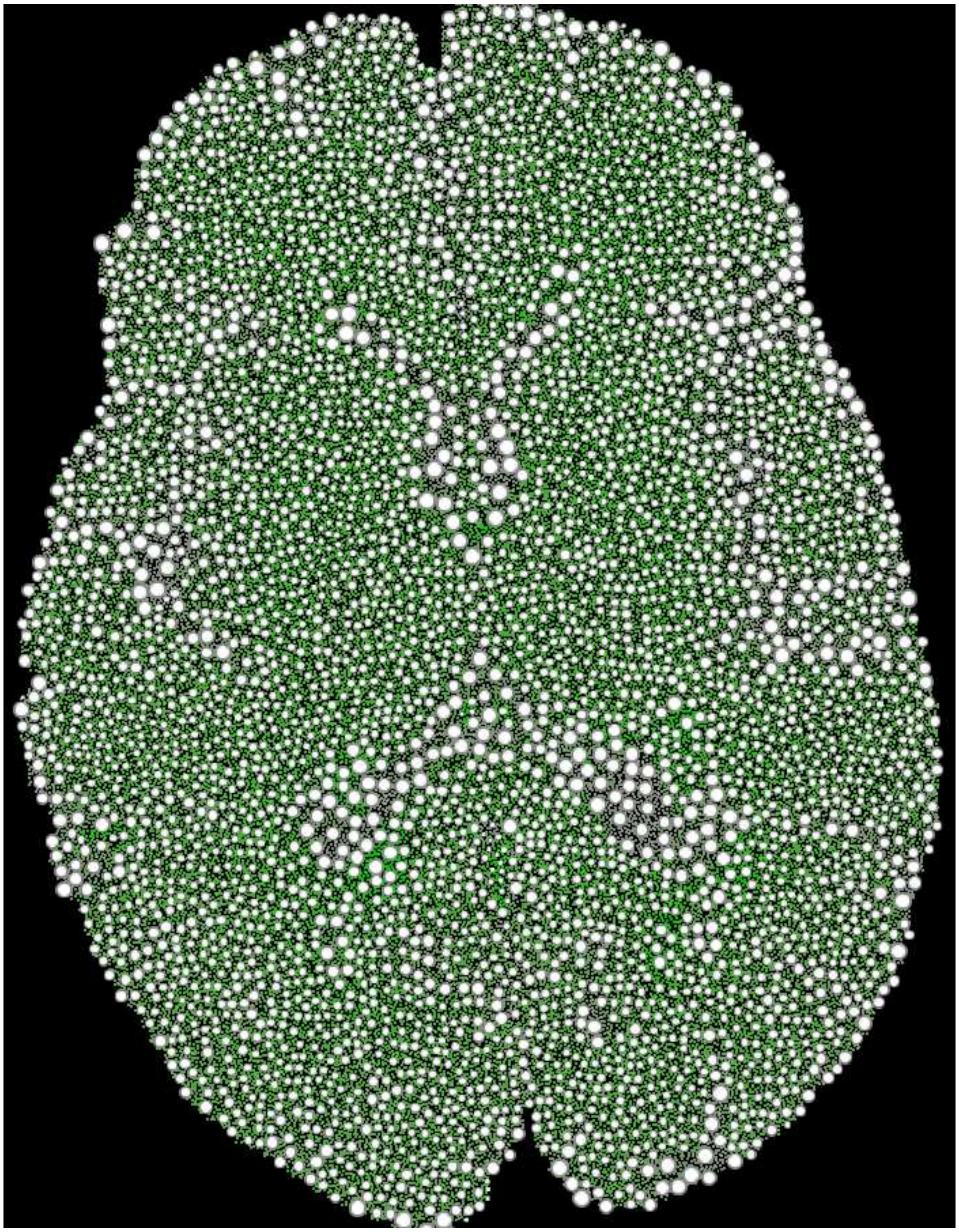
The problem this created was that only a small portion of our full parameter space was explored. Furthermore, solutions of the type reached during the experiment do not provide valid information towards creating multilayered visualization methods for more than two values. Methods containing a sparse bottom layer did not get evaluated for either saliency variability or legibility. Similarly, the results from [House et al., 2006] cannot be applied to the visualization of three overlaid surfaces.

Our initial reaction to this was to change the experimental design to present three scalar data values simultaneously displayed using three separate layers. The hypothesis would be



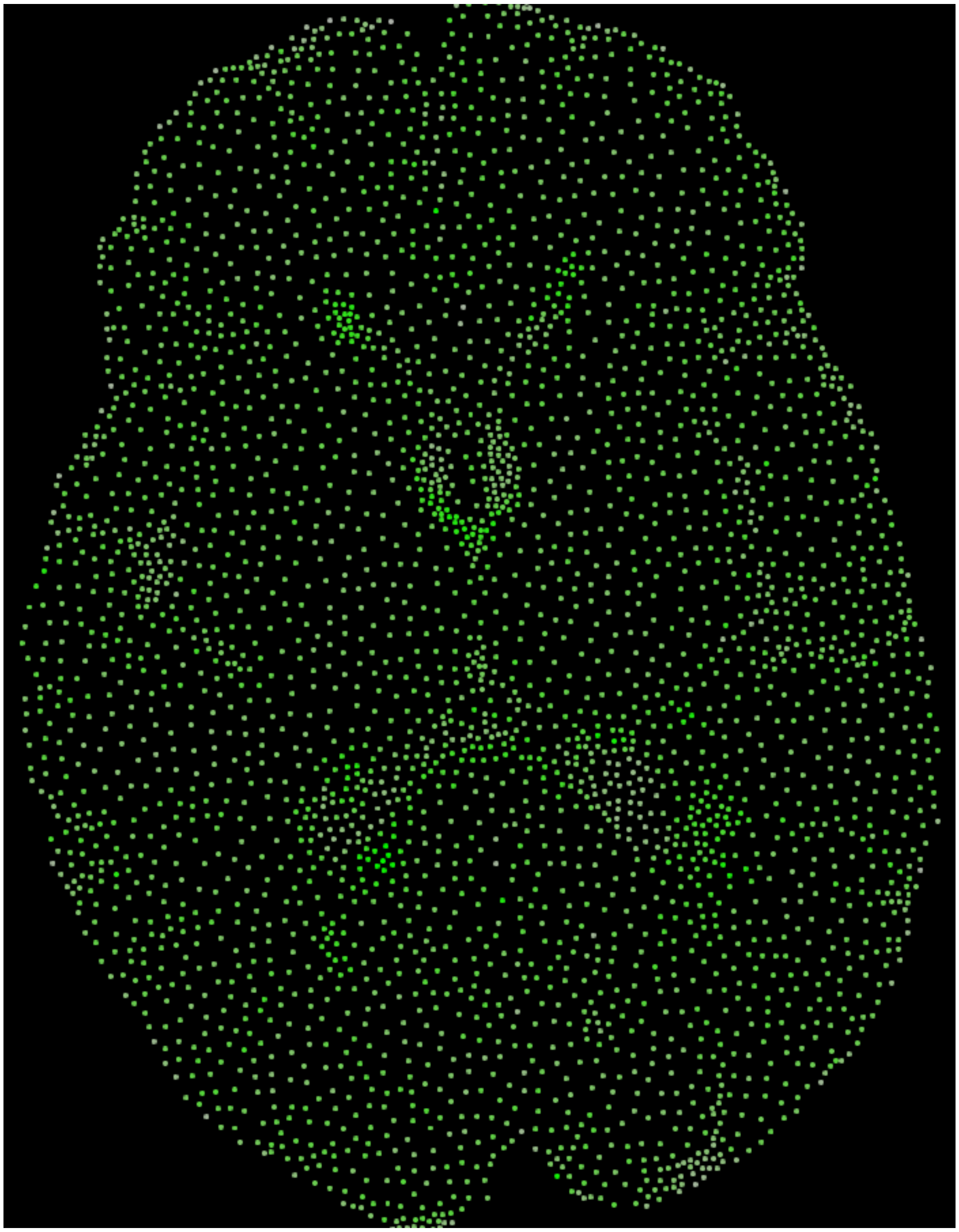
$$v = \{l_0, l_1\} = \{\{(0,0,0,0,0,3,0),(0,0,1,0,0,(2,10),0)\},$$
$$\{(0,1,0,0,0,0,0),(0.33,(0,1),0.6,0,0,8,7.5)\}\}$$

Figure 6.5: A result from the equally salient task for SAT and SIZ. This combination has a predicted saliency value of $w_2(v_2, v_5) = -0.183$ with the available range being $w_2 \in [-0.698, 0.141]$, given the interface options. While our prediction indicates a slight dominance of SAT, this compensates for the task's goal of maintaining legibility for both data values, which our predictive model does not take into account. Interference values are $w_3 = (v_2|v_5) = 0.342$ for how much SIZ interferes the reading of SAT, and $w_3 = (v_5|v_2) = 0.405$ for the opposite case. These values being so close confirm this combination as an equal saliency choice. Data value id's in the parameterization come from Figure 6.2.



