Gesture and Speech-Based Maps to Support Use of GIS for Crisis Management: A User Study

Sven Fuhrmann, Alan M. MacEachren, Jianwei Dou, Ke Wang, and Adrian Cox

GeoVISTA Center, Department of Geography, The Pennsylvania State University, University Park, PA 16802, USA, {fuhrmann, maceachren jxd941, kxw921, aac137}@psu.edu

Abstract

Geographic Information Systems (GIS) have the potential to provide critical support during crisis situations. Unfortunately, most GIS require expert knowledge to operate and do not support collaboration between decision makers. A human-centered systems approach was used to develop the prototype of a multimodal, dialog-assisted visual environment for geoinformation (DAVE_G) that will support collaborative crisis management by individuals without special GIS training. Since current research about multimodal map and GIS usage is at it's beginning, an initial user study with the DAVE_G prototype was conducted. The goals of the study were to investigate the basic proof of concept, collect information about user performance, gather information about overall system usability, and solicit user reactions. The initial study indicates that novice users are able to learn generic GIS interaction functions using a speech and gesture-based interface in a fairly short time and feel comfortable applying them.

Keywords

Multimodal Interaction, Collaboration, Human-Centered Systems, Crisis Management, GIS, Geovisualization

1. Introduction

Geoinformation technologies have become common tools for crisis management, e.g. the Federal Emergency Management Agency (FEMA) began using Geographic Information Systems (GIS) on a regular basis in the early 1990's when Hurricane Andrew threatened the east coast (FEMA 2003). Today, GIS are used in all five emergency management phases: planning, mitigation, preparedness, response and recovery (Johnson 2000).

Crisis management involves a collaborative effort in all phases and is often time critical, especially during emergency responses in which all responders (police, fire, public works, etc.) need to be coordinated by emergency management coordinators and decision makers. Currently, this domain faces the challenge of needing expert knowledge in order to operate the geoinformation technology, since excessive menus and command structures prevent decision makers from using them directly. Typically a GIS technician is asked by decision makers to produce a map, he/she runs to the office and returns with a paper map (Brewer in press). Another limitation of geoinformation technology is that GIS are mostly applied on single user desktop workstations that do not easily support collaborative work. The small screens and interface designed for a single seated user hinder group discussions and decisions during crises.

To address the limitations of current GIS in crisis management applications, we are developing a dialog-assisted visual environment for geoinformation (DAVE_G). We have developed the DAVE_G prototype to support a set of real-world scenarios that focus on domain-specific tasks of emergency management specialists in response situations.

2. Designing the DAVE_G prototype

We have approached the DAVE_G development from a human-centered systems perspective that focuses on users, their knowledge, and their tasks as fundamental input to the tool design process. Domain knowledge (including spoken domain-specific phrases) and tasks of emergency managers were captured using a cognitive systems engineering (CSE) approach (Rasmussen et al. 1994). Some of the knowledge elicited was represented in concept maps that detailed connections among information, technology, people, and decision-making tasks (Brewer 2002). These concept maps provided a basis for DAVE_G scenario development. The two scenarios we have used thus far to frame development are a hurricane threat in Florida and a chemical incident in the New York metro area. Besides task identification, the designed scenarios were essential for identifying generic and domain-specific GIS requests that the system must support.

Research about multimodal interfaces (Lamel et al. 1998; Rasmussen et al. 1994) emphasizes the use of natural and conversational human-computer interfaces, e.g. speech and free-hand gestures but little research has focused on multimodal GIS interaction (Schlaisich & Egenhofer 2001). Cohen and colleagues (1997) developed a pen and speech-based multimodal interface to generate interactive maps, since research showed that speech-only interfaces often provided a vague concept for spatial relations. The usability assessment of Cohen's (1997) multimodal research approach showed that participants were able to express spatial relations more efficiently than with speech-only interfaces. In a following research study Cohen, et al. (2000) indicated that users of interactive multimodal maps even solved tasks faster and had shorter error correction times than users of common GUI-based maps.

Figure 1. The DAVE_G prototype

Sharma, et al. (1999) and Kettebekov and Sharma (2001) solved a variety of speech-gesture capture problems, analyzing weather broadcasts in which individuals talk and gesture at large screen maps. Their research is being leveraged in the DAVE_G development.

3. Functionalities and architecture of DAVE_G

DAVE_G provides data querying, navigating and drawing functions (Figure 1). Two types of requests can be distinguished; one that relies on spatial references that need to be specified by gesturing (e.g. pointing and outlining), and a second that allows requests to be expressed solely by speech. The current prototype of DAVE_G features the following requests

- Data query (show/hide features, select/highlight features)
- Buffering (create/show buffers),
- Map navigation (pan left/right/up/down, center at, zoom in/out/area/full extend) and
- Drawing (points, circle, line, free hand).

For fulfilling a successful task, e.g. visualizing a theme, the user does not need to provide all required information at once. DAVE_G incrementally collects information about the intended task by supporting a human-computer-based dialog. As an example, the user might want to display population data, DAVE_G might have different kinds of population information available (e.g., census tracts, counties), thus DAVE_G would ask which of the available population data the user wants to have displayed. Cai, et al. (2003) describe the dialog approach.

