

A User Study of Visualization Effectiveness Using EEG and Cognitive Load

E. W. Anderson¹, K. C. Potter¹, L. E. Matzen², J. F. Shepherd², G. A. Preston³, and C. T. Silva¹

¹SCI Institute, University of Utah, USA

²Sandia National Laboratories, USA

³Utah State Hospital, USA

Abstract

Effectively evaluating visualization techniques is a difficult task often assessed through feedback from user studies and expert evaluations. This work presents an alternative approach to visualization evaluation in which brain activity is passively recorded using electroencephalography (EEG). These measurements are used to compare different visualization techniques in terms of the burden they place on a viewer's cognitive resources. In this paper, EEG signals and response times are recorded while users interpret different representations of data distributions. This information is processed to provide insight into the cognitive load imposed on the viewer. This paper describes the design of the user study performed, the extraction of cognitive load measures from EEG data, and how those measures are used to quantitatively evaluate the effectiveness of visualizations.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: General—Human Factors, Evaluation, Electroencephalography

1. Introduction

Efficient visualizations facilitate the understanding of data sets through an appropriate choice of visual metaphor. Within the field of visualization, there exist numerous display strategies, many of which can be applied to similar types of data. These various techniques often create distinct imagery, emphasizing particular data characteristics or visualization goals. In most cases, several rendering techniques are appropriate; however, some methods may present salient information more quickly and accurately. The choice of best visualization technique for a particular data set is difficult to make. The visualization expert must not only determine an appropriate technique for the type of data, but also ensure the chosen method will answer the questions posed by domain experts. The difficulty of this choice is exacerbated by the lack of exhaustive visualization evaluation detailing the effectiveness of methods for particular types of inquiry.

Often, evaluation of visualization techniques is conducted through expert assessments and user studies, which typically judge a visualization using verbal feedback and user performance. While some measures of usability and effectiveness are relatively easy to quantify, such as increases in users' response speed or decreases in their error rates, others are problematic. For example, it is difficult to assess improved understanding and insight because those metrics tend to be highly subjective. Approaches to evaluation which rely on verbal feedback can be influenced by personal preference, user expectations, cultural biases within scientific fields, and resistance to change. The work described in

this paper strives to evaluate visualization techniques objectively by using passive, non-invasive monitoring devices to measure the burden placed on a user's cognitive resources.

The study we present in this paper explores the amount of work, defined by cognitive load, needed to interpret a visualization. We evaluate some simple visualization methods by measuring the brain activity through electroencephalography (EEG). A framework is defined for the processing and analysis of the acquired EEG sensor data which allows for the interpretation of difficulty of a visualization task. We believe the results of this study to be an important advancement of objective visualization evaluation.

This work offers the following contributions to the field of visualization analysis and evaluation:

- The use of EEG to inspect brain activity while interpreting visualizations.
- The use of cognitive load as a objective measure of visualization effectiveness
- The formulation of cognitive load based on its spatial, spectral, and temporal organization.
- The use of working memory as an estimation of cognitive load.

2. Visualization Evaluation: A Review

A substantial barrier to the evaluation of visualization techniques is the complexity of the task. Not only must a technique appropriately portray the data, but it also must sufficiently outperform equivalent rendering techniques. While appropriate measures for these requirements are difficult to

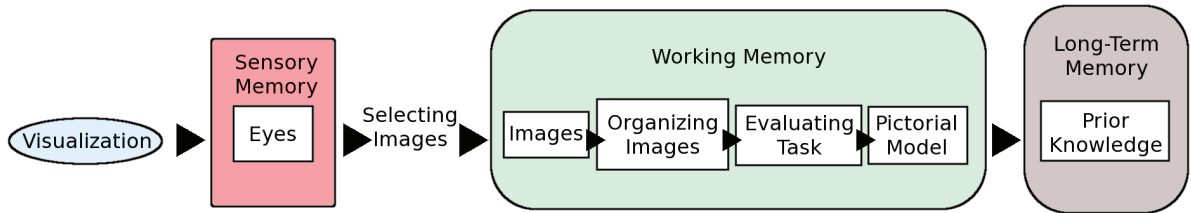


Figure 1: The cognitive and memory model of a single trial

formulate, there exists the additional challenge that most visualization problems are highly application dependent; visualization techniques that are validated as effective for one particular type of problem may not perform well for another one, even if the two are similar. Many visualization techniques are presented with evaluations which rely on technical improvements such as speedups, or the management of larger data sets. However, the use of human factors, user studies or expert evaluations is becoming more common.

User studies are effective ways of evaluating everything from visualization methods [SZB*09, LKJ*05] to complex environments such as airplane cockpits [SW94] and surgical simulators [RBBS06]. These classes of user studies generally use post-experiment surveys in conjunction with timing and task-related data to form a foundation for additional statistical analysis. These user studies leverage both empirical data collected during the user task as well as subjective data collected after the experiment.

While user studies have become an important tool in the assessment of visualization methods, they are not always the best evaluation technique. Kosara, et al. [KHI*03] show that user studies are effective at answering specific questions, such as “Does a specific method of streamline rendering show vorticities better than others?” Similarly, Cleveland and McGill [CM84] use evaluation studies to answer focused questions about data visualized in different ways.

Human factors play an important role in the study of the impact of scientific visualization on research. They are particularly important during the evaluation of visualization systems. An example of this type of system is Kosara, et al. using semantic depth of field [KMH01] in which renderings strive to induce perceptual changes in the user. Tory and Möller [TM04] offer a thorough discussion of human factors in both user study methods, and also in visualization design.

3. Cognitive Load: A Review

Cognition is defined as the process of knowledge acquisition and reasoning, and is responsible for our understanding of visualizations through ingestion and interpretation of an image. Working memory is a central construct of the cognitive process, and the burden placed on working memory and cognitive load can be used as a means to measure the efficacy of a visualization. Inspecting brain activity during a cognitive task offers an opportunity to assess the cognitive performance associated with various visualization methods.