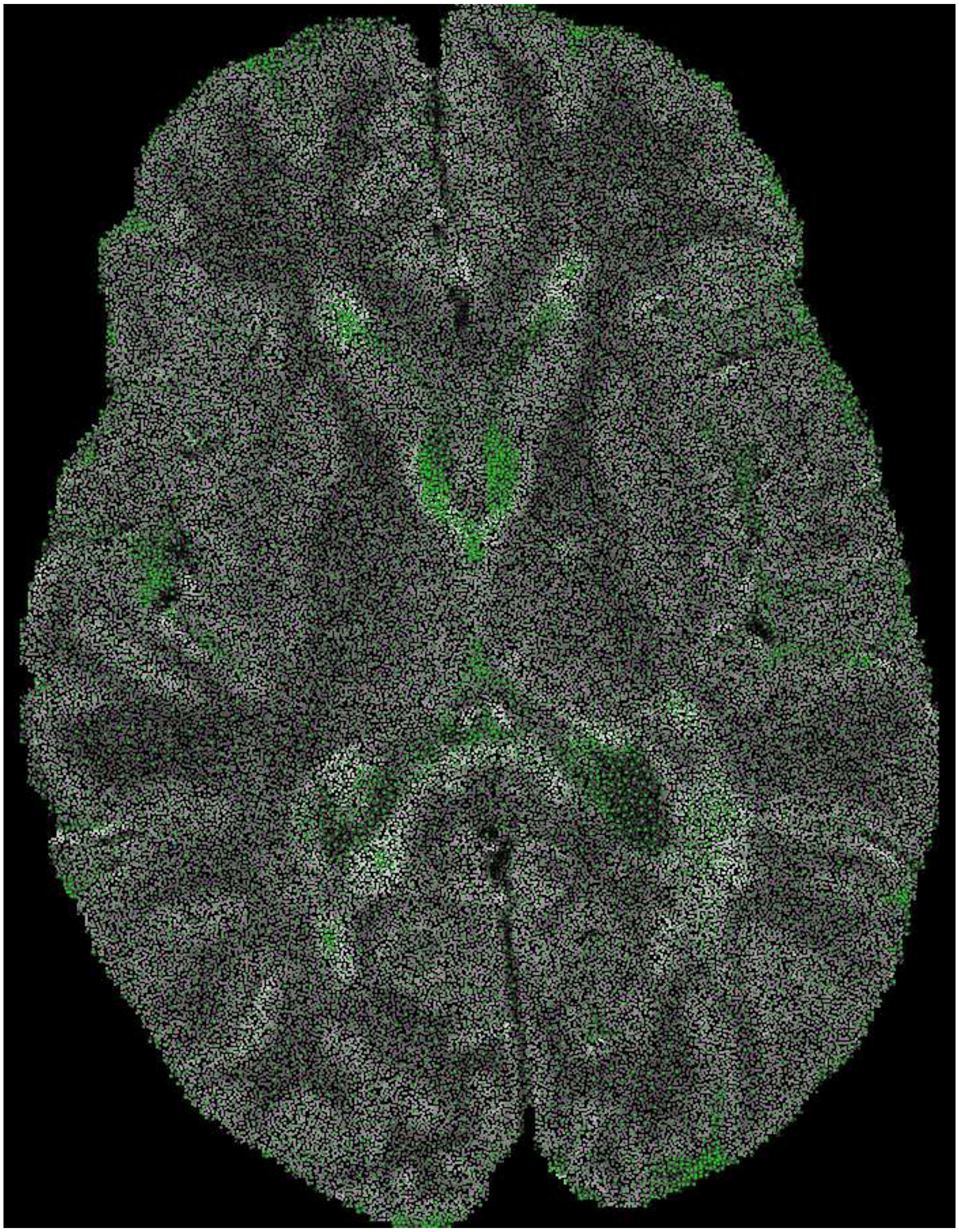
$$v = \{l_0, l_1\} = \{\{(0,3,0,0,0,0,0),(0.33,(0,1),0.6,0,0,2,0)\},$$
$$\{(0,0,0,0,0,1,0),(0,0,1,0,0,(2,10),2.5)\}\}$$

Figure 6.6: A result from the task to highlight one of the variables (SIZ) for a SAT-SIZ combination. In this case, the predicted saliency is $w_2(v_2, v_5) = 0.442 \in [-0.121, 0.714]$. Again, although our model would accommodate a more salient solution for SIZ (a fully opaque layer with 0 distance between icons), designers compromised to allow for some legibility of SAT. The interference values clearly reflect this dominance. While SIZ interferes with SAT quite significantly, $w_3(v_2|v_5) = 0.858$, the opposite is not the case, $w_3(v_5|v_2) = 0.120$. Again, data value id's in the parameterization come from Figure 6.2.



$v = \{(0, 2, 0, 0, 0, 0, 1), (0.33, (0, 1), 0.6, 0, 0, 4, (0, 10))\}$

Figure 6.7: A result from the equally salient task for SAT and SPA using a single layer. For this combination, our model predicts a small range of saliencies available given the interface options: $w_2(v_2, v_6) = -0.177 \in [-0.249, -0.009]$. This indicates a very unstable combination with a tendency for SAT saliency. In this case, experts were clearly limited by the available options, since they could only control the size of the icons. Again, enabling some legibility for SAT threw off our predictions.



$$v = \{l_0, l_1\} = \{\{(0,0,2,0,0,0,0),(0,0,(0,1),0,0,2,0)\},$$
$$\{(0,0,0,0,0,0,1),(0.33,1,0.6,0,0,2,(0,10))\}\}$$

Figure 6.8: A result from the task to highlight one of the variables (LIGHT) for a LIGHT-SPA combination. In this case, experts went to a completely opaque bottom layer that would be salient over a sparse, yet readable, top layer. Our predicted saliency value confirms this: $w_2(v_3, v_6) = -0.127 \in [-0.127, 0.008]$, and interference results indicate SPA does not interfere much with the reading of LIGHT, $w_3 = (v_3|v_6) = 0.007$, while the opaque layer does not compromise the reading of SPA either, $w_3 = (v_6|v_3) = 0.234$.

that the bottom layer would serve as this solid (although still icon-based) background and the other two would provide us the two-layer interaction information we were looking for. As we will explain below, the issue with the dataset dependence impeded our pursuit of this new methodology.

Derivative Information

In designing this experiment, we hoped to identify one or two good combinations per pair of methods and record how moving around those would affect saliency and legibility, but the participants' strategies did not fit our expectations. After they had performed the three subtasks from point 6 of our protocol (modifying the parameters of the visualizations so either variable 1, 2, or both were salient in the combinations display) we were not able to evaluate our models' derivative information, since we could not identify any particular paths of exploration. The process they followed was not what we expected: After almost exhaustively exploring all options from the interface, they reached a solution for the first subtask. They then went back to a neutral combination of parameters and explored the space of parameters again searching for a solution to the second subtask, and so on.

Queried about the reason they were doing this they declared that, although the visual dimensions were familiar, and the datasets were clear enough, their joint use in the combination image was not at all common, making them explore the full space every time to be confident of their response.

Legibility Information

The main goal for this study was to obtain a reading of the intuition designers have, from their experience and design knowledge, as to how the combination of methods work together to represent data effectively. As mentioned before, we had a difficult time keeping participants on-course through our protocol and, for this reason, their comments about legibility were quite sparse and not at all exhaustive enough to complete our scorecards.

Most of the comments regarding our legibility measure were constrained by the particular characteristics of the dataset utilized. The remarkable spatial similarity of the three scalar variables provided many examples of the various situations indicated in our scorecards, but participants commented they had a very difficult time answering the very specific questions from tasks 4 and 5. They acknowledge it is a difficult problem to solve, but without more tools to explore a solution, such as fine-tuning parts of the display by hand, they could not reliably provide general comments valid for other situations. This leads us to the next and perhaps more important result from the experiment.

Dataset Spatial Distribution Dependence

All our participants designed individual solutions based on the particular pair of data variables being displayed. We had hoped that they would try to obtain good overall combinations of parameters that would apply to any pair of variables, but that was not the case.

As it can be seen from Fig. 6.2, since all three MRI scalar values come from the same human brain dataset, they all share similar physical characteristics. This, we hoped, created the ideal setting for exploring how data resolution and spatial feature resolution characteristics were preserved in two-layer combinations.

Participants declared, in fact, these coincidences distracted them too much from trying to answer the tasks. They understood the setting and why this was the case, but adapting to it in such short time period and with the very constrained parameter space they had to work with, did not provide an adequate problem-solving situation. Furthermore, the exploratory goal of our combination display did not help in their design process.

Even though they realized the goal was not to highlight any particular features from the dataset, the many spatial coincidences from the variables made it hard to balance the solutions. While some areas were fulfilling the task and making one variable more salient than another, other areas would show the opposite effect. Once again, the limitations of the interface and parameter space did not let them perform local adjustments that could have potentially solved those issues.

We enquired whether datasets such as the weather data we have used for Fig. 5.17 would have helped, since the spatial distribution of values is different among variables of that dataset. This is where they all agreed on mentioning our main take-home message from the experiment: No matter what data we used, they would explore the space of possibilities and design a visualization that fit the particular combination of data variables. They understood our goal of extracting a general set of rules from their design process, but they all asked for a full set of design tools and an unconstrained space to really show that process.

Two Explanations for the Results

This last comment creates a dilemma for the evaluation of our experimental hypothesis. On the one hand this data dependence issue would make it impossible to generate a general set of legibility scores independent from some measure of spatial data correlation. This, in fact, would indicate that the wrong hypothesis was tested, and a new hypothesis, which would include those data correlations, would need to be posed and evaluated.

On the other hand, the solutions participants reached for the saliency tasks, where an almost solid layer of icons was permanently placed on the bottom layer, along with the limitations our parameter space imposed on their design creativity, would indicate our experiment did not have the capability of finding a significant result for our hypothesis, even though one might have existed. A redesign of the experimental protocol and setup would be granted.

After analyzing all the comments and solutions from our expert visual designers, we believe both conditions are true: the current hypothesis cannot be evaluated and the experimental methodology was not powerful enough to detect a significant result. The next section explores these two issues.

6.1.3 Attempting a New Approach

Our methodology for the experiment was based on our exploration of this small portion of the visualization design space, which contained only four visual dimensions. This, we hypothesized, would make it easier to explore both perceptually and with the help of expert designers. Our results indicate that that very limitation impeded the effective use of visual design experts.

This result has similar connotations to the one from our vector field visualization experiment, where it seemed we under-taxed our experts. In this case, we had a complex problem for them, but the experimental setup did not allow them to solve that problem effectively. We, again, under-taxed their expertise by not providing them the right tools.

Following that conclusion, we could think of approaching our exploration from the top down. That is, we could present the visual problem to the designers and let them generate solutions for us using their own tools. We could still record their comments and annotate their process as much as possible to extract the necessary pieces of data to inform our model. This obviously creates the problem of generating a powerful enough protocol so we could obtain valid data from the potentially huge space designers could explore. Also, trying to extract information about our particular dimensions of interest would be biased by many factors, mostly different for each subject, that we could not fully control: interfaces, user interactions, or combinations of visual dimensions used, to name a few. This approach assumes our hypothesis can in fact be evaluated with a more complex experimental methodology.

Note that this is the traditional way in which visual designers collaborate with visualization scientists. Since our goal from the beginning was to generate a quantitative model of utility, our approach was necessarily bottom-up, hoping to generalize to complex situations what we could learn from simpler quantifiable ones.

On the other hand, if we conclude that we cannot disprove our current null hypothesis, then we believe including the dataset as part of the model could be part of a new experimental hypothesis. We will explain in the next chapter our ideas on how to do this. The main issue with this solution is that, a successful modeling of this combined space (visual dimensions plus data characteristics) could lead to visualization solutions that would be different for the same types of data. For example, using the weather data as our multivalued dataset, this new model could suggest an optimal visualization method combination for the mean temperature and average wind speed data variable combination, based on the particular spatial relationships between both variables. Now, if we changed the dataset to be that of another region of the globe, those relationships would most likely change, potentially leading to a different choice of visualization method. This is precisely the issue with this approach: even when the same type of variables are used (temperature and wind speed), a meteorologist looking at these data would be given a different visualization for each part of the world, or even for each time of day when conditions change in the same region.

