Impact of Time-based Visualization on Situation Awareness

Mariëlle den Hengst
Delft University of Technology
The Netherlands
Email: m.denhengst@tbm.tudelft.nl

Michael McQuaid
University of Arizona
Tucson, AZ, USA
Email: mmcquaid@eller.arizona.edu

Jiang Zhu
University of Arizona
Tucson, AZ, USA
Email: jiangzhu@cs.arizona.edu

Abstract

We have developed a time-based visualization system and tested it in accordance with a well-developed theory of situation awareness during a simulation of an actual emergency. The client-server system is fed by a collaborative server, and uses linguistic parsing, suffix tree clustering, and multidimensional scaling to develop a changing picture of textual data related to an unfolding event. We provide a detailed account of situation awareness theory and use it to explain the effect of the visualization on perceptions of facts, comprehension of the situation, and projection of near term status. We developed a simulation of a tragic shooting at the University of Arizona, performing requirements elicitation of officials involved in the actual event, and preparing a repeatable scenario based on, but differing from, the event. We ran the simulation with about 200 management students, half of whom used the visualization system and half of whom used a competing text-based system, and found that the visualization system aided comprehension and projection, but had no effect on perception of facts.

1 Introduction

Most visualization research focuses on information that can be expressed numerically. Nevertheless, a growing literature treats visualization of textual data. Visualization systems offer appealing summaries of textual data, but we need to know about the impacts of these systems on work processes.

In particular, we would like to be able to understand and explain the impact visualization has on managerial work in terms of a well-tested theory. Over the past two decades, human factors researchers have built an impressive body of work about situation awareness, initially mostly in aviation, and this work has recently spread to other areas, particularly areas where workers interact with computer-based systems that represent the state of the world in some way. These characteristics suggest that visualization may be well partnered with situation awareness for an explanatory analysis.

Our approach has been to develop a candidate visualization system and test according to the theory of situation awareness. In succeeding sections, we describe visualization, our implementation, our experiment, our results, and our conclusions.

2 Visualization

Visualization offers a powerful framework for partnering information technology with the human brain, about half of which is devoted to vision processing. The term visualization was coined in a seminal 1987 paper [21] advocating a federally funded initiative to support a research stream, visualization in scientific computing, already recognized as distinct from computer graphics, but unnamed for several years. The notion of visualization for this question and related work is based on a definition suggested by [4], where some of the discussion and citations might lead to the following definition (not directly stated in the article): *Visualization is transformation and analysis to aid in the formation of a mental picture of symbolic data. Such a picture is simple, persistent, and complete.*

Taxonomies of visualization, such as [3], typically classify visualization based on

- available technology,
- dimensionality of representation,

- scale (e.g., continuous / discrete) of visualized material,
- scope of visualized material, and
- domain of visualized material.

The first two classes are closely linked: 3D visualization didn't become prevalent until special hardware became widely available. The third class is also closely related as the data and the number of values data can take on informs the dimensionality. The most varied data typically visualized is unstructured text, which is also most resistant to increasing scale in the sense of dataset size. Scope in this context refers to scope of control, such as desktop (a single user's visualization of material under his/her control), workgroup (material under a manager's control), enterprise (material under control of managers with congruent goals but different and possibly conflicting subgoals), and public (such as all websites about automobiles).

One effect of such taxonomies is to see where research has proliferated. The class called "scale" above has a very skewed sample—with much more work available on data that can take on more limited values, better understood data, numeric data. Much visualization research has explored quantitative information, with many opportunities remaining for visualizing text information. Of the fewer initiatives for visualizing text, many have focused on maximizing the size of the target data set, often in conjunction with the question of whether we can visualize the entire WWW. One outcome of this kind of research has been the recognition that such large-scale visualizations are extremely imprecise and, though often beautiful in appearance, of limited use in the workplace.

Categorization has been a focus of cognitive science as the fundamental activity used by humans seeking to understand their environment. It may be the case that a tool that helps humans to categorize phenomena may help them to understand those phenomena. Visualization has been used to aid in categorization of material from face-to-face electronic meetings [22].

There is not clear evidence that visualization systems for text have offered value to managers. For example, [16] calls for "new ideas about how to display large, abstract information spaces intuitively." She asserts that "clustering can result in the juxtaposition of very different levels of description within a single display."

Hearst offers a pessimistic review of evaluations of graphical overviews, saying for instance that "Although intuitively appealing, graphical overviews of large document spaces have yet to be shown to be useful and understandable for users. In fact, evaluations that have been conducted so far provide negative evidence as to their usefulness."

Many researchers in Management Information Systems have tried to explain the relationship between graphical displays and decision making, such as [32], [33], and [7].