Cognitive load and working memory are linked concepts [Eng02]. Working memory is the aspect of short-term memory responsible for the retrieval, processing, and integration of data during executive decision making [Bad92]. Figure 1 represents the general sensory and cognitive pathway used during the interpretation of a visualization. Imagery is first processed by the visual system and is then organized and evaluated by the working memory and cognition centers. Prior knowledge is then used to determine the appropriate cognitive schema for data interpretation.

The capacity and performance of the neural circuitry that implements working memory plays a vital role in cognitive activities. Figure 2 depicts the relationship between working memory capacity and the various types of cognitive load present during a single trial [PRS03]. It is useful to distinguish between working memory performance and task performance. Task performance is typified by a participants external performance of a task; for example, the time it takes to complete the task or the ratio of incorrect responses to correct ones. Working memory performance is measured by the spectral changes in the alpha and theta frequency bands as measured by EEG (as described by Klimesch [Kli99]). As Figure 2 shows, the various cognitive load sub-types remain constant for a given task, but the working memory capacity and performance are inversely proportional. This relationship provides a measurable quantity that is used to determine the overall cognitive load associated with the task.

3.1. Working Memory

Working memory is responsible for the retrieval, manipulation, and processing of task-related information and has functional importance to a variety of cognitive activities including learning, reasoning, and comprehension [Bad92]. It is often useful to think of the working memory system in terms of a computer architecture in which working memory acts as the central processing unit (CPU) with direct connections to temporary data buffers (RAM) in the form of short-term memory, and external communications (IO) through sensory perceptions and resulting reactions [Bad92]. Of course, the actual working memory system is much more complex than a computer, and therefore dividing up the processes of the system is not always possible, as many of the functions occur across the same neural substrate [CPB*97]. Although a strict spatial segmentation of the brain in terms of working memory activity is impossible, Braver, et al. show

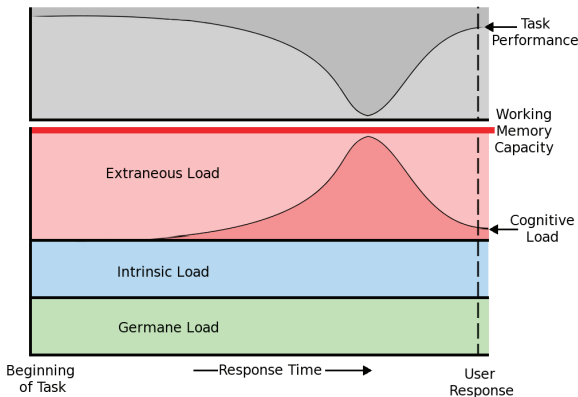


Figure 2: The combination of germane, intrinsic, and extraneous load to form working memory capacity and the impact of higher cognitive load (bottom curve) on task performance (top curve). Note that cognitive load peaks prior to the user's response to the task.

that the working memory processing is measurable in the prefrontal cortex of the brain [BCN*97] while Constantinidis, et al. explore a more complete neural circuit for spatial working memory [CW04]. Working memory is also divided into visuo-spatial, phonological, and executive sub-systems [Bad83]. In this work, our processing techniques focus on the visuo-spatial and executive working memory circuits by weighting contributions from the prefrontal cortices more heavily than those of the parietal regions.

3.2. Cognitive Load Theory

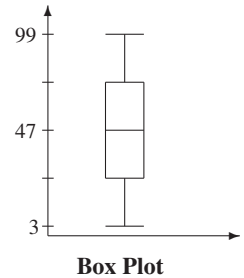
Cognitive load theory [Swe05] describes the relationship between the capacity of working memory and the cognitive demands of a particular task. The core of the theory is that people have a limited cognitive capacity during learning and problem solving tasks. The way in which information is presented can affect the amount of load placed on the working memory system and thus affect performance [Eng02]. Cognitive load theory distinguishes three types of cognitive load: germane, intrinsic, and extraneous [CS91]; each distinctly affecting learning and decision making. The combination of the three types characterize the overall cognitive load [SJB07] (Figure 2).

Germane Cognitive Load: Germane cognitive load is the load devoted to learning new cognitive schema [Swe05]. These schema are internal representations formed in the learning process which are used over and over and may be relevant to many tasks. Once these cognitive schema are in place, the contribution of germane cognitive load to the overall load is minimal.

Intrinsic Cognitive Load: Intrinsic cognitive load describes the demands on working memory capacity generated by the innate complexity of the information being examined [Swe05]. This load represents the portion of overall cognitive load that is influenced by the difficulty of the

3, 5, 28, 78, 72, 40, 52, 37
76, 6, 26, 68, 96, 70, 66, 75
34, 33, 20, 74, 36, 85, 99, 51
99, 33, 18, 38, 14, 18, 37, 53
25, 8, 69, 85, 25, 65, 30, 28
12, 87, 59, 54, 6, 30, 16, 59
97, 66, 23, 84, 87, 76, 36, 15
97, 87, 93, 12, 70, 56, 94, 97

Table of Data



Box Plot

Figure 3: An example of extraneous cognitive load. Both figures represent the underlying data; however, the visual nature the box plot facilitates understanding by taxing the working memory system less than the numerical description.

underlying task at hand and cannot be manipulated by the design of the task. An example of intrinsic cognitive load is the inherent challenge involved in adding two numbers compared to the greater challenge in solving more advanced arithmetic problems.

Extraneous Cognitive Load: Extraneous cognitive load measures the additional load placed on users by the design of a task [PRS03]. This type of load can be controlled by the way information is presented [SJB07]. For example, Figure 3 shows two ways to describe data. On the left, is a numerical description and on the right is a visual one. The box plot quickly gives a summary of the data through a visual presentation, while the numerical display requires more extraneous cognitive load to extract the properties of the data.

3.3. Measuring Cognitive Load

One method of measuring the various types of cognitive loads is by using task completion time and accuracy. Another method of measuring cognitive load is the NASA-TLX test [HKD*99]. This test describes cognitive load in terms of subjective responses to a post-experiment survey. However, EEG-based processing is capable of determining cognitive load magnitude by analyzing the temporal, spectral, and spatial patterns of brain activity. The Aegis simulation environment [BLR*05] was evaluated using EEG to monitor the amplitude of brain activity induced by situational properties of the task. In this way, cognitive strains placed on the participants involved in the study were measured.