We will further elaborate on these possible scenarios in our conclusion chapter, when we will put these options in the context of our overall research.

6.1.4 Experiment Conclusions

This experiment has provided valuable confirmation of our previous experimental results and important information about the methodology to use when engaging expert visual designers in visualization evaluations such as this. In light of our results, we could consider this experiment a pilot study that evaluated a potentially effective methodology to evaluate legibility factors in multivalued scientific visualization methods.

It has offered us a glimpse into the visualization design process, albeit biased by our limited exploration space. This type of exploration required visual design experts to almost exhaustively search the space every time a new set of variables needed to be displayed. The limited design space also generated frustration among our participants and dampened their effectiveness performing our tasks. In summary, we should not limit their design creativity if we expect to extract good data from examining their design process.

Finally, the protocol design assumed also that questioning the participants and guiding their exploration would help obtain consistent data. Although that is the aim of experimental protocol analysis techniques, the protocol itself broke the concentration of our participants, hence not really performing as they normally would. We believe the visual design field is a very difficult area to explore and to try to quantify. That said, successful protocols have been developed to explore, for example, users' ability to control complex user interface designs, but we estimate it would require years to develop a deep enough protocol to put some mathematical structure around our visualization design process.

6.2 Evaluation on Practical Applications

We present here the results of an informal evaluation of our models and multilayered visualization methods. In it, researchers from the Department of Geology at Brown University utilized our visualization methods to explore data collected by the 2001 Mars Odyssey Gamma Ray Spectrometer (GRS), which included concentrations of hydrogen (H), chlorine (Cl), silicon (Si), potassium (K), iron (Fe), thorium (Th), and a potassium-thorium ratio ($KvsTh$) from the near-surface of the planet. Figure 6.9 shows these data using grayscale values on a simple cylindrical projection of the surface of Mars (from -180 to 180 degrees of latitude horizontally, and from -60 to 60 degrees of longitude vertically), cropped at high latitude values where data collection is not reliable. The same figure also shows the estimated standard deviation values (σ) for all seven data values. Since only qualitative relationships are sought, no numerical values are attached to these images or any visualizations shown here.

GRS is a significant component of NASA's Mars Surveyor Program. It is an ongoing initiative to explore Mars through scientific instruments aboard orbiters, landers, and rovers. GRS is really a suite of three instruments designed to analyze the chemical composition of the Martian surface. GRS also has the capability of detecting water in shallow subsurface depths.

The goal of visualizing these elements together was to confirm whether strong correlations exist among the different values, in particular with hydrogen, which is used as an indicator for the existence of water on or below the planet's surface. Current state-of-the-art methods to explore these correlations rely on 2D scatter plots that show concentration correlations without any spatial information attached (see Fig. 6.10 (a)) or, at the most, displays combining two elements as shown in Fig. 6.10 (b). Apart from these, side by side comparisons are used and, sometimes, RGB combined images displaying three elements simultaneously and focusing on finding areas with pure red, green, blue, white or black colors, since any other intermediate combinations would be difficult to interpret.

Our goal during this informal session was to determine whether providing users with explicit control of the perceptual relationships among the different methods would overcome the loss in spatial feature resolution. This loss comes from the discretization of the data into icons. However, the main benefit of this discretization is the possibility of layering icons and showing more data values simultaneously. Two researchers attended the session in which we explored different combinations of data values.

The first display we generated is shown in Fig. 6.11. It shows the concentration of H

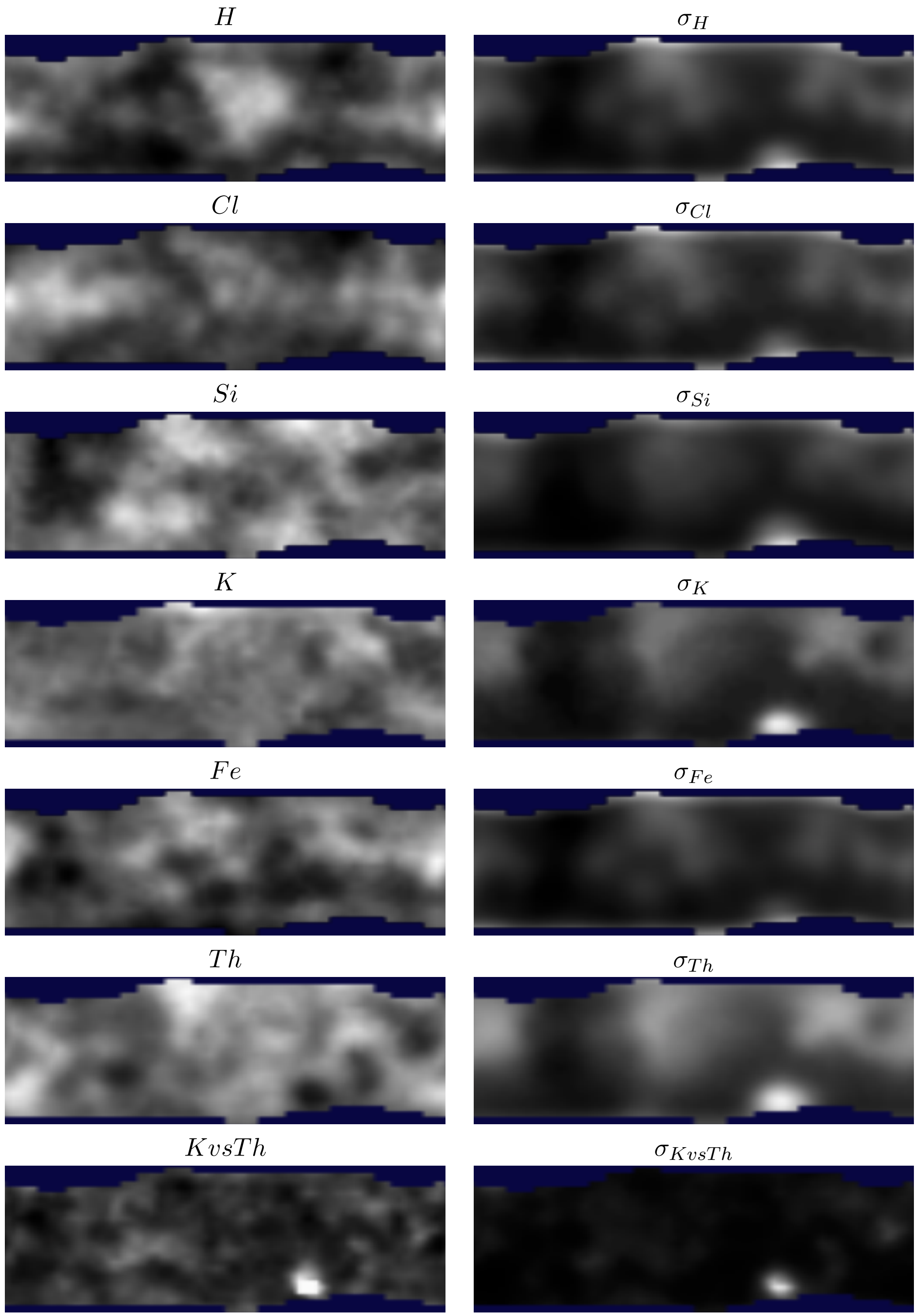
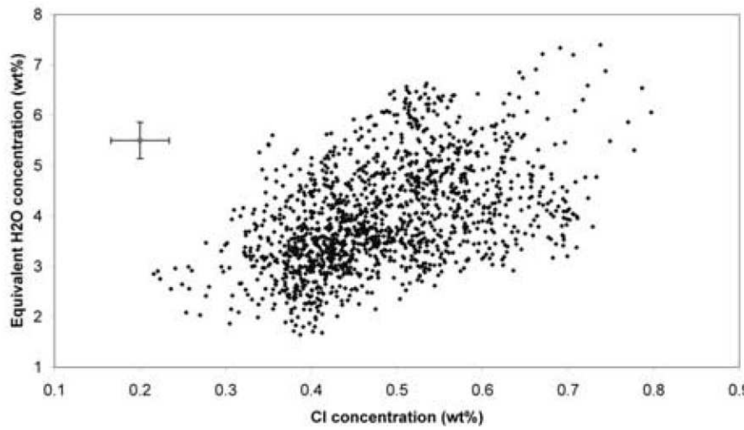
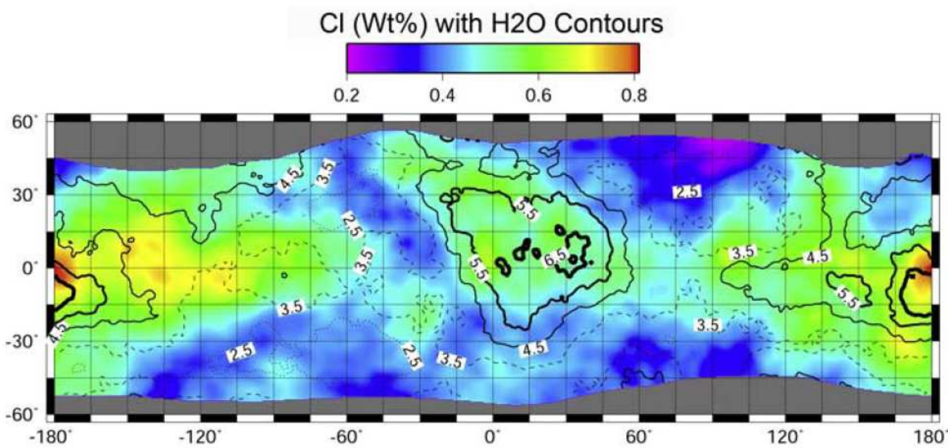


Figure 6.9: Concentrations and related measurement errors of various elements on the surface of Mars as captured by a gamma ray spectrometer. Low values are black and high values white.



a)



b)

Figure 6.10: (a) Correlation scatter plot of Cl and H concentrations on the surface of Mars. This plot lacks spatial information, and is sometimes complemented by (b) using a combination of a non-perceptually-uniform rainbow scale for the Cl concentrations and contour lines for the H concentrations. Both images from [Keller et al., 2006].