One important omission from the above work is a larger view of tasks involving decision making. [8, p. 466] contrasted the breadth of focus on decision making between two cultures, describing Japanese managers as focusing on framing the problem, and American managers as forgetting to do so and being content on solving problems that have already presented themselves in unambiguous terms. [34] expands on this notion, describing a framework for making sense of a situation. [23] see a major limitation of Weick's approach. Weick does not model any kind of organizational learning or knowledge creation or knowledge management in the sensemaking approach. The organization is in a state of perpetual confusion, punctuated by flashes of illumination about individual situations.

Another way to frame visualization requirements is to consider the key questions recognized in the information visualization community. [3, p. 640] lists eight key issues. Three of these issues relate to visualization systems for managers: collaborative visualization, characterization of information visualization at the operator level, and perceptual analysis of dynamic information displays. Collaborative visualization can be addressed by structuring the visualization to support a collaborative process like emergency response, operators can include icons and their relationships, and perceptual analysis can be addressed in terms of situation awareness, which has been empirically linked to decision making quality.

The visualization system proposed here depicts relationships between texts in an emergency response scenario, shown in a manner intended to ease the managerial burden of comprehension as the situation, and the texts reporting the situation, evolve.

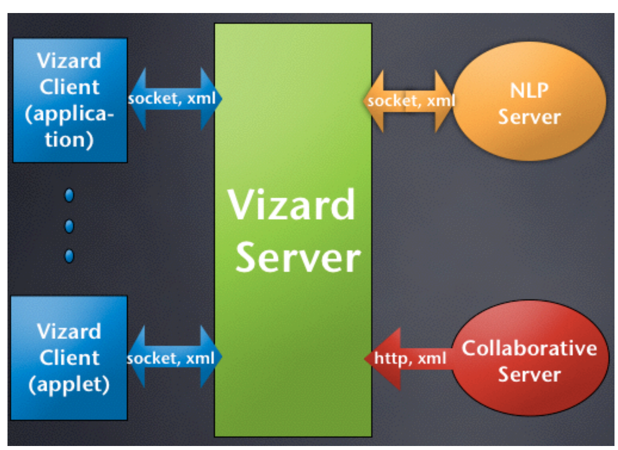

Figure 1. Implementation overview.

3 Implementation

Our implementation of a time-based text visualization system is a client-server system called Vizard. The server can be characterized by four main features, two of which are developed externally to this project, as shown in figure 1. These external features are a collaborative computing system and a natural language processing system. The internal features are characterized by their two main algorithms, suffix tree clustering and multidimensional scaling, located in the clustering job controller and clustering engine shown in figure 2. We discuss these four features in turn.

For this experiment, text input came from a subsystem we developed to simulate a forthcoming collaborative server by GroupSystems.com. This is an interrupt-driven system that injects XML-formatted texts at intervals determined by timestamps on the texts, as if the texts were received from the GroupSystems.com server in its XML format.

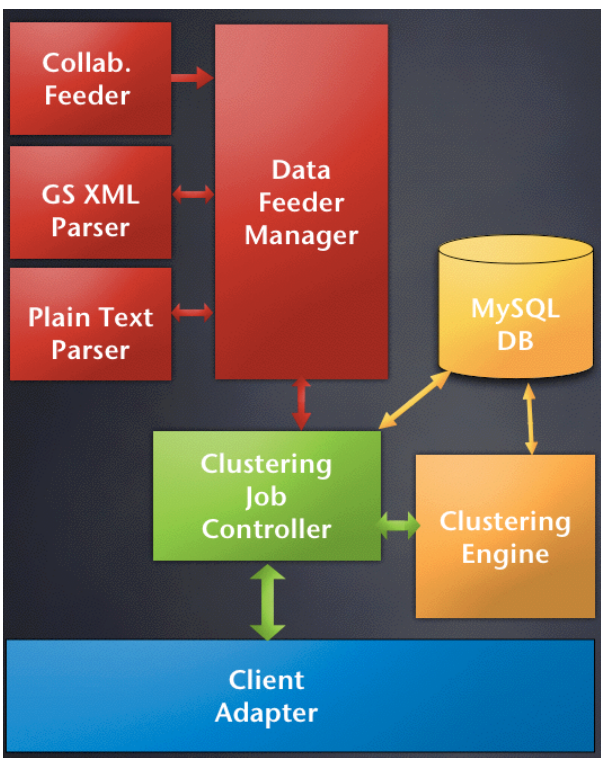

Figure 2. Implementation server overview.

We communicate with a linguistic analysis server, Connexor FDG (Functional Dependency Grammar) developed by Connexor OY of Helsinki, Finland. This tool provides several text parsing facilities, including dependency functions, functional tagging, surface syntactic tagging, morphological tagging, and various other facilities, including named entity recognition. The tool and its antecedents are described in many studies, such as [29] and [28].

Parsed texts are clustered, using the suffix tree clustering algorithm of [36], whose main features include $\mathcal{O}(N)$ time complexity for construction of the suffix tree—each incoming text is added to its appropriate place in the tree without recomputing—and cluster overlap, allowing a text to appear in more than one cluster.