In our study, we employ EEG to measure brain activity related to cognitive load and working memory; however, other physiological measures, such as pupil dilation or galvanic skin response have also proven useful in assessing cognitive load [SRT*07, KTH11]. Physiological measures in user studies do not always attempt to measure cognitive stresses directly. Recently, eye tracking technology has shown great utility in studying topics ranging from graph comprehension [CS98] to the use of contextual cues in visualization [PCVDW01]. However, it is still unclear to what degree these techniques capture cognitive responses elicited by visualization.

We exploit the spatial, temporal, and spectral organization of the neural circuits subserving working memory to measure its performance, as in [CPB*97]. The neural circuitry is monitored throughout the experiment using EEG. Although brain activity used to measure cognition is not visible in the raw EEG data, each data channel is processed to extract the spectral components associated with cognition and specifically, working memory [Kli99]. By measuring the performance of working memory, we measure the overall cognitive load imposed on a user in real-time. This real-time measurement cannot easily distinguish one cognitive load sub-type from another; however, the processing techniques allow us to make temporally sensitive analyses.

4. User Study of Cognitive Load

This user study is designed to evaluate different visualization techniques by measuring the amount of extraneous cognitive load each rendering imposes on the viewer. Because extraneous cognitive load is influenced by the way in which information is presented to the viewer, measuring its differences between visualization types provides insight into how the presentation of the data affects working memory and cognition. In order to reduce the complexity of this task, we have chosen to use simple visualization methods in this study. To this end, we compare variations of the *box plot* to see which is most effective in displaying a statistical data distribution.

The box plot is a graphical data analysis construct used to visually describe the distribution of a data set by indicating the minimum, median, and maximum data values, as well as the interquartile range (the range between the 25th and 75th percentile). The canonical box plot [Tuk77], (Figure 4a), does this by encompassing the central 50% of the data with a box, indicating the median with a crossbar, and extending lines out to the minimum and maximum values. Due to the box plot's simplistic representation of the underlying data, its use has become prolific in the scientific community, most notably to express error or variability within a data set. The extensive use of the box plot has supported various visual modifications, such as reducing the number of lines used to depict the plot [PKRJ10, Tuf83] (Figure 4b-c), or adding information about the density of the underlying data distribution [Ben88, PKRJ10, HN98] (Figure 4d-f).

The collection of box plots shown in Figure 4 were compared in this study to determine the extraneous cognitive load of each plot type. The plots were created based on 500 different normal distributions of size 100. For each distribution, the mean and standard deviation were picked uniformly random from [0,1] and [0.25, 0.75] respectively. For a single trial, two distributions are chosen and displayed using two types of box plots and the participant is asked to choose which of the distributions has a larger interquartile range.

4.1. Extracting Extraneous Cognitive load

EEG measures of cognition account only for overall load through the tracking of working memory performance; however, our interest lies in measuring extraneous cognitive load.

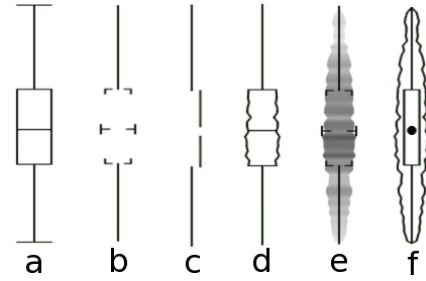


Figure 4: The plots used in the study. The left 3 plots are variations of the box plot: a) The Box Plot [Tuk77], b) Abbreviated Box Plot [PKRJ10], c) Interquartile Plot [Tuf83]. The right 3 are box plots with additional density information: d) Vase Plot [Ben88], e) Density Plot [PKRJ10], f) Violin Plot [HN98].

In order to extract extraneous cognitive load from overall cognitive load, the design of the user study must effectively control for the other cognitive load sub-types.

Germane cognitive load is controlled for by collecting subjective data relating to participant expertise. In a post-experiment survey, each participant rates their ability in interpreting the visualizations, and this information is used to approximate germane load on a per-user basis. The responses to each question on the survey are given on a Likert scale [Lik32], which asks respondents to specify their level of agreement to a statement. The survey questions are specifically designed to capture both user expertise in the interpretation of statistical data as well as the aesthetic qualities of each visualization technique. To negate the cognitive contribution of germane load, participants were required to be familiar with one-dimensional distribution data, and thus had pre-formed cognitive schemas. Germane cognitive load per participant was then judged to be negligible.

Intrinsic cognitive load is represented by *task difficulty*. When comparing various types of box plots, task difficulty refers to the complexity intrinsically present in deciphering differences in the interquartile range of two data sets, independent of the plotting method. When comparing images, the task is facilitated by examining common reference points within the two images. In the case of assessing which of two box plots has a larger interquartile range, the relevant common reference points are the locations of the first and third quartiles, and the median. The greater the similarity between the medians, the better the correspondence between the images, making the underlying task easier. However, as the distributions' interquartile ranges become similar, determining the distribution with the larger range is more difficult.

The measure of task difficulty takes into account both the interquartile range, *IQR*, defined as the difference between the first and third quartiles, $IQR = Q_3 - Q_1$, and the median, \tilde{m} , of the two underlying data distributions. Since we restrict the range of the generated distribution to be [0, 1], we can define task difficulty between two distributions, i and j , as

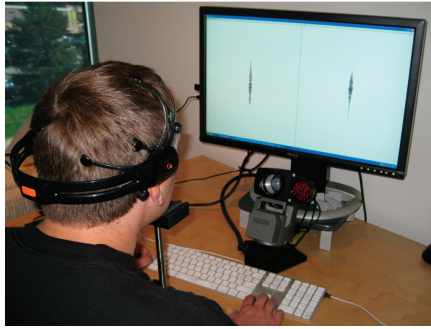


Figure 5: A participant is fit with the EEG headset to monitor brain activity for the duration of the 100 trial experiment. Distribution visualization pairs are presented side-by-side during each trial and a keyboard is used to enter responses.