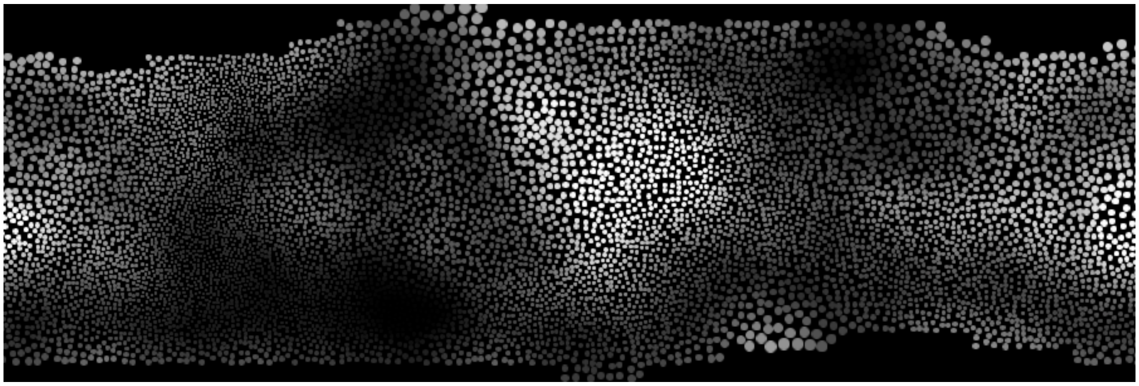


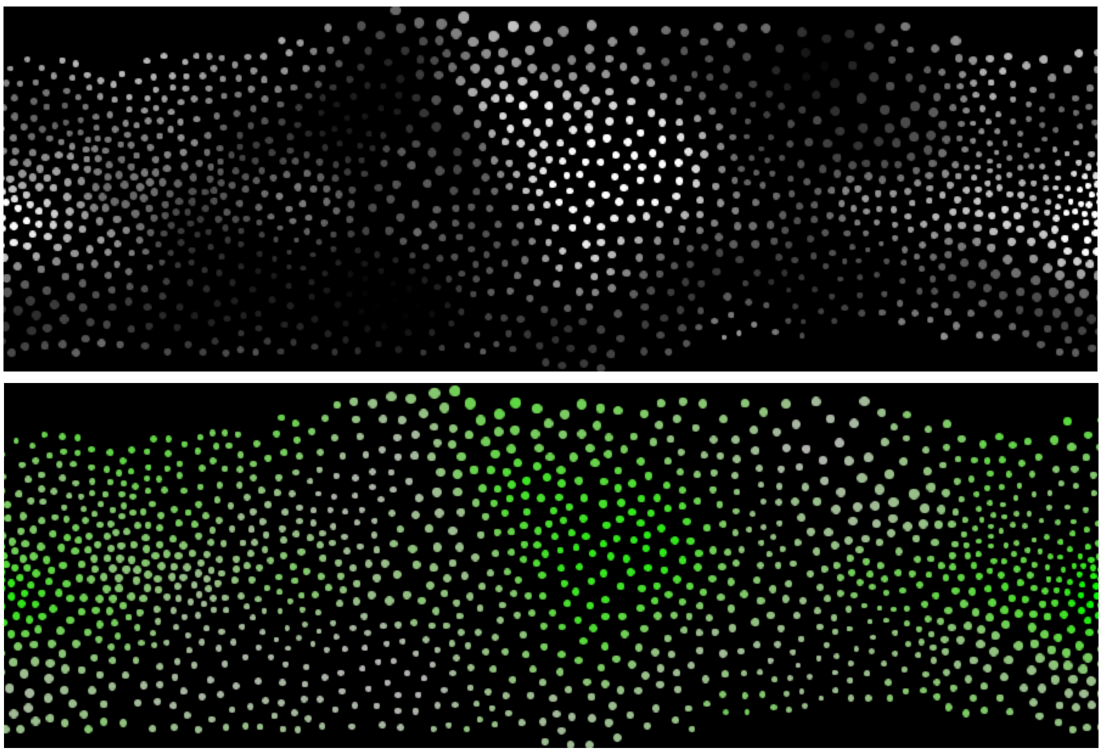
Figure 6.11: Concentration of H mapped to lightness and its corresponding σ_H mapped to size. Observe how is mostly at the higher latitudes where data is less reliable.

as lightness changes and its estimated σ as size changes. Given the resolution of the given data, we kept the spacing to a minimum so we would loose as little information as possible (both spatial feature resolution and data resolution of lightness decrease as spacing values increase). Also, since the actual concentration values should dominate the composition but not interfere with the reading of the variance values, we chose this combination of lightness and size. Our models predict lightness as more salient and interference to be low between both methods.

This initial display was regarded as effective by both our users. They declared the loss of resolution was a problem, but recognized the simplicity and immediate reading of both variables that this icon-based visualization created.

The next step was to include some of the other concentration values in the display. Figure 6.12 shows two visualization of the same three elements (H, Cl, and K). The first display (Fig. 6.12 (a)) was based on our initial combination, so we kept H as lightness and included Cl as spacing changes and K as size changes. These choices came from the better data resolution performance of spacing versus size, and the relatively low data resolution of the K concentration values. With this representation we kept the dominance of the H component but our models predict a relatively high interference with the spacing changes. This was recognized by our users and we switched the visualization to the one shown in Fig. 6.12 (b). The only change here was to switch the H data values to saturation changes. This, our models predict, lowers the saliency of that data value, but also decreases the interference with the other methods.

At this point both users agreed this was a completely novel and effective way of looking at their data. They were able to recognize topographical features of the surface of Mars based on our visualizations. It is important to note that they were only seeing these resulting visualization and never the raw data or any other image that could serve as a reference.



a)

b)

Figure 6.12: Two visualizations of the same three elements. (a) Shows H as lightness, Cl as spacing, and K as size changes. In (b) we changed H to saturation to decrease its dominance of the composition and the interference with the other two data values, as predicted by our models.

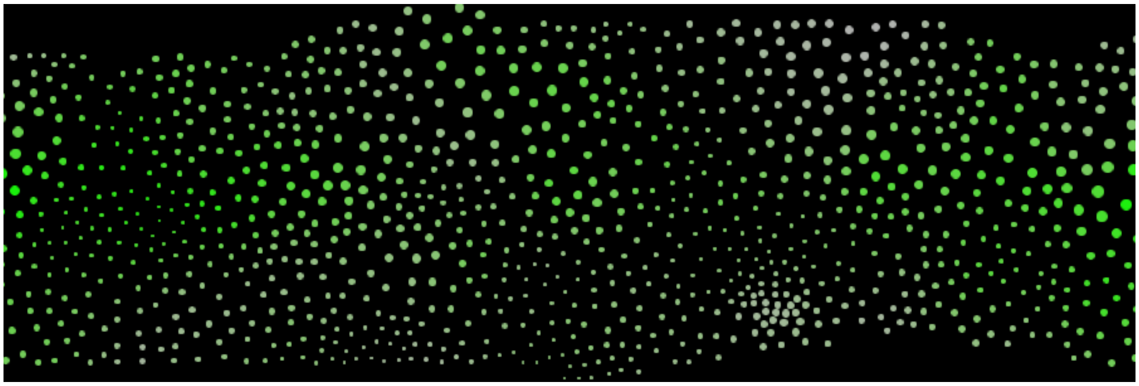


Figure 6.13: Concentration of Cl mapped to saturation, Fe mapped to size, and the ratio $KvsTh$ mapped to spacing.

They both had extensive experience looking at these data in other formats. They quickly became comfortable reading the icon characteristics as data values and were able to confirm expected areas of correlation and non-correlation among the different variables.

In order to bring in some of the other data values, we created the display shown in Fig. 6.13. This uses the same parameterizations as Fig. 6.12 (b) but using Cl, Fe, and $KvsTh$ as the data values being represented. In this case, spacing is used for the data value with lower data resolution requirements ($KvsTh$) to minimize the predicted greater saliency that using size would cause. As it can be observed, even taking this measure, the particular spatial distribution of high values for $KvsTh$ creates a very salient feature. This effect, as we mentioned before, cannot be predicted by our models. Our users commented on the effectiveness of this display also, in particular the equal saliency and low interference between saturation and size changes, making it easy to read both data values simultaneously.

As a final step, users asked to correlate these three values to the concentration of H. Figure 6.14 shows the two options we provided to them using multiple layers. In Fig. 6.14 (a) we included the raw data directly underneath the previous display. This achieved the initial goal of highlighting the H concentration, but created too much interference with all other methods present. While our models do not include full color planes, we can estimate design factor values by plugging in low size and spacing values into our models. In order to minimize overall interference and create a more balanced display in terms of saliency we increased both size and spacing values of the bottom layer (shown in Fig. 6.14 (b)). We manually optimized the final parameterizations by looking at the predicted values from our models for each pair of methods. The combination shown provides the best balance between equal saliency and low interference, while utilizing the best methods based on the particular resolution characteristics of each data value.

Users were excited about the possibilities of this, for them, new visualization approach,

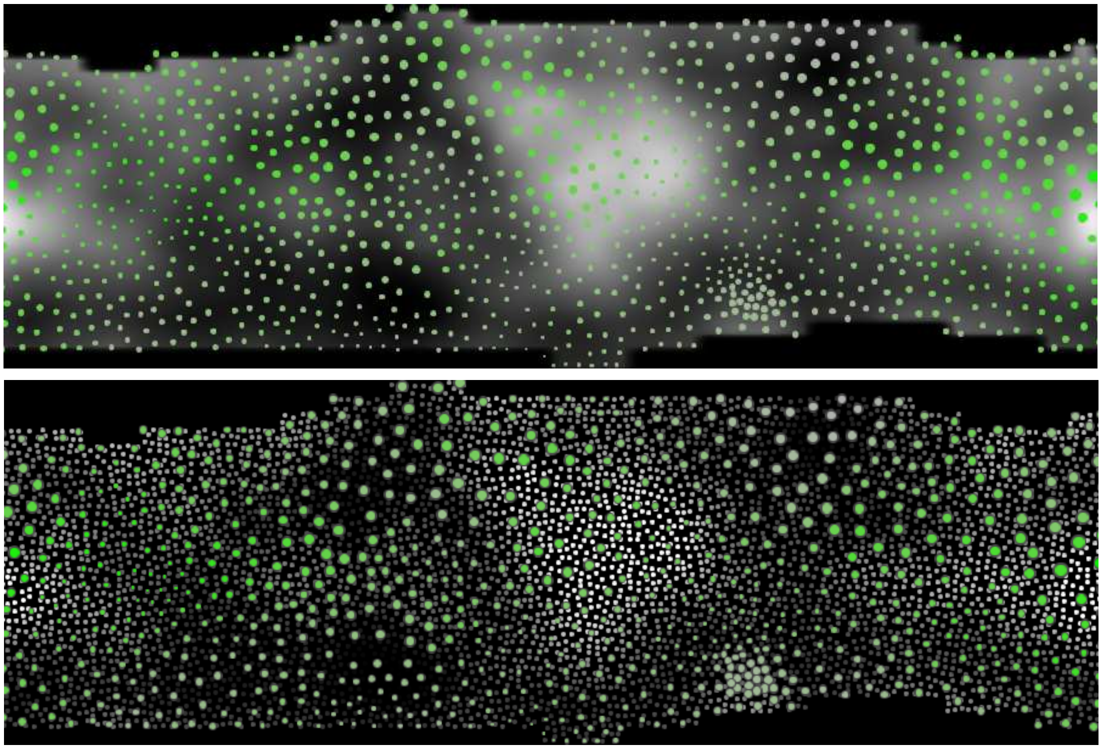
a)
b)