The suffix tree clustering process is as follows.

- Regard each text as a string.
- Build a suffix tree of these strings.
- Each leaf contains an ID and a starting position within the original string.
- Regard every node with 2 or more leaves as a "base cluster."
- Assign a score to each base cluster, B, based on the cluster size, $|B|$ and the length of the cluster label, $|P|$

$$s(B) = |B| \cdot f(|P|)$$

where

$$f(|P|) = \begin{cases} 1, & \text{if } |P| = 1 \\ 1/|P|, & \text{if } 2 \le |P| \le 6 \\ 1/7, & \text{otherwise} \end{cases}$$

- Combine base clusters whenever

$$\left(\frac{|B_m \cap B_n|}{|B_m|} > 1/2\right) \wedge \left(\frac{|B_m \cap B_n|}{|B_n|} > 1/2\right)$$

for every pair of clusters m and n.

Finally, each set of clusters at the same level is displayed in a 2D array computed by a non-metric multidimensional scaling algorithm, described in [22] as follows.

- Prepare a matrix of dissimilarities between texts using the typical Jaccard coefficient described in [25, section 16.2.2]

$$S_{D_i,D_j} = \frac{\sum_{k=1}^{L}(w_{ik}w_{jk})}{\sum_{k=1}^{L} w_{ik}^2 + \sum_{k=1}^{L} w_{jk}^2 - \sum_{k=1}^{L}(w_{ik}w_{jk})}$$

where S_{D_i,D_j} is the similarity between texts i and j. D is a text, w_{ik} is the weight applied to collocation k in text i, which is usually either the number of occurrences of the collocation in the text or [0, 1] signifying the presence or absence of the collocation. L is the total number of collocations in a text. Each dissimilarity is $1 - S_{D_i,D_j}$. Collocations are text objects identified by Connexor FDG such as noun phrases, named entities, and idiomatic expressions.

- Obtain an initial configuration using singular value decomposition (SVD) as described in [13].
- Measure the information loss between the original matrix and the initial configuration using Kruskal's STRESS metric [19], defined as

$$S = \sqrt{\frac{\sum(d_{ij} - \hat{d}_{ij})^2}{\sum d_{ij}^2}}$$

  where the d_{ij} are inter point distances in the present configuration, and the $\hat{d}_{ij}$ are monotonically related to the values in the original dissimilarity matrix δ_{ij} so that $\hat{d}_{ij} \leq \hat{d}_{i'j'}$ whenever $\hat{\delta}_{ij} < \hat{\delta}_{i'j'}$.
- Find a new configuration with smaller information loss than the initial configuration using an isotonic regression algorithm [15] to obtain fitted distances and a conjugate gradient descent algorithm from [24] to optimize.
- Repeat until a threshold information loss is reached or until a threshold number of iterations are performed.

The resulting displays are available to the client, which is implemented both as a Java applet and as a Java application. The client offers numerous capabilities, of which the following are relevant here. A left hand side pane provides a hierarchical list of the all the clusters and texts within the those clusters. The list can be arranged by timestamp or alphabetically. A right hand side pane shows the 2D array described above for either the root or any selected sub-cluster. The user can navigate the texts clicking on either pane, but the right hand side pane only allows drilling deeper, since only one level is shown at any given time. The left hand side pane shows the entire hierarchy at all times, any of which can be clicked on. Figure 3 shows a screenshot of a portion of the client's right hand side pane during a military exercise. Each cluster is labeled by the two most distinguishing collocations in the cluster, separated by a comma and followed by the number of texts in the cluster in parentheses.

4 Situation Awareness

4.1 Introduction

Having discussed the concept of visualization in the previous section, we now turn to the concept of the situation

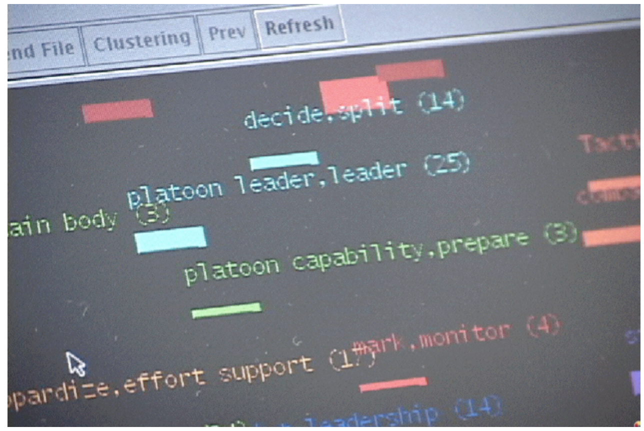

Figure 3. Client screenshot during a military exercise.

being portrayed by the visualization. Specifically, the visualization is of the relationships between texts pertinent to a situation. This concept of situation leads to a way to evaluate the effectiveness of the visualization.