$d(i, j) = 0.5(1 - |IQR_i - IQR_j| + |\tilde{m}_i - \tilde{m}_j|)$. By formulating task difficulty in this way, we are guaranteed that each single trial has a difficulty in the range $[0, 1]$ in which 1 represents the highest degree of difficulty. In practice, task difficulty and thus intrinsic cognitive load, was uniformly distributed in the range $[0.4, 0.8]$.

5. Data Analysis

Investigating the effects of different visualization techniques in terms of cognitive load requires the analysis of the various data products generated during the experiment. Time series data collected by EEG hardware must be rigorously processed to extract relevant working memory and cognitive load measures. Similarly, specific values acquired from user interaction must be manipulated to determine the task difficulty and reaction times experienced during each trial. Finally, each of the various data products must be statistically analyzed to ensure cognitive load measures are appropriate for visualization evaluation.

5.1. Data Acquisition

A group of 17 individuals consisting of 10 males and 7 females participated in the user study. The user study consists of 100 independent single trials preceded by a resting period of one minute during which baseline values for EEG are collected. Figure 5 shows a participant during a single trial of the experiment. Each trial begins with a two second period in which no images are shown, and is followed by the display of two box plots, side by side, as shown as the stimulus at the top of Figure 6. The participant is asked to choose the plot with the largest interquartile range as quickly as possible, and respond by pressing the appropriate directional arrow button on a standard keyboard.

Timing and response data is recorded during the experiment through custom-written display and acquisition software. A timer with 10 microsecond resolution was used to record response times during each of the single trials. In addition to the timing data used to determine reaction time,

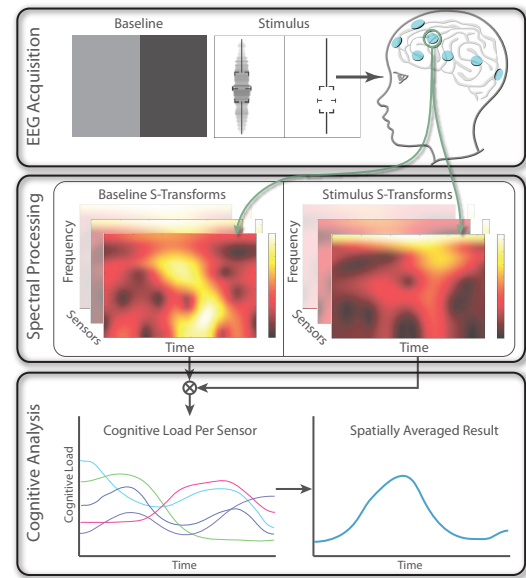


Figure 6: The experimental data collection and analysis workflow. EEG is collected during each of the 100 trials and then segmented into Baseline and Stimulus Epochs. These epochs are then processed using the S-Transform for each sensor. The resulting time-frequency planes are further processed to extract the gravity frequency and energy density for the theta and alpha bands of frequencies in each epoch. These values are combined in the Cognitive Analysis resulting in a single time series of cognitive load for each sensor. These time series are then combined through spatially-aware averaging to form the overall cognitive load for the trial.

each distribution's central moments and the response given by the participant are recorded for later analysis.

EEG data is collected at 128 Hz from an Emotiv EPOC wireless EEG headset (<http://www.emotiv.com>). The Emotiv headset exposes 14 data channels with two bipolar reference electrodes spatially organized using the International 10–20 system, as seen in Figure 7. The Emotiv Software Development Kit (SDK) provides a packet count functionality to ensure no data is lost, a writable marker trace to ease single trial segmentation tasks, and real-time sensor contact to ensure quality measurements.

During the experiment, a unique marker value is inserted into the marker trace to signal the end of the one minute resting period. Additional markers are inserted to record the onset of each new trial, the presentation of each pair of distributions, and the user response which signals the end of a single trial. The EEG record is then segmented, using the marker trace, into the resting segment, used as a baseline measurement of brain activity, and 100 single trials. A single trial includes a 2 second resting period used to form inter-trial baseline measurements, followed by the presentation of the distribution pair. Each trial may be of variable length due to

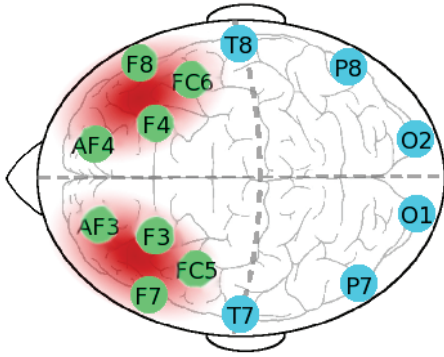


Figure 7: Sensor placement around the prefrontal cortex of the 14 data channels in the Emotiv EEG. The regions in red show the Gaussian weighting used to emphasize the regions of the brain most related to working memory.

reaction time differences, so a window of 1.0 seconds surrounding the user response forms the trial epoch.

Since our cognitive load measure is computed from EEG, care must be taken to account for the spatial organization of the brain. Rowe, et al. discuss the roll of the prefrontal cortex of the brain in various aspects of working memory [RTJ*00]. The spatial activation sites were found to be quite localized; however, EEG experiences volume conduction causing activity generated at a single point to be measured at multiple sensors. To help account for this, spatial averaging was performed using Gaussian weights centered at the prefrontal cortex on each brain hemisphere defined in the 10-20 electrode placement system, as shown in Figure 7. The parametrization of the Gaussian was set to encompass sensors F7 and F3 and their contra-lateral pair F4 and F8 in the first standard deviation. There were no substantial differences between the left and right hemispheres found during later analysis.

5.2. EEG Signal Analysis

The first step in processing the raw EEG signals is to segment the 14 time series (one for each sensor) into individual trials. Next, each trial is divided into the inter-trial baseline and the trial stimulation. Both of these tasks use the markers inserted into the EEG record, as discussed in Section 5.1. The baseline and stimulus signals are then transformed, using the S-Transform to determine the power change and frequency shift induced by the stimulation. These values are used to calculate the cognitive load experienced at each of the 14 sensors for the trial in question. Spatially averaging these 14 values gives a single measurement for cognitive load. Figure 6 shows the workflow of the experiment from data collection through analysis.