Figure 6.14: Two visualizations of the same four elements. (a) Shows H as background lightness, Cl as saturation, Fe as size, and $KvsTh$ as spacing changes. This option makes the background layer too salient with respect to the other values. In (b) we discretized the H data value and chose a combination of size and spacing for the bottom layer that would bring pairwise saliency values between methods closer together while, at the same time would minimize interference.

and our goal of using our models to achieve effective solutions efficiently was fulfilled. Furthermore, a more important conclusion of this experiment was the realization that, even when the original continuous datasets are discretized, users accept the loss of resolution in exchange for a quick and intuitive way of correlating their values. They especially praised the simplicity with which we controlled both the mappings and parameters to achieve different saliencies, and which data values should be be mapped to which visual dimensions based on their particular features.

In conclusion, we successfully utilized our models to predict the performance of the different methods for a real visualization problem. Although we performed all the optimizations and parameter tweaking by hand, the process we followed serves as an example of how changing requirements can be plugged into our models to control the knobs of our visualization software.

Chapter 7

Discussion and Conclusion

Through a variety of experiments, this dissertation has advanced the state of the art in scientific visualization by evaluating how perceptual studies and collaboration with visual designers can help predict the properties of visualization methods and ultimately their effectiveness. Our contributions include new experimental methods and perception models that can be used to guide future research and improve the utility of multivalued visualization methods.

This final chapter will summarize the main conclusions from all our experiments, and will propose some potentially rewarding future lines of research, including our ideas on how to solve some of the issues that we have identified along the way. We will also dedicate a section of this chapter to provide a series of guidelines for designing visualization evaluation experiments. Even though some of those might not be novel, we believe it is useful to complement our contributions with a checklist that reflects the lessons we have learned.

7.1 Research Summary

One guiding principle in designing our research plan was to complete our model from the bottom-up: we should first understand and quantify how the individual dimensions that form our visualizations work before trying to approach multivalued scenarios, where perceptual interactions would play a significant role. Our hypothesis was:

> Measuring the perceptual capabilities of several icon-based scientific visualization methods for simple single-valued scalar datasets in 2D, and combining that with subjective evaluations of complex multilayered methods representing multivalued datasets, we can generate a predictive model of utility of a space of visualization methods.

A key contribution of this dissertation was showing that expert visual designers can be effective evaluators of scientific visualization methods. Our first experiment showed that their subjective critiques can be significantly correlated with the results from more traditional objective quantitative studies based on task performance. In addition, and this is one of the benefits of using these expert visual designers as experimental subjects, they were able to provide us with reasons for certain visual dimension interactions, and to indicate how a change in these dimensions would further affect the utility of a method and the effective reading of the data variables.

We continued this line of research further by conducting a second study to evaluate the utility of individual 2D visualization methods in terms of a set of design factors, which were subjectively rated by expert visual design educators. We successfully characterized a total of 33 visualization methods using 11 different visual dimensions and 6 different design factors for representing single-variable continuous scalar datasets. While the ranking results obtained were informative, they were not nearly as exhaustive as we needed them to become the basis for our utility prediction model.

Therefore we approached the evaluation of our hypothesis from a perceptual standpoint, leaving the use of visual designer subjective critiques for a later stage, when we needed to evaluate complex visual displays, difficult to test perceptually. Our next experiments were aimed at quantifying the utility, in terms of some of our design factors, of a subset of the visual dimensions we set off to model. We could not try to explore such a vast space of visual dimensions exhaustively through psychophysical studies. Since the number of variables to be controlled was just unmanageable, we evaluated four of them: icon saturation, icon lightness, icon size and icon spacing.

The experiments evaluated some individual properties of these dimensions, such as data resolution and spatial feature resolution, and also properties of the perceptual relationships between them, such as saliency, and interference. We were able to successfully obtain predictive mathematical models based on variables such as the size and spacing of the icons, their color, the number of layers used in the display and their order.

Once these models were completed, we designed a study that, using expert visual designers to evaluate real two-valued dataset visualizations, would allow us to quantify the loss in data resolution and spatial feature resolution when two methods were combined: i.e. to quantify the loss or gain in legibility. We had the baseline capabilities for each method when utilized individually, and we knew how much one method affected the time required to understand another or which one would dominate, but we needed to know the actually change in legibility to be able to propose a model for higher order combinations.

With this experiment we were not able to disprove the null hypothesis, as we explained in the previous chapter, and it left us with several open questions that will need to be answered and evaluated to move forward with this research. We did, however, validate the models we generated by obtaining anecdotal evidence from designers during the experiment and through informal demonstrations of the predictions of our models with scientists exploring multivalued datasets.

7.2 Open Issues and the Development of a New Hypothesis

As we mentioned before, there are two possible explanations why an experiment fails to disprove its null hypothesis: the hypothesis is false or the experiment was not designed in such a way that could capture a significant result if one existed.

We believe both options played a role in the outcome of the last experiment and hence in the whole project.

7.2.1 Investigating Data Dependence

The main reason our initial hypothesis could be false is that, from the beginning, we eliminated the dataset itself from the argument. Our aim was to obtain a general model that could be applied to any type of multivalued scalar dataset. We tried to capture some of the characteristics of those datasets by including data resolution and spatial feature resolution as two of our design factors.

Indeed, this provided, we believed, the link we needed between the visualization methods and the actual information needs they were required to address. In fact, while studying the individual methods these offered sufficient information to determine their utility. When we introduced two-valued visualizations, it became clear we needed some spatial data correlation information to effectively address the modeling of the methods' utilities. The intuition for this is as follows.

Let us suppose we have successfully completed the last experiment and we have obtained a full legibility model for all pairwise interactions between methods. That means, and limiting this discussion to data resolution legibility, we know how different levels of one method affect the reading of all levels from another. In order to apply this knowledge we need to know how the spatial distribution of values from a data variable coincides with the distribution of another. Here is where our hypothesis fails. We do not account for such data correlation information to be part of the model.

Our hypothesis does not completely fail since, instead, it errs on the side of caution. Without including the data, the only prediction of utility we can make is based on the overall ranges of the data variables. Given those, we can just apply our model to the extreme values and use those as a very conservative estimate of correlations. In fact, the coincidence of the extreme values may never occur and we could be dismissing perfectly valid solutions because we are not looking at the specific datasets being visualized. Figure 7.1 describes our attempt to obtain this data correlation information. Having that we could evaluate visualization method utilities with much more narrow and realistic constraints.

Note how this is precisely what expert visual designers were doing in our last experiment. They tweaked each combination of methods to fit the particular pair of data variables being visualized. We decided we could not take those results because they were not representing a general methodology but individual instances. This leads us to a difficult research situation.

On the one hand, if we do not address this data dependence we believe we would not be able to find a general model that would apply to any multivalued dataset. But, if we do address it, the model itself will generate too much variability in the choice of methods. As explained in Section 6.1.3, users would need to adapt constantly to new visualization displays for similar types of data, with important consequences for their research effectiveness.

We were not able to continue our research precisely because of this issue. Nevertheless, important lessons were learned and new perceptual models developed. We will now address the second possible reason we failed to validate our initial hypothesis: the experimental methodology.

7.2.2 Experimental Methodology and Guidelines

Our strategy in designing our last experiment was based on our interest in studying a particular portion of the space of visualization methods. We also utilized the same constraints we had for the perceptual experiments: fixed ranges for the methods involved and limited number of levels for the independent variables.

Our results show that these limitations impeded our experts to fully provide their expertise to solve our visual problem. The fact that subjects explored the full space of parameters every time signaled an unfamiliarity with both the dataset and the limited space we allowed them to explore. Participants declared the tasks were easy to understand and the dimensions used were clear, but the lack of freedom for exploring other options frustrated them and ultimately made their confidence drop.

From the initial experiment in this project, we have already supported the well-established idea of collaborating with visual designers for scientific visualization design and evaluation. It is indeed a hot topic in visualization literature, along with the use of perceptual knowledge. But while there is a whole area of science dedicated to the study of perception, very little has been done to try to capture design knowledge experimentally.