Information Science as a discipline concerns itself with information seeking behavior in humans, so it makes sense to look here for insight into the notion of what a situation is, since the visualization we construct can be construed as a vehicle for information presented to an information seeker.

Situation as a concept began to attain prominence in Information Science only in the past few years. [5] notes that she provides the first review of the concept in Information Science literature. Cool describes six major theoretical perspectives on the situation concept, the Problematic Situation, Social Interaction Theory, Situated Action, Situation Awareness, the Person-in-Situation Model, and Situation Environments. We select one of these, Situation Awareness, that best fits the evaluation of visualization. We describe that concept in detail at the theoretical level and conclude with a technique for measuring it.

4.2 Theory of Situation Awareness

[12] presents a theoretical model of situation awareness (SA). Following is a discussion of this model with respect to the present study. Endsley frequently studies flight crews, but strives to extend the model to other fields and convincingly presents constructs that may be generalized to other fields. Endsley does report on SA studies of operators of flexible manufacturing systems, refineries, and power plants. Other SA researchers have studied firefighters, police units, military command personnel, and physicians [14].

Of particular relevance to the present study is the research on tactical commanders [18], and its finding that "recognizing the situation provided the challenge to the decision maker" [18, p. 1110].

A wide body of research exists to link problem framing or problem presentation to decision making processes [2, 17, 30, 31], as well as research showing that decision makers act first to classify and understand a situation before proceeding to action selection as reviewed by [20]. This prior research motivates a model defining the SA constructs as closely as possible to the constructs in the prior research, benefitting from the demonstrated links.

[12] further develops a previously described model [9, 10], which defines SA as a state of knowledge, and divides that state into three levels, perception of elements in current situation, comprehension of current situation, and projection of future status. This construct explicitly excludes less dynamic knowledge, such as doctrine, rules, procedures, and templates. SA affects decision making, which affects performance of actions. SA is affected by the actual state of the environment, as well as task / system factors and individual factors. Task / system factors include system capability, interface design, stress, workload, complexity, and automation. Individual factors include goals and objectives, preconceptions, long term memory stores, information processing mechanisms, automaticity, abilities, experience, and training. Figure 4 depicts the relationships between these constructs.

This richly delineated model gives the lie to [35, p. 4], who claimed that situation awareness was the buzzword of the nineties, without a clear definition, and marked a turning point from scattered recognition of like ideas, to the basis for a research agenda.

[12] distills the model into the following narrative definition: situation awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future.

In the present study, we simulated a tragic shooting at the University of Arizona. Subjects were asked to prepare the University president to make a public statement based incoming information about the shooting. In this case, the first level includes facts of the situation such as that shots have been fired, that the shots were fired in the College of Nursing building, and so on. The second level includes comprehension of the meaning of these facts, such as that the shooter may present a danger to the entire University community. The third level includes projections we can make, that the police will either apprehend the shooter or that the shooter will escape, and that any new reports of violence in the vicinity of the College of Nursing may be connected, and that we can expect to see frantic movement by large numbers of students away from the building.

Hence, when a report disputes the police version of the event, saying that a car has backfired, rejection of that report is a matter of comprehension. When a report insists that the evacuation of students is actually a campus riot, rejection of that report is a matter of comprehension. Recognition of the need to reassure and inform members of the campus community is a matter of projection, as is recognition of the need to notify next of kin of University community members slain.

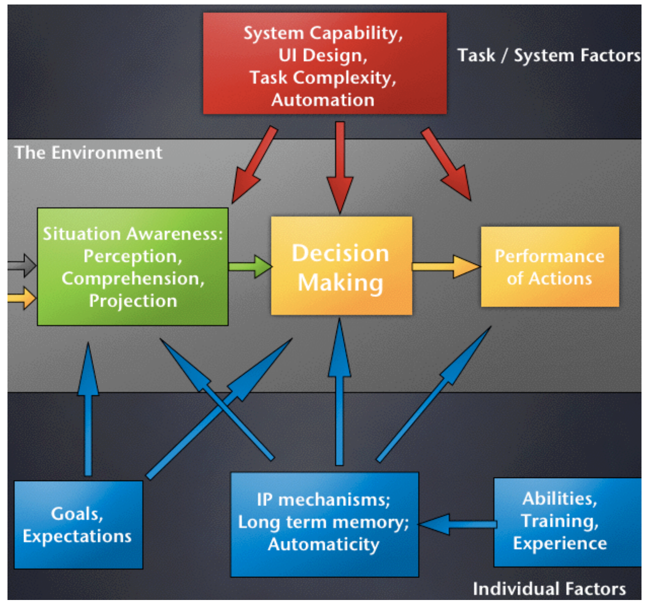

Figure 4. Situation Awareness Model.