5.2.1. Artifact Detection and Removal

Since EEG measures voltages at the scalp, there are many possible sources for data contamination that must be ad-

dress. Artifacts related to eye blinks and other muscle movements in addition to physical movements of the sensors themselves must be removed before the EEG traces can be processed. We have adapted work by Berka, et al. to decontaminate EEG signals generated by Emotiv hardware [BLC*04] and rely on the Emotiv SDK to automatically detect eye blinks. Since muscle contraction and control are generally governed outside of the frequency range of interest [SPK*97], we are able to use frequency band limiting procedures such as low-pass, high-pass and notch filters to adequately remove these signal components. If, after removing EEG artifacts, the energy densities of the alpha or theta frequency bands are changed by more than 20% of their original values, the trial is removed from all further analysis. This criterion is informed by the bad-channel removal method discussed by Anderson, et al. [APS07]. In this study, we threw out 3% of the trials due to excessive signal degradation from movement and 1.5% due to high change in spectral densities, totalling 4.47% of the total trials being removed from further analysis.

5.2.2. Spectral Decomposition of Cognitive Load

In order to understand cognitive load, we must examine the spectral characteristics of the EEG signals. Based on the work of Klimesch [Kli99], we focus our analysis on the alpha (7.5 – 12.5 Hz) and theta (4 – 7.5 Hz) frequency bands, which have been identified as reflecting cognitive and memory performance. We use the S-Transform [Sto07] to decompose the signal into an appropriate time-frequency representation. The S-Transform was chosen over other transformations because it offers adaptive spectral and temporal resolution similar to the Wavelet Transform and is a direct mapping to the complex Fourier Domain.

To be able to properly assess the spectral evolution of EEG associated with working memory, each trial is processed with respect to its own inter-trial rest period. The individual alpha and theta frequencies are determined for both the trial and rest period and their amplitudes measured [Kli99]. By comparing these values, a shift of both the individual frequencies as well as their amplitudes are revealed. The degree of change in these amplitudes, weighted by the amount of shift in the frequency domain, determine the working memory and cognitive load characteristics for each single trial, as described in Equation 2.

Our computation of cognitive load derived from EEG uses the individual mean frequencies in both the alpha and theta frequency bands. The mean frequency is computed as:

$$f(\omega) = \frac{\sum_{i=0}^{n-1} I_{\omega(i)} f_{\omega(i)}}{\sum_{i=0}^{n-1} I_{\omega(i)}} \quad (1)$$

where ω is the frequency band in question, n is the number of frequency bins in ω , f_i is the frequency at bin i and I_i is the energy density of ω at frequency bin i . This formulation

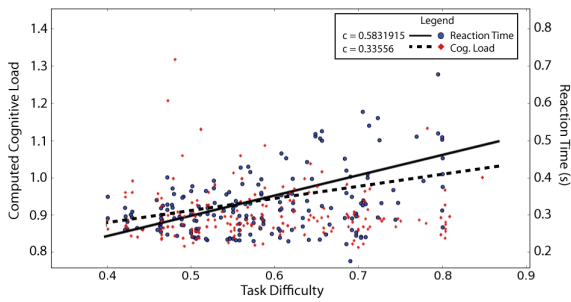


Figure 8: Results from this experiment suggest a correlation between greater task difficulty and higher cognitive load. Here, task difficulty is plotted against computed cognitive load and reaction time for each valid trial across all participants.

of mean frequency is used to compute the frequency shifts in both the alpha and theta wavebands. The frequency shift of a waveband is given by $f_i(\omega) - f_b(\omega)$ where f_i is the frequency content determined from EEG collection during each trial and f_b is the frequency content collected during inter-trial rest periods. Additionally, the change in energy density in a waveband, $\Delta|f(\omega)|$, is the difference of energy densities at the mean frequencies: $\Delta|f(\omega)| = |f_i(\omega)| - |f_b(\omega)|$.

Klimesch identified working memory performance decreases during task-related stimulation expressed as theta power decreases with simultaneous alpha power increases with respect to baseline measurements [Kli99]. We form our model of cognitive load per trial, $L(t)$, as the combination of frequency and power changes in both the alpha and theta bands.

$$L(t) = \Delta|f_i(\alpha)||f_i(\alpha) - \Delta|f_i(\theta)||f_i(\theta) \quad (2)$$

6. Cognitive Load User Study Results

Using direct inspection of brain activity during a visualization task provides us with additional empirical data regarding the effectiveness of different rendering methods. Because EEG measurements are not corrupted by the participant's subjectivity or the benefit of hindsight, as may be the case during post-experiment surveys, they are well-suited for determining the effectiveness of visualization.

Based on our EEG recordings and subsequent analysis, the canonical Box Plot was found to place the least amount of strain on the user's cognitive resources for the task at hand. Table 1 shows the computed cognitive load for each plot type using both Gaussian and constant spatial averaging. The table indicates the Box plot and the Density Plot incurred the lowest cognitive load scores (in bold) using Gaussian and constant weighting, respectively. This result highlights the effect of the spatial averaging on overall cognitive load. Using Gaussian weights helps account for the brain's natural spatial organization, providing a more reliable measure. Interestingly, the Violin and Interquartile plots induced

| | Box | Abbrv. | Interquartile | Vase | Density | Violin |
|----------|--------------|--------|---------------|-------|--------------|--------------|
| Constant | 1.101 | 1.284 | 1.214 | 1.571 | 0.830 | <i>1.619</i> |
| Gaussian | 0.815 | 0.833 | 1.563 | 1.203 | 1.285 | 1.492 |

Table 1: Computed cognitive load for each plot type. Constant and Gaussian spatial averaging are shown. Lowest cognitive load scores are highlighted in bold while highest scores are italicized.

the highest cognitive load (in italics). This may be due to greater visual complexity or the reduction of distinguishable visual elements; however, the validation of such claims warrants additional study.