The goal is to try and learn from the visual process these experts go through when solving a visual communication problem. The key to do this, we believe, is to provide them with the right tools to do their job while controlling their exploration enough to be

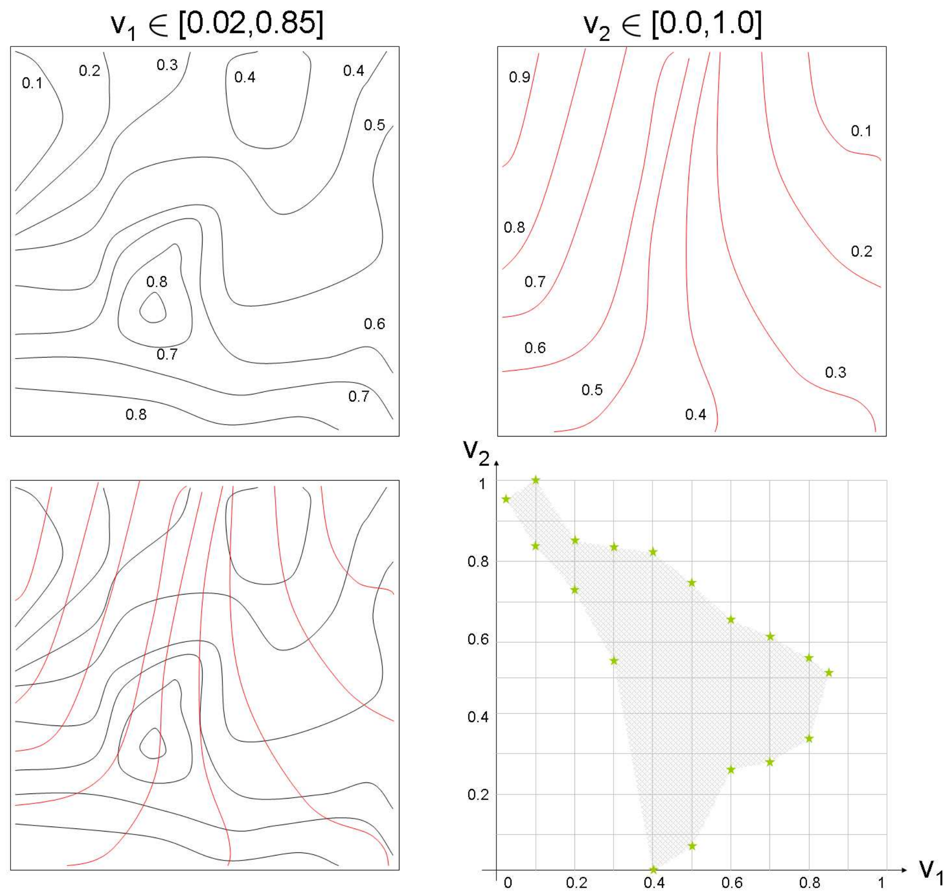


Figure 7.1: Given the two data variables, v_1 and v_2 on the top, we can delimit the value correlations across the range of the display by superimposing both variables (bottom left) and plotting what values of v_2 coincide with what values of v_1. Instead of assuming the worst case scenarios of just crossing the limits of the value ranges, this construction would give us much more narrow constraints. An optimization scheme could be set up based on this information and the capabilities (data resolution, in this case) of the methods chosen to represent these data. The result of the optimization would be the best ranges for those methods that produce the most (or any other specific goal) perceivable levels across the display.

analyzable. This could indeed prove to be an impossible quest: trying to capture on a fixed protocol something that, although it is undoubtly based on perception rules and established design knowledge, has a large component of experience, inspiration, and other influences that designers can let transpire into their work.

In summary, in our case we definitely constrained our experts too much, but we believe we needed to do so in order to understand their process enough to turn their responses into quantitative results. We are, once again, presented with two options: constrain and not get generalizable results, or do not constrain and not be able to quantify the results.

The same as before, this research has advanced the knowledge in our field by testing one of those strategies. We estimate that a full overhaul of the evaluation methodology would take many years to develop. It would require exhaustive observation of visual designers in their own workspaces and careful dissection of the process they follow. This still does not guarantee success, since the key to developing a useful utility model depends on understanding the individual dimensions of visualization and their capabilities, and they will be difficult to isolate on a top-down approach.

Nevertheless, our efforts were not without reward. In the process of evaluating our methodologies we have learned valuable lessons, and we hope future endeavors can benefit from our experience. Along these lines, we would like to highlight the main guidelines we believe must be followed when designing this type of experiment using expert visual designers. Note that some of these have not been tested and, therefore, are our hypothesis of what to try next.

Experimental Guidelines

This is by no means an exhaustive list but a set of guidelines we believe are the most important to take into account when designing new experiments.

- *Use comparative critique:* Providing participants with several displays to evaluate by contrast. This is an established technique in art and design [Feldman, 1994], and one that we successfully used in our scalar datasets experiment with expert visual designers. This enhances the quality of the participants comments and provides initial ranking information. It is hard to do this on paper due to printing calibration issues and, on a computer monitor, they must be high resolution and as large as possible.

- *Recruit experienced design educators:* We hypothesize that their critiques can be better suited for protocol analysis than expert non-educators or students. Throughout this dissertation we have mentioned the use of experienced educators as our subjects. Their approach to critiquing designs includes their experience teaching composition and communication concepts to students, and during our second experiment, they accommodated more easily to the thinking-out-loud protocol. It is, however, hard to schedule their participation but they are, in general, very enthusiastic about investigating lines of collaboration between art and science.

- *Allow for multiple subjects:* We believe we could improve our results by running two or more experts simultaneously in the same experiment. The most important benefit from this is that they could explain to each other task goals or other experiment related issues in a language they are familiar with. Practical limitations such as scheduling issues can sometimes impede this approach, such as in our case.

- *Create challenging tasks:* Simple tasks will not engage subjects and generate only general results with no new insights. Too complex tasks will, in turn, be difficult to control and analyze because of the multitude of elements present. A balance must be found to successfully complete studies of this kind.

- *Record real design session:* Allowing experts to use their own preferred tools improves their confidence in the results and allows them to more easily think out loud. We did this, in part, when we asked designers to create a new method during our first experiment with 2D vector visualization methods.

- *Include interaction:* Interactive exploration is known to enhance insight, and our research is no exception. However, care must be taken to eliminate as many limitations as possible. Our last experiment included interactive controls for our visual dimensions and, although it was welcomed by our participants, the experimental constraints provoqued frustration and lack of confidence in their results. As a guideline, we believe interaction should be limited to those elements that can be fully controlled with no constraints.

- *Combine objective and subjective studies:* The first can provide exact quantification of variables that can be directly used to generate mathematical models. The second can validate those results and help define what type of functions are needed for those models.

- *Present the ground truth:* Although showing the original data is in itself a visualization, which would contain the very perceptual issues we are trying to explore, it is necessary to incorporate this into the design. The ground truth is, after all, the subject matter for the experts' critiques. Experienced participants can abstract, to a point, from perceptual artifacts and concentrate in general trends and features.

7.2.3 A New Hypothesis

We can now attempt to provide a new hypothesis that we believe would guide a new set of experiments:

> By analyzing how expert visual designers solve complex multivalued visualization problems using their own tools, we can obtain a description of the design process and design psychophysical experiments that quantify the particular interactions among visual dimensions that experts pay attention to. The process description will then serve as a guide to generate a mathematical model that explains and predicts the utility of different combinations of visualization methods.

We believe this hypothesis constitutes a possible solution to the issues found in this research. It proposes a top-down approach to obtain both the key visual dimensions to consider and the interactions that experts pay attention to when presented with complex situations. We go from complex displays to identifying the process elements and analyzing them in detail.

This hypothesis also helps in two key aspects. First, it will engage experts where they are most effective: presented with challenging problems and allowed to use their own tools and techniques. Second, it proposes a solution to solve the data dependence problem.

Indeed our proposal acknowledges that it is possible to find a general model independent of the dataset. Presenting several experts with different problems to solve and trying to obtain commonalities among the process they follow is the part where this hypothesis might fail to produce the expected results. But, as opposed to our current approach, this challenge is placed at the beginning and the hypothesis can be adjusted based on the initial findings.

7.3 Impact of this Dissertation

Our hypothesis aimed at establishing a basis for a theory of visualization. This theory would be based on the quantification of the utility of different visualization methods.

Our results contribute to that goal by providing several models for the effective use of some 2D visualization methods. We have also produced a knowledge base for the design, execution, and analysis of evaluation studies that use expert visual designers as the main participants. We hope the visualization community will benefit from this body of work in its continuing quest for its theoretical foundations.

The experiments we have conducted add up to be an important methodological framework with which other visual dimensions can be explored. Through those experiments much has been learned that the visualization, perception, and visual design communities can build upon. Even if the results do not add up to a full fledge model to be plugged into a visualization software, our individual experimental results will help non-expert users in their search for an effective visualization, by providing some indication about probable directions of improvement for their visualizations, and by shedding some light as to what methods to use in what situations. This will allow researchers to concentrate on data analysis instead of visualization creation.

We recognize that this dissertation is just scratching the surface of the complex problem of effective visualization design. Our hope is that new lines of research, involving collaboration with visual design and art experts, as well as perceptual psychologists, will develop around the basis we had created with these initial results.

7.4 Future Directions of Enquiry

While developing our investigation, there were many options and ideas that we could not follow for different reasons. Nevertheless, we believe they are promising directions to be explored, with significant contributions at the end.

- *Optimization:*

  Upon a successful characterization of the methods' utility for multivalued cases, the next step is to match up those capabilities with the requirements from scientists wanting to visualize their data. An optimization process can be designed to weight the possible solutions and offer users the best options based on their demands. Furthermore, principal component analysis could be applied to try and discern the correlation among different visual dimensions. These components would provide knobs similar to the ones proposed in our idealized screenshots of our project's vision.

- *Double mapping:*

  The practical limits of experimental design constrained our investigations to a handful of visual dimensions and limited discretizations of the continuous axes defined by each one. Although we have taken into account and explored the different implications of single versus multiple layers of icons, we could not study the case where more than one visual dimension were mapped to the same data variable. It is recognized in the perceptual literature that synergistic relationships might occur when multiple visual cues are combined, forming emergent features. We would like to explore how multiple simultaneous mappings would affect the expressiveness of high order visualization methods. Indeed, in cases where some visual dimensions in a layer are *free*, maybe mapping them to already mapped data variables would enhance the overall utility of the visualization.

- *Genetic algorithms:*

  Given the high dimensionality of the space of visualization methods we are working with, an efficient search strategy is difficult to design. In particular, since our data are based on low order interactions and we are trying to look for effective solutions for higher order problems, the guiding of that search will not be efficient. Our original idea for this dissertation was to implement a genetic algorithm approach for this problem. Each visualization method would be an individual in our population, and the genome would be built with the different visual dimensions we are interested in using. The definition of an evaluation function that could select surviving individuals for each generation led to the current dissertation. We still believe, as do some other researchers in the field [House and Ware, 2002; House et al., 2006], that a GA approach would provide good results and an efficient exploration of the vast extent of the visualization space. The evaluation function would be based on the interactions and model for them defined in this dissertation.

- *Perceptual mapping:*

  Given that our datasets are scalar fields in 2D, we could analyze the local contrast between every pair of data variables by using a grayscale representation. Given a set of design goals, we could use local measurements of contrast to see how to optimally fit a single mapping (data to visual dimension) that would maintain the saliency requirement across the image. One hypothesis that would need to be evaluated would be that changing the mapping across the image to favor perception would not affect the reading of the data values. In other words, if we know that certain values of lightness and spacing conflict with each other by decreasing the data resolution of lightness, we could tweak the lightness mappings in those areas to perceptually maintain data resolution. This is a very complex and potentially risky proposition, since it basically means that we are creating a mathematically incorrect visualization (the mapping is not constant across the display) that is arguably perceived correctly.