4.3 Measurement of Situation Awareness

[11] provides a comparison of measurement techniques and seeks to establish the validity of Situation Awareness Global Assessment Technique (SAGAT). This persuasive study, in part, led to the adoption of SAGAT for the present study. Here, we will review only the use of questionnaires.

[11] treats questionnaires in general, then delves into the author's contribution, Situation Awareness Global Assessment Technique, or SAGAT. Questionnaires can be classified according to when they are administered, either after or during the activity. Questionnaires administered during the activity may be administered either without interruption of the activity, or while the activity is frozen. Obviously, the last class can only be conducted in a simulation. Post facto questionnaires suffer from the obvious limitation that much of what is sought consists of the contents of working memory which has changed after the conclusion of the activity and is notoriously difficult to recall in any circumstances. In particular, when misperceptions are corrected during the activity, successful performance may dictate immediately discarding them. Endsley concludes that a posthoc question-

naire can only reliably provide evidence about the subject's SA at that point in time.

Questionnaires administered during an activity are intrusive in that they add to the workload of the activity and redirect attention during the activity. The remedy proposed by Endsley is to freeze the activity for a questionnaire based on a requirements analysis of the tested activity.

Endsley conducted two studies to validate SAGAT. These studies investigated (1) the period of time during which a questionnaire remains an effective measure of SA before changes in working memory, as discussed above, alter the outcome, and (2) whether the freeze technique is intrusive, altering the subject's SA. The first study found that SAGAT could extract SA for up to 5 or 6 minutes after beginning a freeze. The second study was not able to find any impact on performance from freezes.

Can these results be generalized to include other cases, such as the present study? Both of Endsley's experiments involved pilots and aviation SA. Aviation is the original domain of SA research and the most extensively reported in literature. SA is relatively straightforward and deeply studied in this domain, consisting mostly of spatial information and information represented as numerical measurements. In the present study, and its more general parent, "crisis situations," plenty of situation elements are spatial or interpreted as numerical measurements, but the vast majority of signals consist of textual information. This is the key difference between the present study and the study validating SAGAT, so it's reasonable to ask whether there is anything about words that may change the outcome.

Luckily, we can appeal to recent psychological research, such as [27], suggesting that any differences we find will not be due to differences in human processing of words and numbers. [6] found, using magnetic resonance imaging, that numerical thinking such as performing arithmetic uses the same brain centers as tasks involving only words. Further, they discovered that students forced to attempt mathematical reasoning in second languages were markedly disadvantaged.

[26] endeavors to situate natural language in a central role in differentiating humans and human thought from other creatures and whatever thought they may possess, drawing on the above-mentioned research.

Spelke recounts numerous studies showing that performance of humans and non-human animals diverge only upon the acquisition of language. Non-human animals possess similar, frequently better core systems including numerosity, geometry, object mechanics. The result is a convincing case that the combinatorial properties of natural language allow humans to combine concepts from more than one core system. Of particular interest to the present study, the subsystem involving numerosity can produce no better performance than rats, who can evidently count to precisely 16 to obtain a food pellet, until the child begins to acquire language. This matters a great deal to the present study because the bulk of the early study of SA is in aviation where measurement of SA has been represented as numerical and spatial information. Since we can conclude that natural language is essential to the use of core cognitive systems, we can also conclude that we can measure the same processing engine without recourse to any specific core system.

This is not to say that measurements will be the same, any more than one candidate's verbal and math SAT scores will be the same. All that this notion of the centrality of natural language means is that we're measuring the same processing engine doing two different things and accessing two different resources.

5 Experiment

5.1 Introduction

Having described a system and a theory for testing the system in previous sections, we turn now to a description of the actual experiment. Our overall goal is to to implement the theory for a legitimate managerial task in such a manner as to legitimately evaluate Situation Awareness during the task. So our operational goals, described in the next two sections, are to (1) faithfully develop the task through requirements elicitation, and to (2) faithfully develop the evaluation.

A tragic incident on the University of Arizona campus formed the basis for faithfully developing the managerial task. In 2002, a failing University of Arizona nursing student shot and killed two nursing professors during classes, and brought University operations to a halt as students fled the Nursing school and surrounding buildings, and a police SWAT team took control of the area searching for the shooter, who took his own life.

In the aftermath of this crisis, we interviewed numerous University officials and concluded that the office of the University president faced a managerial task that could be easily simulated and easily understood by management students on campus.

5.2 Requirements Elicitation

SAGAT (Situation Awareness General Assessment Technique) described previously, calls for requirements elicitation, where experts in the field describe salient cues and their impact on situation awareness. These descriptions also provide guidelines for determining what are *not* salient cues and what are *not* appropriate impacts on situation awareness.

Our approach to this requirements elicitation was to conduct unstructured interviews with officials, asking them about incoming information and its impact on their actions. We were assisted by the presence of on-campus forums about the crisis, for which several officials provided notes of phone calls, partial printouts of email messages, and in one case, a phone bill providing time and duration of calls.