Reaction time is important in determining working memory performance and capacity [Ste69]. While reaction time cannot measure working memory performance directly, it is an appropriate means of capturing the aggregated performance and capacity of working memory. As the role of reaction time in determining working memory performance is well-explored [APS07, PAS*10], we focus our analysis on the assessment of brain activity via EEG processing.

Figure 8 plots the computed cognitive load and the reaction time from this experiment against the task difficulty for each trial spanning all participants in the user study. The figure suggests correlation between task difficulty and both reaction time as well as the measurement of cognitive load; as the difficulty of the task increases, so does the computed cognitive load and reaction times. However, there is a relatively large variance in both cognitive load and reaction times, particularly in the investigation of high-difficulty tasks. One explanation for this large variance is an incorrect model for task difficulty. The computed task difficulty (Section 5) uses only the median and interquartile range of each distribution. Exploring different formulations for task difficulty may result in a more robust correlation between each trial's computed difficulty and the cognitive load computed. Additionally, our cognitive load measure weights contributions from the alpha and theta frequencies equally. It is possible that a more advantageous combination of theta and alpha spectral changes exist, but adequately exploring the nuances of these formulations is beyond the scope of this paper.

6.1. Statistical Analysis

In order to determine significant correlation between the measured data and visualization type, we employ paired 2-tailed T-tests. T-tests were used to determine significance of spectral properties departing from baseline measurements taken as well as spectral differences between visualization types. All statistical tests used the null hypothesis that there is no significant change between the two distributions being analyzed. Each distribution tested was inspected to verify it was not multimodal prior to analysis.

Table 2 displays the maximum significance values (p-values) as computed for cognitive load by the 2-tailed T-tests. Of particular interest are the high degrees of similarity

between the Box Plot and Abbreviated Box Plot (Box and Abbrev. in Table 2) and the Violin and Interquartile Plots (Violin and Interquartile in Table 2). All tests were performed with cognitive loads computed using Gaussian weights as discussed in Section 5.1.

7. Discussion

In this study, we explored different methods of visualizing distribution data. For each method under consideration, the cognitive load associated with interpreting the interquartile range was determined. While each of the visualizations used for this study displayed the interquartile range of a distribution in some way, not each rendering displayed the same amount of data associated with each underlying data set. For example, the Violin Plot rendered the sample density as described by its histogram whereas the Box Plot did not. These differences enable a different set of questions to be asked about these visualizations that cannot be asked about other visual representations. This study, like others, focuses on the effectiveness of visualization method with respect to a single subset of appropriate interpretation tasks.

Until recently, the expense of EEG technology greatly limited its application in the field of user studies. The Emotiv EPOC headset used in this experiment provided a cost-effective means of EEG acquisition. However, although this system conforms to the international 10-20 standard for electrode placement, getting each electrode in the proper position is important and non-trivial. Additionally, the analysis and interpretation of EEG data remains difficult, requiring training and expertise.

The visualizations and interpretations required during this user study were purposefully chosen to be elementary. The simplicity of this study allowed participants to be chosen from a wide range of potential candidates in order to minimize the potential for schema creation and overrepresentation of germane cognitive load. In addition to controlling germane cognitive load, this decision allowed us to more completely regulate and estimate the contribution of intrinsic cognitive load during each single trial. By acknowledging and controlling these two parameters, we were able to more thoroughly process the resulting data without substantially complicating the analysis.

Minimizing the visualization and task complexity eased requirements for the analysis and processing steps used in this study; however, the experimental design was still difficult. After determining the appropriate visualizations to use during the experiment, finding the proper interpretation task proved to be arduous. Using too simple an interpretation task did not create enough cognitive load to substantially influence working memory performance. Meanwhile, employing too complex a task induced cognitive overload, complicating analysis. Cognitive overload was identified by the movement of the individual alpha frequency outside of the 8–12 Hz band of frequencies, following the results of Klimesch and Gevins, et al. [Kli99, GS00].

| | Box | Abbrev. | Interquartile | Vase | Density |
|---------------|-------|---------|---------------|--------|---------|
| Violin | 0.001 | 0.001 | 0.134 | 0.0015 | 0.0015 |
| Density | 0.001 | 0.001 | 0.003 | 0.002 | x |
| Vase | 0.001 | 0.001 | 0.0015 | x | x |
| Interquartile | 0.001 | 0.001 | x | x | x |
| Abbrev. | 0.216 | x | x | x | x |

Table 2: Pairwise significance values for cognitive load of the Box Plot (Box), Abbreviated Box Plot (Abbrev.), Interquartile Plot (Interquartile), Vase Plot (Vase), Density Plot (Density) and Violin Plot (Violin). While most significance values are below 0.01, some pairs of comparisons generated similar distributions. The Box Plot and abbreviated version score similarly as do the Interquartile and Violin Plots.

Much work has been done to explore the effects of practice on cognitive measures, as the introduction of these effects often confound analysis. Berry, et al. [BZR*09] found that practice does not expand the capacity of working memory and cognition, as was previously thought, but instead improves the efficiency of data encoding. This finding implies that the inverse relationship between available working memory capacity and cognitive load is maintained regardless of practice during an experiment. The spectral dynamics of practice effects in cognition were explored by Gevins, et al. [GS97]. Practice was found to decrease reaction time, but also increase spectral organization. The spectral changes induced by practice comprised an increase in power and frequency modulation *prior* to the task onset. To mitigate the effects of practice in this study, we re-evaluate baseline conditions during the rest period before each trial begins. While this helps minimize the practice effect in analysis for this study, re-evaluating baseline performance may not be possible in more complex, or time-sensitive experiments.

The temporal, spatial, and spectral organization of brain activity enable both analysis and interpretation. Despite an adequate tool set for the processing and general analysis of EEG signals, their interpretation requires domain experts. The multidisciplinary nature of this study was essential for proper examination of the results we collected. Without the close collaboration between computer scientists, neuroscientists, and psychiatrists, the success of this study would have been jeopardized.