- *Extensions to other visual dimensions in 2D*:

  One of the main contributions of this dissertation is that the methodology used to gather, analyze, and model the perceptual and visual design knowledge about the four dimensions we chose can be utilized to include others such as:

  - Icon hue.
  - Icon motion.
  - Streamline-type representation (for vector field visualization.)

- Ellipsoid-type representations (for tensor field visualization.)

- *Second-order effects:*

  Size and spacing are used both as independent variables and factors in our models. We have not investigated, because that is not the purpose of a dissertation in Computer Science, what are the processes used by our perceptual system to read these dimensions. We believe our eyes and brain obtain some brightness and contrast change information from size and spacing changes. If this is true, treating them as independent factors from lightness is not completely correct.

  Another effect would be appearance of three-dimensionality that some dimensions produce. Size, for example, when representing a linear dataset, appears to suggest a plane fading in the distance or coming out of the screen. Spacing or saturation can also generate such illusions. Similarly to the brightness effects, it is conceivable, but out of the scope of this thesis, that our perceptual system uses those illusory cues to help read the dimensions we are interested in. In that case, our model will implicitly include these effects, but possible interferences with other dimensions will not be detected.

  We recognize the importance of these effects, some of them very obvious, but we decided not to include them in this dissertation knowing they are a limitation for the applicability of our model. Nevertheless, we believe including these effects in the model's mathematical definitions would not be complicated, although the experimental designs to gather the appropriate measures would be.

- *Icon orientation:*

  After our experience trying to use orientation to represent scalar fields, we believe it should not be used for that purpose. Subjects had a lot of trouble going beyond the sense of "flow" it conveys, and developed strategies to read the scalar field, but only for our simple linear datasets. Even expert designers could not read orientation for general multi-valued datasets. This "flow" effect can also be considered a second-order effect but, in this case, it is our hypothesis that it is clear it interferes very much negatively in the legibility of orientation.

  On the other hand, orientation could be used as a differentiating factor for multilayered methods. Including this factor would only improve the model we present here and provide a more comprehensive coverage of visualization options.

- *Other multivalued visualization options:*

We have not explored the possibility that users will not accept the loss of legibility inherent in the discretization of a continuous dataset. The purpose of it is to enhance the multivalued capabilities, allowing for multiple layers. A very interesting avenue of study would be to compare the ability to find spatial correlations among variables using techniques such as side-by-side visualization, in-place image flipping (where the user clicks a button to switch between data variables in the same visual display using a single method), and our own icon-based multilayered approach.

- *Display Size and Conditions:*

  In our experiments room illumination, monitor calibration, and subject positioning were controlled to a first approximation. This introduces a limitation on how our results apply in other conditions. Although we cannot certainly expect the exact values to be maintained for other display form factors or illumination conditions, we believe our model will provide a valuable approximation to effective visualizations. Note that, even for conditions exactly matching our experiments, we do not claim to find a single best visualization, but a solution that puts users closer to an effective visual display of their data.

- *Extensions to 3D:*

  The same reasons that motivated us to create a predictive model of utility for 2D visualization apply for the 3D case. There are many perceptual artifacts that combine together to facilitate or impede effective visualization of phenomena in 3D. There is a clear need to understand how to quantify and harness those artifacts to create efficient and effective visualization methods.

  Our experimental methodology, measuring the individual contributions of each visual dimension and combining those in a predictive model, can be directly applied to the three dimensional case. The definition of the dimensions, however, is harder since there are more cues involved. The same goes for the design of the experiments, where it will be more difficult to isolate individual elements to obtain unbiased effect measurements, while maintaining a realistic view of how these methods would be used in practice. The same way we did on the 2D case, there needs to be a balance between isolating individual contributions and successfully generalizing to practical cases where many factors interact.

  We believe that this very complexity, and the vast amount of combinations possible when more cues are added, will allow visual designers and artists to really show their potential as visualization evaluators. With their holistic approach to critiquing, they are able to evaluate how the different elements present in a visual display participate in the overall composition.

7.5 Conclusion

Acquiring and using expert visual design knowledge and perceptual interaction data at the scope proposed has been intellectually challenging, and the framework we presented in this thesis to do it advances the state of the art in scientific visualization. Better visualizations have the potential to advance science more quickly by improving our understanding of physical and biological phenomena, applied science, and engineering. This dissertation enhances the channels of collaboration in education and research among the disciplines of cognitive science, visual design, art and scientific visualization. It advances our understanding about the areas in which each discipline influences the visualization design process and the quality of the final product: effective scientific visualizations.

"The prize is the pleasure of finding the thing out, the kick in the discovery, the observation that other people use it [my work] –those are the real things, the honors are unreal to me."
– Richard P. Feynman, in *The Pleasure of Finding Things Out.*

Bibliography

Acevedo, D., Chen, J., and Laidlaw, D. H. (2007a). Modeling perceptual dominance among visual cues in multilayered icon-based scientific visualizations. IEEE Visualization'07, Poster Compendium.

Acevedo, D., Jackson, C., Laidlaw, D. H., and Drury, F. (2005). Using visual design expertise to characterize the effectiveness of 2D scientific visualization methods. In *Proceedings IEEE Visualization, Poster Compendium (BEST POSTER AWARD)*.

Acevedo, D., Jackson, C., Laidlaw, D. H., and Drury, F. (2007b). Using visual design experts in critique-based evaluation of 2d vector visualization methods. *IEEE Transactions on Visualization and Computer Graphics*. In Review.

Acevedo, D. and Laidlaw, D. H. (2006). Subjective quantification of perceptual interactions among some 2D scientific visualization methods. *IEEE Transactions on Visualization and Computer Graphics (Proceedings Visualization / Information Visualization)*, 12(5).

Acevedo, D., Vote, E., Laidlaw, D. H., and Joukowsky, M. (2001). Archaeological data visualization in VR: Analysis of lamp finds at the Great Temple of Petra, a case study. In *Proceedings IEEE Visualization 2001 (BEST CASE-STUDY AWARD)*, pages 493–496.

Acevedo, D., Zhang, S., Laidlaw, D. H., and Bull, C. (2004). Color rapid prototyping for diffusion tensor MRI visualization. In *Proceedings of MICCAI 2004 Short Papers*.

Andrienko, G. and Andrienko, N. (1999). Data characterization schema for intelligent support in visual data analysis. In *Lecture Notes in Computer Science*, pages 349–366. Springer-verlag.

Bergen, J. R. (1991). Theories of visual texture perception. In *Spatial Vision*. CRC, Boca Raton, FL.

Bergen, J. R. and Landy, M. S. (1991). Computational modeling of visual texture segregation. In *Computational Models of Visual Processing*, pages 253–271. MIT Press, Cambridge, MA.

Bertin, J. (1983). *Semiology of Graphics.* University of Wisconsin Press.

Bokinsky, A. A. (2003). *Multivariate Data Visualization with Data-Driven Spots.* PhD thesis, University of North Carolina at Chapel Hill.

Brooks, F. P. (1996). The computer scientist as a toolsmith II. *Communications of the ACM*, 39(3):61–68.

Cabral, B. and Leedom, L. C. (1993). Imaging vector fields using line integral convolution. In Kajiya, J. T., editor, *Computer Graphics (SIGGRAPH '93 Proceedings)*, volume 27, pages 263–272.

Callaghan, T. C. (1984). Dimensional interaction of hue and brightness in preattentive field segregation. *Perception and Psychopysics*, 36(1):25–34.

Callaghan, T. C. (1989). Interference and dominance in texture segregation: Hue, geometric form, and line orientation. *Perception and Psychopysics*, 46(4):299–311.

Card, S. K. and Mackinlay, J. (1997). The structure of the information visualization design space. In *Proceedings of IEEE Symposium on Information Visualization*, pages 92–99.

Carswell, C. M. and Wickens, C. (1990). The perceptual interaction of graphical attributes: Configurality, stimulus homogeneity, and object integration. *Perception and Psychopysics*, 47:157–168.

Casner, S. M. (1991). A task-analytic approach to the automated design of graphic presentations. *ACM Transactions on Graphics*, 10(2):111–151.

Chi, E. H. (2000). A taxonomy of visualization techniques using the data state reference model. In *Proceedings of InfoVis 2000*, pages 69–76.

Cleveland, W. S. and McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):531–554.

Cody, R. P. and Smith, J. K. (2006). *Applied statistics and the SAS programming language.* Pearson Prentice Hall, Upper Saddle River, N.J.

Dastani, M. (2002). The role of visual perception in data visualization. *Journal of Visual Languages and Computing*, 13(6):601–622.

Demiralp, C., Jackson, C., Karelitz, D., Zhang, S., and Laidlaw, D. H. (2006). Cave and fishtank virtual-reality displays: A qualitative and quantitative comparison. *IEEE Transactions on Visualization and Computer Graphics*, 12(3):323–330.

Dippé, M. A. Z. and Wold, E. H. (1985). Antialiasing through stochastic sampling. In Barsky, B. A., editor, *Computer Graphics (SIGGRAPH '85 Proceedings)*, volume 19, pages 69–78.

Eick, S. G. (1995). Engineering perceptually effective visualizations for abstract data. In *Scientific Visualization Overviews, Methodologies and Techniques*, pages 191–210. IEEE Computer Science Press.

Ellis, W. D. (1939). *A source book of gestalt psychology*. Harcourt, Brace, and company, New York, NY.

Feldman, E. B. (1994). *Practical Art Criticism*. Prentice Hall.

Grossberg, S. (2006). The art of seeing and painting. Technical Report CAS/CNS-2006-011, Department of Cognitive and Neural Systems, Boston University.

Hanrahan, P. (2005). Teaching visualization. *Computer Graphics*, 39(1):4–5.

Healey, C. G. (1996). Choosing effective colours for data visualization. In *Proceedings of IEEE Visualization'96*, San Francisco, CA.

Healey, C. G., Amant, R. S., and Elhaddad, M. S. (1999). Via: A perceptual visualization assistant. In *28th Workshop on Advanced Imagery Pattern Recognition*.

Healey, C. G., Booth, K. S., and Enns, J. T. (1993). Harnessing preattentive processes for multivariate data visualization. In *Proceedings of Graphics Interface*, pages 107–117, Toronto, Canada.

Healey, C. G., Booth, K. S., and Enns, J. T. (1996). High-speed visual estimation using preattentive processing. *ACM Transactions on Human-Computer Interaction*, 3(2):107–135.