Conveniently, officials readily discussed their perception of facts, their comprehension of facts, and projections of future states based on the facts, corresponding neatly to the three theoretical levels of situation awareness. As a result, it is easy to organize our discussion of their requirements by level.

We then used our notes and campus maps to construct artificial records of incoming email and phone calls to resemble both the messages containing salient cues and the messages containing obfuscation of salient cues, erroneous information or erroneous inferences drawn from information.

The problem can be seen as a classification problem, where each item in a list can be classified as relevant, irrelevant, or incorrect, and as a perception, comprehension, or projection. Space prohibits an exhaustive treatment in a conference paper, so examples are given below.

5.2.1 Perception

Relevant perceptions of the event can be classified according to attributes of the victim(s), the shooter(s), the location(s), response(s) by officials or others, and witnesses or bystanders.

Example: victim(s). Major facts about the victims include their number, their status (injured or deceased), and their relationship to the university. There were also relevant facts about the location of victims when shot, and the sex of the victims.

5.2.2 Comprehension

Relevant comprehensions of the event can be similarly classified according to attributes of the victim(s), the shooter(s), the location(s), response(s) by officials or others, and witnesses or bystanders, but also the wider University community, including the geographical relationship of the community to the shooting location.

Example: summarizing responses. Many police responses can be regarded as presenting a snapshot summary of the situation as known by the University Police. Police responses indicating that they control a given area or are searching a given area, or providing other details can be used to infer degree of public safety and as a check against other sources of information because they are presumed to be more reliable and to be focused on relevant issues.

5.2.3 Projection

Projection was related to specific salient cues, including immediate physical safety, breakdown of the chain of command, breakdown of communications media (cell phone / landline), breakdown of language communications, erroneous and redundant communications, emergency procedures, and triage.

As a salient cue example, we describe language breakdown messages. Messages describing difficulties between speakers of Spanish or Chinese and English prompted the realization that multilingual announcements were necessary, something not obvious beforehand because the University's official language is English. It turns out that when people hear gunfire, the panic runs so deep that the official language can't be expected to communicate emergency information to everyone.

5.3 Experimental Design

To develop the evaluation, we developed a method from the Situation Awareness literature previously discussed, and described in detail in the Experimental Design section.

The experimental subjects were all management students taking a required upper division information technology course.

The subjects were randomly assigned to one of two equally sized groups. Each group worked for 45 minutes. One group received incoming information through the visualization system, while the other group received information from a text-based system. Both systems were fed exactly the same information from the same server at the same time. Both systems allowed access to all previously disseminated information. The text based system could *only* present texts in time order, whereas the visualization system presented the same texts in a hierarchical outline form by default in the left pane, and colored blocks representing clusters of texts in the right pane. The block colors had no meaning except to simplify the associations between labels and blocks in congested areas.

All subjects had access to GroupSystems, a collaboration tool, and a web browser, preset to two windows. All subjects were presented with brief verbal descriptions of the available tools, and a monitor walked around the room, answering tool-oriented questions, but refraining from answering any content questions. Every five minutes, a camera sweep captured the active windows for each subject. Four times during the simulation, participants were asked to stop their activities and answer a SAGAT questionnaire. At the end of the simulation, subjects were asked to complete a SART questionnaire after an additional SAGAT questionnaire.

The subjects were instructed to complete the following task during the simulation. Subjects were to consider themselves as advisors to the University of Arizona President, and to pretend that they were occupying a conference room in the President's office at the beginning of the crisis. Their task was, understanding that the President had to make a public appearance in an hour, to provide President with three lists he would need for his appearance. First, he would need a list of the relevant facts of the crisis as we know them at the time he must speak. Second, he would need a list of the relevant stakeholders whose needs he must address. Third, he would need a list of the most important topics to address.

Collaboratively, the subjects completed these three lists at the end of the simulation. In addition, following the notion pioneered in [1], the subjects completed in process versions of these three lists individually, four times during the simulation as their SAGAT freezes, taking two minutes for each freeze.

We use these lists to test three hypotheses as follow.

H_1 Subjects using the visualization system will score about the same on questions of perception as subjects using a text-based system, since the visualization system does nothing to highlight facts.

H_2 Subjects using the visualization system will score higher on questions of comprehension than subjects using a text-based system, since the visualization system organizes information into clusters, aiding the comprehension process.

H_3 Subjects using the visualization system will score higher on questions of projection than subjects using a text-based system, since the projection process can rely on the same clusters that have been helpful in comprehension.

Based on interviews with officials, we developed a simulation scenario consisting of messages to be released by the server at the specified intervals during the simulation. Both systems received 142 messages in a 45 minute time period.