8. Conclusions

This work is not the first user study to take cognitive load into account during exploration [RTJ*00], but to the best of our knowledge it is the first study directly measuring brain activity using EEG to study cognitive load across multiple visualization types. Measurements of cognitive load during user studies provide a mechanism for objectively evaluating interpretation difficulty of visualizations. The evaluation method presented here forms the basis for a new and potentially powerful physiological measurement for the evaluation of visualization techniques during user studies.

Although the traditional methods for determining cognitive load are applicable to general user studies, the methods implemented here measure cognitive activity in a more direct manner. By inspecting brain activity to determine cognitive load, we prevent the corruption of the measurements by insights gained after the trial as is possible with many post-hoc methods. Unfortunately, this type of cognitive measure is highly sensitive to the specific tasks presented to the participant. In the case of this study, all tasks focus on determining interquartile ranges resulting in analysis that is valid only with respect to this task. Due to this specificity, it is clear that cognitive load derived from EEG is more difficult to apply to user studies of more complex tasks that cannot be simply divided.

Such specificity in user studies is not a new or unexpected result [KHI*03]. In this view, user studies should be used to measure specific relationships of visualization and perception. This work adds to this paradigm of user evaluations by contributing an additional measure relating visualization to cognition. However, because the study of brain activity through EEG is itself complex, its direct application to the evaluation of broad or complex visualization tasks may be limited. We foresee the greatest impact of this work to be in the assessment and subsequent evaluation of specific choices within a single visualization technique.

9. Future Work

Additional studies exploring the relationship between cognition, working memory, and the visual system may provide further insights into human factors in scientific visualization. Such studies would require the quantification of visual complexity, and focus on both the working memory centers and the visual system [EBJ*88].

This study presents a basis on which other studies may build. Of particular interest to the visualization community is the investigation of cognitive load from more advanced visualization techniques. Additional experiments will investigate the same data representation methods used in this study with respect to a wider range of interpretation tasks. Also, future experiments will be designed to incorporate 2 and 3-dimensional scalar and vector fields to determine the cognitive differences associated with each visualization technique. Additionally, studying cognitive implications of visualization with respect to a large collection of specific tasks may result in a more profound understanding of the cognitive effects of more complex systems not directly addressable by user studies involving EEG.

Other techniques have shown promise in the measurement of cognitive performance. Eye tracking and pupillary responses [KTH11] may provide additional insights into cognitive load with respect to visualization studies. Future experiments must be performed to properly determine the benefits and drawbacks associated with each approach to physiological measurements, with particular attention given to the appropriate application of the different techniques.

This user study framework can also be applied to the design of new visualization techniques. By examining extraneous and visual cognitive loads during the development of visualization methods, more optimal design choices may become apparent. By examining and minimizing the overall cognitive load associated with new visualization techniques, methods may be developed that are more easily adopted by domain-specific users and students first learning the science.

Acknowledgements The authors would like to thank the anonymous reviewers for their insightful comments. We also thank Dr. Laura McNamara for discussions on experimental design, and Dr. Joel Daniels II for his help and advice. This work was supported in part by grants from the National Science Foundation (IIS-0905385, CNS-0855167, IIS-0844546, ATM-0835821, CNS-0751152, OCE-0424602, CNS-0514485, IIS-0513692, CNS-0524096, CCF-0401498, OISE-0405402, CCF-0528201, CNS-0551724, CNS-0615194), Award No. KUS-C1-016-04, made by King Abdullah University of Science and Technology (KAUST), the Department of Energy, and IBM Faculty Awards.

References

- [APS07] ANDERSON E. W., PRESTON G. A., SILVA C. T.: Towards development of a circuit based treatment for impaired working memory: A multidisciplinary approach. *IEEE/EMBS Conference on Neural Engineering* (2007), 302–305.
- [Bad83] BADDELEY A.: *Working Memory*. Oxford University Press, 1983.
- [Bad92] BADDELEY A.: Working memory: The interface between memory and cognition. *J. of Cognitive Neuroscience* 4, 3 (1992), 281–288.
- [BCN*97] BRAVER T. S., COHEN J. D., NYSTROM L. E., JONIDES J., SMITH E. E., NOLL D. C.: A parametric study of prefrontal cortex involvement in human working memory. *NeuroImage* 5, 1 (1997), 49–62.
- [Ben88] BENJAMINI Y.: Opening the box of a boxplot. *The American Statistician* 42, 4 (1988), 257–262.
- [BLC*04] BERKA C., LEVENDOWSKI D. J., CVETINOVIC M. M., PETROVIC M. M., DAVIS G., LUMICAO M. N., ZIVKOVIC V. T., POPOVIC M. V., OLMSTEAD R.: Real-time analysis of eeg indices of alertness, cognition and memory with a wireless eeg headset. *Int'l J. of HCI* 17 (2004), 151–170.
- [BLR*05] BERKA C., LEVENDOWSKI D. J., RAMSEY C. K., DAVIS G., LUMICAO M. N., STANNEY K., REEVES L., REGLI S. H., TREMOULET P. D., STIBLER K.: Evaluation of an eeg-workload model in the aegis simulation environment. *Proceedings of SPIE* (2005), 90–99.
- [BZR*09] BERRY S., ZANTO T. P., RUTMAN A. M., CLAPP W. C., GAZZALEY A.: Practice-related improvement in working memory is modulated by changes in processing external interference. *J. of Neurophysiology* 102 (2009), 1779–1789.
- [CM84] CLEVELAND W. S., MCGILL R.: Graphical perception: Theory, experimentation and the application to the development of graphical methods. *J. of American Statistical Association* 79 (1984), 531–554.
- [CPB*97] COHEN J. D., PERLSTEIN W. M., BRAVER T. S.,