Healey, C. G. and Enns, J. T. (1999). Large datasets at a glance: Combining textures and colors in scientific visualization. *IEEE Transactions on Visualization and Computer Graphics*, 5(2):145–167.

Healey, C. G., Enns, J. T., Tateosian, L., and Rempel, M. (2004). Perceptually-based brush strokes for nonphotorealistic visualization. *Transactions on Graphics*, 23(1).

Heydayat, A., Sloane, N., and Stufken, J. (1999). *Orthogonal arrays : theory and applications*. Springer.

Hibbard, B. (2004). The top five problems that motivated my work. *IEEE Computer Graphics and Applications*, 24(6):9–13.

House, D. and Ware, C. (2002). A method for perceptual optimization of complex visualizations. In *Proceedings of Advanced Visual Interface*.

House, D. H., Bair, A. S., and Ware, C. (2006). An approach to the perceptual optimization of complex visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 12(4):509–521.

Interrante, V., Fuchs, H., and Pizer, S. M. (1997). Conveying shape of smoothly curving transparent surfaces via texture. *IEEE Transactions on Visualization and Computer Graphics*, 3(2):98–117.

Jackson, C., Acevedo, D., Laidlaw, D. H., Drury, F., Vote, E., and Keefe, D. (2003). Designer-critiqued comparison of 2D vector visualization methods: A pilot study. In *SIGGRAPH 2003 Sketches and Applications*. ACM SIGGRAPH.

Jankun-Kelly, T. (2003). *Visualizing Visualization: A Model and Framework for Visualization Exploration*. PhD thesis, Center for Image Processing and Integrated Computing, University of California, Davis.

Johnson, C. (2004). Top scientific visualization research problems. *IEEE Computer Graphics and Applications*, 24(4):13–17.

Keefe, D., Karelitz, D., Vote, E., and Laidlaw, D. H. (2005). Artistic collaboration in designing VR visualizations. *IEEE Computer Graphics and Applications*, 25(2):18–23.

Keller, J. M., Boynton, W. V., Karunatillake, S., Baker, V. R., Dohm, J. M., Evans, L. G., Finch, M. J., Hahn, B. C., Hamara, D. K., Janes, D. M., Kerry, K. E., Newsom, H. E., Reedy, R. C., Sprague, A. L., Squyres, S. W., Starr, R. D., Taylor, G. J., and Williams, R. M. S. (2006). Equatorial and midlatitude distribution of chlorine measured by mars odyssey GRS. *Journal of Geophysicial Research*, 111.

Kirby, M., Keefe, D., and Laidlaw, D. H. (2004). Painting and visualization. In *Visualization Handbook*. Academic Press.

Kirby, M., Marmanis, H., and Laidlaw, D. H. (1999a). Visualizing multivalued data from 2D incompressible flows using concepts from painting. In *Proceedings of IEEE Visualization 1999*, pages 333–340.

Kirby, R. M., Marmanis, H., and Laidlaw, D. H. (1999b). Visualizing multivalued data from 2d incompressible flows using concepts from painting. In *Proceedings Visualization '99*. IEEE Computer Society Press.

Kosara, R., Healey, C. G., Interrante, V., Laidlaw, D. H., and Ware, C. (2003). User studies: Why, how, and when. *Computer Graphics and Applications*, 23(4):20–25.

Laidlaw, D. H. (2001). Loose, artistic "textures" for visualization. *IEEE Computer Graphics and Applications*, 21(2):6–9.

Laidlaw, D. H., davidkremers, Toga, A., Drury, F., and Jacobs, R. E. (2004). Applying lessons from visual art to exploration of the brain. Panel in Thirty-Seventh Annual Winter Conference on Brain Research.

Laidlaw, D. H., Kirby, M., Jackson, C., Davidson, J. S., Miller, T., DaSilva, M., Warren, W., and Tarr, M. (2005). Comparing 2D vector field visualization methods: A user study. *In IEEE Transactions on Visualization and Computer Graphics*, 11(1):59–70.

Lamming, D. (1991). Constrast sensitivity. In *Vision and Visual Dysfunction*.

Landy, M. and Movshon, J. (1991). *Computational Models of Visual Processing*. MIT Press, Cambridge, MA.

Landy, M. S. and Bergen, J. R. (1991). Texture segregation and orientation gradient. *Vision Research*, 31(4):679–691.

Lange, S., Schumann, H., Muller, W., and Kromker, D. (1995). Problem-oriented visualisation of multi-dimensional data sets. In *Proceedings of the International Symposium and Scientific Visualization*, pages 1–15.

Laper, N. (1995). *Mix and Match: A Construction Kit for Scientific Visualization*. PhD thesis, University of California Santa Cruz.

MacEachren, A. and Kraak, M.-J. (1997). Exploratory cartographic visualization: Advancing the agenda. *Computers and Geosciences*, 23(4):335–343.

Mackinlay, J. (1986). Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics*, 5(2):110–141.

McCleary Jr., G. F. (1983). An effective graphic vocabulary. *IEEE Computer Graphics and Applications*, 3(2):46–53.

McCool, M. and Fiume, E. (1992). Hierarchical poisson disk sampling distributions. In *Proceedings of ACM SIGGRAPH*, pages 94–105.

Miceli, K. D. (1992). A framework for the design of effective graphics for scientific visualization. Technical Report RNR-92-035, NASA Ames Research Center.

Nagappan, R. (2001). A compositional model for multidimensional data visualisation. In *In Proceedings of SPIE, Visual Data Exploration and Analysis VIII*, pages 156–167.

Nielsen, J. (1992). Finding usability problems through heuristic evaluation. In *Proceedings of CHI*, pages 373–380.

Norberg, U. M. and Rayner, J. M. V. (1987). Ecological morphology and flight in bats. wing adaptations, flight performance, foraging strategy and echolocation. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 316:337–419.

Nowell, L. T. (1997). *Graphical Encoding for Information Visualization: Using Icon Color, Shape, and Size to Convey Nominal and Quantitative Data*. PhD thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

Rhyne, T.-M. (2003). Does the difference between information and scientific visualization really matter? *IEEE Computer Graphics and Applications*, 23(3):6–8.

Robertson, P. K. (1991). A methodology for choosing data representations. *IEEE Computer Graphics and Applications*, 11(3):56–67.

Salisbury, L. D. P. (2001). *Automatic Visual Display Design and Creation*. PhD thesis, Department of Computer Science and Engineering, University of Washington.

Sayre, H. M. (1995). *Writing about Art*. Prentice Hall, 2nd edition.

Senay, H. and Ignatius, E. (1994). A knowledge-based system for visualization design. *IEEE Computer Graphics and Applications*, 14(6):36–47.

Shneiderman, B. (1996). The eyes have it. a task by data type taxonomy for information visualizations. In *Proceedings of IEEE Symposium on Visual Languages*, pages 336–343, Boulder, CO.

Sobel, J., Forsberg, A., Laidlaw, D. H., Zeleznik, R., Keefe, D., Pivkin, I., Karniadakis, G., Richardson, P., and Swartz, S. (2004). Particle flurries: Synoptic 3D pulsatile flow visualization. *IEEE Computer Graphics and Applications*, 24(2):76–85.

Springmeyer, R. R., Blattner, M. R., and Max, N. L. (1992). A characterization of the scientific data analysis process. In *Proceedings of the Second IEEE Visualization Conference*, pages 235–242.

Swan, J. E., Interrante, V., Laidlaw, D. H., Rhyne, T.-M., and Munzner, T. (1999). Visualization needs more visual design! Sensory design issues as a driving problem for visualization research. In *Proceedings of IEEE Visualization Conference*, pages 485–490, San Francisco, California.

Taylor II, R. M. (2002). Visualizing multiple scalar fields on the same surface. *IEEE Computer Graphics and Applications*, 22(2):6–10.

Thomas, J. (2005). *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Center.

Tory, M. and Moller, T. (2004). Human factors in visualization research. *IEEE Transactions on Visualization and Computer Graphics*, 10(1):72–84.

Tory, M. and Moller, T. (2005). Evaluating visualizations: Do expert reviews work? *IEEE Computer Graphics and Applications*, 25(5):8–11.

Tufte, E. (1983). *The Visual Display of Quantitative Information*. Graphics Press.

Tufte, E. (1990). *Envisioning Information*. Graphics Press.

Tufte, E. (1997). *Visual Explanations*. Graphics Press.

Turk, G. and Banks, D. (1996). Image-guided streamline placement. In *Proceedings of SIGGRAPH 96*, pages 453–460. ACM SIGGRAPH.

van Wijk, J. (2006). Bridging the gaps. *IEEE Computer Graphics and Applications*, 26(6):6–9.

van Wijk, J. J. (2005). The value of visualization. In *Proceedings of IEEE Visualization*, pages 79–86, Minneapolis, MN.

Vote, E., Acevedo, D., Jackson, C., Sobel, J., and Laidlaw, D. H. (2003). Design-by-example: A schema for designing visualizations using examples from art. In *SIGGRAPH 2003 Sketches and Applications*. ACM SIGGRAPH.

Vote, E., Acevedo, D., Laidlaw, D. H., and Joukowsky, M. (2002). Discovering Petra: Archaeological analysis in VR. *IEEE Computer Graphics and Applications*, 22(5):38–50.

Wallschlaeger, C. and Busic-Snyder, C. (1992). *Basic Visual Concepts and Principles for Artists, Architects and Designers*. McGraw Hill.

Ware, C. (2004). *Information Visualization. Perception for Design*. Elsevier, 2nd edition.

Watanabe, T. and Cavanagh, P. (1996). Texture laciness: the texture equivalent of transparency? *Perception*, 25(3):293–303.

Watson, B. (2006). Broadening our collaboration with design. *IEEE Computer Graphics and Applications*, 26(5):18–21.

Weigle, C., Emigh, W., Liu, G., Russell M. Taylor II, Enns, J. T., and Healey, C. G. (2000). Oriented sliver textures: A technique for local value estimation of multiple scalar fields. In *Proceedings of Graphics Interface*.

Wolfe, J. M. (1998). Visual search. In *Attention*. University College London Press, London, UK.

Woodhouse, I., Turner, D., and Laidlaw, D. H. (2002). Improving the visualization of polarimetric response in SAR images: from pixels to images. In *Proceedings of IEEE IGARSS*.

Zhang, S., Demiralp, C., and Laidlaw, D. H. (2003). Visualizing diffusion tensor MR images using streamtubes and streamsurfaces. *IEEE Transactions on Visualization and Computer Graphics*, 9(4):454–462.