6 Results

Lists for each subject and each freeze were scored by the experimenters according to the information gathered in the requirements elicitation phase. To date, about a thousand of the six thousand collected statements have been processed and compared to statements gathered in requirements elicitation. Each statement was scored 1 to signify conformance, 0 to signify irrelevance, or -1 to signify erroneous situation awareness.

Testing was grouped by freeze and level, since empirical studies of situation awareness have consistently shown statistical independence on these attributes.

Statistical tests of the scores were conducted using the R statistical language. The main R outputs are reproduced below.

The following t-test is for the difference in *perception* scores for the visualization users (V) and the text users (T). We do not see evidence that they differ.

```
Welch Two Sample t-test

data:
    score[team == "V" & SA == "perc"]
and
    score[team == "T" & SA == "perc"]
t = 0.4413, df = 360.957,
p-value = 0.6593
alternative hypothesis:
true difference in means
is not equal to 0
95 percent confidence interval:
 -0.1136398  0.1793958
sample estimates:
mean of x mean of y
0.6722222 0.6393443
```

The following t-test is for the difference in *comprehension* scores for the visualization users (V) and the text users (T). We see strong evidence, $p < 0.0016$, that they differ.

```
Welch Two Sample t-test

data:
  score[team == "V" & SA == "comp"]
and
  score[team == "T" & SA == "comp"]
t = 3.1892, df = 245.658,
p-value = 0.001612
alternative hypothesis:
true difference in means
is not equal to 0
95 percent confidence interval:
 0.1061147 0.4488989
sample estimates:
mean of x mean of y
0.7886179 0.5111111
```

The following t-test is for the difference in *projection* scores for the visualization users (V) and the text users (T). We see some evidence, $p < 0.0886$, that they differ.

These results appear to confirm all three hypotheses, that visualization does nothing to alter a subject's perception of facts, but helps to build comprehension of facts through the clustering and display of related information. We are more reserved about the findings regarding projection, both because of the lesser statistical significance and because of the generally lower scores for all subjects in projection as compared with perception and comprehension.

```
Welch Two Sample t-test

data:
   score[team == "V" & SA == "proj"]
and
   score[team == "T" & SA == "proj"]
t = 1.7088, df = 274.508,
p-value = 0.08861
alternative hypothesis:
true difference in means
is not equal to 0
95 percent confidence interval:
 -0.02279185  0.32261401
sample estimates:
mean of x mean of y
0.6942149 0.5443038
```

7 Conclusions

We conclude from our initial analysis that we have promising evidence that visualization aids situation awareness by managers at the comprehension level. We also have some evidence that visualization aids situation awareness by managers at the projection level. The lesser performance of all subjects at the projection level may mean that this task was of greater difficulty. Certainly we saw less insight into the near term future than into the comprehension of the immediate situation across the board.

We have generated a great deal more data than has yet been analyzed, and may be able to glean better insight about the low projection scores from that data.

We learned a great many user interface improvements that could be made to both the visualization system and the text-based system that would increase their value, making future comparisons more meaningful. These improvements include better messages to clarify the system status, more intuitive responses to user expectations, and better display organization to fit what users do the most.

Having demonstrated the usefulness of visualization for situation awareness, we are excited about moving on to answer questions about different features of visualization. We are now in a position to compare two different visualization systems to learn more about the impact of different portrayals of the clustering and semantic distance information.

Acknowledgments

The authors would like to thank Dr. Queen Esther Booker and her students for help with experimental subjects and Dr. Jay F. Nunamaker and his staff for help with GroupSystems facilities.

References

[1] M. J. Adams, Y. J. Tenney, and R. W. Pew. Situation awareness and the cognitive management of complex systems. *Human Factors*, 37(1):85–104, 1995.

[2] J. Bettman and P. Kakkar. Effects of information presentation format on consumer information acquisition strategies. *Journal of Consumer Research*, 3(4):233–240, 1977.

[3] S. K. Card, J. D. MacKinlay, and B. Schneiderman. *Readings in information visualization: using vision to think*. Morgan Kaufmann, San Francisco, 1999.

[4] B. M. Collins. Data visualization — has it all been seen before? In *Animation and scientific visualization tools and applications*, chapter 1, pages 3–28. Academic Press Limited, London, 1993.

[5] C. Cool. The concept of situation in information science. In M. E. Williams, editor, *Annual Review of Information Science and Technology (ARIST)*, volume 35, chapter 1, pages 5–42. Information Today, Medford, NJ, 2001.

[6] S. Dehaene, E. Spelke, R. Stanescu, P. Pinel, and S. Tsivkin. Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science*, 284:970–974, 1999.

[7] A. R. Dennis and T. A. Carte. Using geographical information systems for decision making: Extending cognitive fit theory to map-based presentations. *Information Systems Research*, 9(2):194–203, June 1998.

[8] P. F. Drucker. *Management: Tasks, Responsibilities, and Practices*. Harper and Row, New York, 1974.