- NYSTROM L. E., NOLL D. C., JONIDES J., SMITH E. E.: Temporal dynamics of brain activation during a working memory task. *Nature* 386 (1997), 604–608.
- [CS91] CHANDLER P., SWELLER J.: Cognitive load theory and the format of instruction. *Cognition and Instruction* 8 (1991), 293–332.
- [CS98] CARPENTER P. A., SHAH P.: A model of the perceptual and conceptual processes in graph comprehension. *J. of Experimental Psychology: Applied* 4, 2 (1998), 75–100.
- [CW04] CONSTANTINIDIS C., WANG X.-J.: A neural circuit basis for spatial working memory. *The Neuroscientist* 10, 6 (2004), 553–565.
- [EBJ*88] ECKHORN R., BAUER R., JORDAN W., BROSCHE M., KRUSE W., MUNK M., REITBOECK H.: Coherent oscillations: A mechanism of feature linking in the visual cortex? *Biological Cybernetics* 60, 2 (1988), 121–130.
- [Eng02] ENGLE R. W.: Working memory capacity as executive attention. *Current Directions in Psychological Science* 11 (2002), 19–23.
- [GS00] GEVINS A., SMITH M. E.: Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. *Cerebral Cortex* 10, 9 (2000), 829–839.
- [GSMY97] GEVINS A., SMITH M. E., McEVOY L., YU D.: High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. *Cerebral Cortex* 7 (1997), 374–385.
- [HKD*99] HITT J. M., KRING J. P., DASKAROLIS E., MORRIS C., MOULOVA M.: Assessing mental workload with subjective measures: An analytical review of the NASA-TLX index since its inception. *Human Factors and Ergonomics Society Annual Meeting* 43 (1999), 1404–1404.
- [HN98] HINTZE J. L., NELSON R. D.: Violin plots: A box plot-density trace synergism. *The American Statistician* 52, 2 (1998), 181–184.
- [KHI*03] KOSARA R., HEALEY C. G., INTERRANTE V., LAIDLAW D. H., WARE C.: Thoughts on user studies: Why, how and when. *IEEE Computer Graphics and Applications* 23, 4 (2003), 20–25.
- [Kli99] KLIMESCH W.: Eeg alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews* 29 (1999), 169–195.
- [KMH01] KOSARA R., MIKSCH S., HAUSER H.: Semantic depth of field. *Proceedings of IEEE INFOVIS* (2001), 97–104.
- [KTH11] KLINGNER J., TVERSKY B., HANRAHAN P.: Effects of visual and verbal presentation on cognitive load in vigilance, memory and arithmetic tasks. *J. of Psychophysiology* 48, 3 (2011), 323–332.
- [Lik32] LIKERT R.: A technique for the measurement of attitudes. *Archives of Psychology* 140 (1932), 1–55.
- [LKJ*05] LAIDLAW D. H., KIRBY R. M., JACKSON C. D., DAVIDSON J. S., MILLER T. S., DA SILVA M., WARREN W. H., TARR M. J.: Comparing 2d vector field visualization methods: A user study. *IEEE Transactions on Visualization and Computer Graphics* 11, 2 (2005), 59–70.
- [PAS*10] PRESTON G. A., ANDERSON E. W., SILVA C. T., GOLDBERG T., WASSERMANN E. M.: Effects of 10 Hz rTMS on the neural efficiency of working memory. *J. of Cognitive Neuroscience* 22, 3 (2010), 447–456.
- [PCVDW01] PIROLLO P., CARD S. K., VAN DER WEGE M. M.: Visual information foraging in a focus + context visualization. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 2001), CHI '01, ACM, pp. 506–513.
- [PKRJ10] POTTER K., KNISS J., RIESENFELD R., JOHNSON C. R.: Visualizing summary statistics and uncertainty. *Computer Graphics Forum* 29, 3 (2010), 823–831.
- [PRS03] PAAS F., RENKL A., SWELLER J.: Cognitive load theory and instructional design: Recent developments. *Educational Psychologist* 38, 1 (2003), 1–4.
- [RBBS06] REITINGER B., BORNIK A., BEICHEL R., SCHMALSTIEG D.: Liver surgery planning using virtual reality. *IEEE Computer Graphics and Applications* 26 (2006), 36–47.
- [RTJ*00] ROWE J. B., TONI I., JOSEPHS O., FRACKOWIAK R. S. J., PASSINGHAM R. E.: The prefrontal cortex: Response selection or maintenance within working memory. *Science* 288, 5471 (2000), 1656–1660.
- [SJB07] SEUFERT T., JÄNEN I., BRÜNKEN R.: The impact of intrinsic cognitive load on the effectiveness of graphical help for coherence information. *Computers in Human Behavior* 23 (2007), 1055–1071.
- [SPK*97] SALENIUS S., PORTIN K., KAJOLA M., SALMELIN R., HARI R.: Cortical control of human motoneuron firing during isometric contraction. *J. of Neurophysiology* 77, 6 (1997), 3401–3405.
- [SRT*07] SHI Y., RUIZ N., TAIB R., CHOI E., CHEN F.: Galvanic skin response (gsr) as an index of cognitive load. In *CHI '07 extended abstracts* (2007), ACM, pp. 2651–2656.
- [Ste69] STERNBERG S.: Memory scanning: Mental processes revealed by reaction-time experiments. *American Scientist* 57 (1969), 421–457.
- [Sto07] STOCKWELL R. G.: A basis for efficient representation of the s-transform. *Digital Signal Processing* 17, 1 (2007), 371–393.
- [SW94] SARTER N. B., WOODS D. D.: Pilot interaction with cockpit automation ii: An experimental study of pilots' model and awareness of the flight management system. *Int'l J. of Aviation Psychology* 4, 1 (1994), 1–28.
- [Swe05] SWELLER J.: *The Cambridge handbook of multimedia learning*. Cambridge University Press, 2005, ch. Implications of cognitive load theory for multimedia learning, pp. 19–30.
- [SZB*09] SANYAL J., ZHANG S., BHATTACHARYA G., AMBURN P., MOORHEAD R. J.: A user study to compare four uncertainty visualization methods for 1d and 2d datasets. *IEEE Transactions on Visualization and Computer Graphics* 15, 6 (2009), 1209–1218.
- [TM04] TORY M., MÖLLER T.: Human factors in visualization research. *IEEE Transactions on Visualization and Computer Graphics* 10, 1 (2004), 72–84.
- [Tuf83] TUFTE E. R.: *The Visual Display of Quantitative Information*. Graphics Press, 1983.
- [Tuk77] TUKEY J. W.: *Exploratory Data Analysis*. Addison-Wesley, 1977.