[9] M. R. Endsley. Situation awareness: The development and application of a theoretical framework. Technical report, HFES, 1993.

[10] M. R. Endsley. Situation awareness in dynamic human decision making: Theory. In *Situational Awareness in Complex Systems*, pages 27–58. Embry-Riddle Aeronautical University Press, 1994.

[11] M. R. Endsley. Measurement of situation awareness in dynamic systems. *Human Factors*, 37(1):65–84, 1995.

[12] M. R. Endsley. Towards a theory of situation awareness. *Human Factors*, 37(1):32–64, 1995.

[13] G. E. Forsythe, M. A. Malcolm, and C. B. Moler. *Computer methods for mathematical computations*. Prentice-Hall, Englewood Cliffs, N.J., 1977.

[14] D. M. Gaba, S. K. Howard, and S. D. Small. Situation awareness in anesthesiology. *Human Factors*, 37(1):20–31, 1995.

[15] S. Grotzinger and C. Witzgall. Projections onto order simplexes. *Applied Mathematics and Optimization*, 12:247–270, 1984.

[16] M. Hearst. User interfaces and visualization. In R. Baeza-Yates and B. Ribeiro-Neto, editors, *Modern Information Retrieval*, chapter 10. Addison Wesley Longman, 1999.

[17] J. Herstein. Keeping the voter's limit in mind: A cognitive processing analysis of decision making in voting. *Journal of Personality and Social Psychology*, 40:843–861, 1981.

[18] G. L. Kaempf, S. Wolf, and T. E. Miller. Decision making in the aegis combat information center. In *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting*, pages 1107–1111, Santa Monica, CA, 1993. Human Factors and Ergonomics Society.

[19] J. B. Kruskal. Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, 29(2):115–129, June 1964.

[20] K. Manktelow and J. Jones. Principles from the psychology of thinking and mental models. In *Applying cognitive psychology to user-interface design*, pages 83–117. Wiley, 1987.

[21] B. McCormick, T. DeFanti, and M. Brown. Visualization in scientific computing — a synopsis. *Computer Graphics*, 21(6):1–10, 1987.

[22] M. J. McQuaid, T.-H. Ong, H. Chen, and J. F. Nunamaker, Jr. Multidimensional scaling for group memory. *Decision Support Systems*, 27(1):163–176, November 1999.

[23] I. Nonaka and H. Takeuchi. *The Knowledge-Creating Company*. Oxford University Press, New York, 1995.

[24] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C, Second Edition*. Cambridge University Press, Cambridge, 1992.

[25] E. M. Rasmussen. Clustering algorithms. In W. B. Frakes and R. Baeza-Yates, editors, *Information Retrieval: Data Structures and Algorithms*, pages 419–442. Prentice Hall, Englewood Cliffs, NJ, 1991.

[26] E. S. Spelke. What makes humans smart? In D. Gentner and S. Goldin-Meadow, editors, *Advances in the Investigation of Language and Thought*, chapter 10, pages 277–311. MIT Press, Cambridge, MA, 2003.

[27] E. S. Spelke and S. Tsivkin. Language and number: A bilingual training study. *Cognition*, 78:45–88, 2001.

[28] T. Strzalkowski, J. Perez-Carballo, J. Karlgren, A. Hulth, P. Tapanainen, and T. Lahtinen. Natural language information retrieval: Trec-8 report. In *Proceedings of the Eighth Text REtrieval Conference (TREC 8)*, Gaithersburg, 1999. NIST.

[29] T. Strzalkowski, G. Stein, G. Wise, J. Perez-Carballo, P. Tapanainen, T. Jarvinen, A. Voutilainen, and J. Karlgren. Natural language information retrieval: Trec-7 report. In *Proceedings of the Seventh Text REtrieval Conference (TREC 7)*, Gaithersburg, 1998. NIST.

[30] G. Sundstrom. Information search and decision making: The effects of information displays. *Acta Psychologica*, 65:165–179, 1987.

[31] A. Tversky and D. Kahneman. The framing of decisions and the psychology of choice. *Science*, 211:453–458, 1981.

[32] I. Vessey. Cognitive fit: Theory-based analyses of the graphs versus tables literature. *Decision Science*, 22(1):219–241, 1991.

[33] I. Vessey and D. Galletta. Cognitive fit: An empirical study of information acquisition. *Information Systems Research*, 2(1):63–84, 1991.

[34] K. E. Weick. *Sensemaking in Organizations*. Sage Publications, Thousand Oaks, CA, 1995.

[35] E. L. Wiener. Life in the second decade of the glass cockpit. In R. S. Jensen and D. Neumeister, editors, *Proceedings of the Seventh International Symposium on Aviation Psychology*, pages 1–11, Columbus, OH, 1993. Ohio State University, Department of Aviation.

[36] O. Zamir and O. Etzioni. Web document clustering: A feasibility demonstration. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 46–54. ACM, 1998.