The DAVE_G architecture is based upon three modules: Human Interaction Handling, Human Collaboration and Dialog Management, and Information Handling (Figure 2). The Human Interaction Handling Module has two components: 1) a human reception control that captures speech and gesture input and generates descriptions of recognized words, phrases, sentences, and gestures to be used for processing, 2) a display control that receives the processed responses of the system, e.g. GIS-based maps and textual messages and coordinates the direct feedback in response to users actions, e.g. cursor movements. Gesture is captured using active cameras that

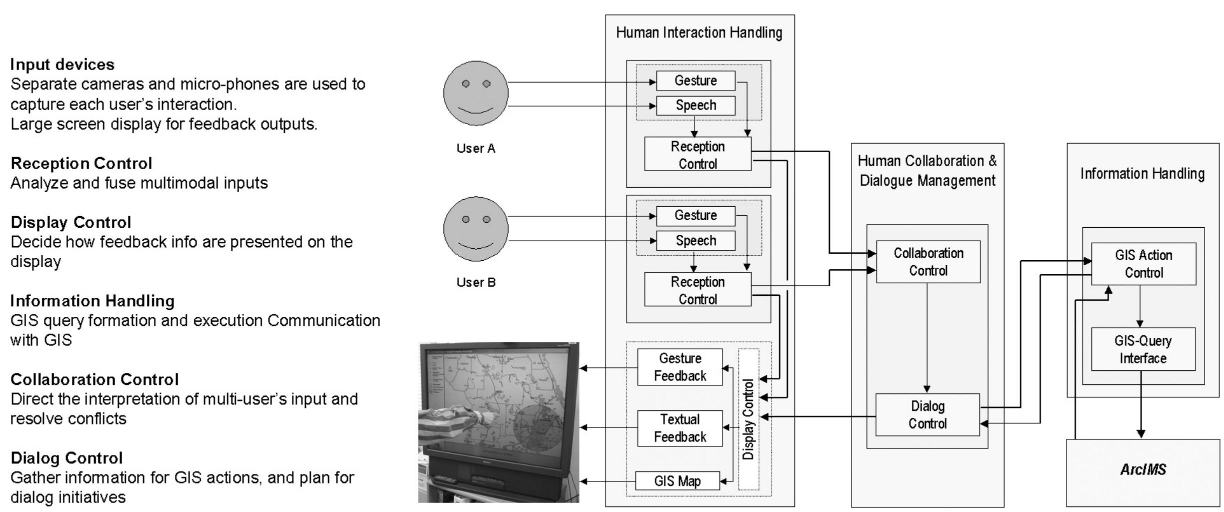

Figure 2. The architecture of DAVE_G

track the user's head and hand skin color and motion. Speech is captured by microphones and processed by speech recognition software.

The Human Collaboration and Dialog Management Module receives the gesture descriptions and verbal utterances from the reception control components. It coordinates the execution of these commands through the collaboration manager and the dialog manager. The collaboration manager is involved in the conflict management if two or more persons are interacting with DAVE_G. The dialog manager checks the received information for consistency and ambiguity and establishes a dialog with the users to resolve interpretation problems. When sufficient information is collected on a task, the information is passed to the Information Handling Module for the formation and execution of a corresponding GIS query. Information returned from the Information Handling Module can be maps and textual messages (successful query) or error messages. This information is passed to the display control.

The Information Handling Module processes the received commands by forming GIS queries and sending the query, through the GIS query interface, to ArcIMS, which is located on a remote server (ESRI 2003). The GIS action control takes care of all GIS related queries and maintains information regarding the current status of the GIS. Successful GIS queries return a map and relevant metadata.

4. Procedure

Geographic Information Systems offer not only ways to query databases or visualize data but also provide a wide range of spatial analysis methods. Since current research about multimodal GIS usage is at it's beginning, an initial user study with the DAVE_G prototype was conducted. The goals of the study were to 1) investigate the basic proof of concept, 2) collect information about user performance, 3) gather information about overall system usability, and 4) solicit user reactions.

Ten Geography graduate students, all familiar with maps and their application to both representing human and physical components of the environment, participated in the study. The study participants volunteered and did not receive any form of reimbursement. Each was familiar with standard computer interfaces. Five participants had three years of GIS experience, the others less than three years or no GIS knowledge.

None of the participants had used DAVE_G before the study. The session for each began by setting up a voice profile for speech recognition (taking an average of eight minutes). Next, participants completed a learning session (approximately 10 minutes) that introduced DAVE_G operation and functionality. The session facilitator (the first author) and the participant followed a predefined scenario in which the basic functions of the system (speech and gesture interaction) were explained. This learning phase ended when the participant clearly stated that he/she felt comfortable using the speech and gesture-based functions. This introductory session was followed by a five-minute interview focusing on the novelty of the participant's experience.

After the short interview, participants completed a guided task involving a scripted scenario in which the interaction commands (requiring speech and gesture interaction) were read by the session facilitator and repeated by the test participant. No questions were answered or additional help, e.g. on how to apply gesturing, etc. was given to the subject during this phase, since the speech and gesture coordination after the learning phase was investigated.

After successful completion of the initial tasks, the participants were given two unguided tasks. Participants had to place a fire marker at a specific highway intersection and retrieve information about a given railroad. In order to solve these tasks subjects had to load data layers, zoom, pan, and identify specific objects. The user study ended with a questionnaire and short interview.

Participant interaction with DAVE_G was video and audio taped using two digital video cameras. One camera focused on capturing the large screen interaction (system requests and responses) while the second camera recorded gestures and body movements of participants. This approach allowed us to analyze the human-computer interaction and provided valuable information for the refinement of the computational gesture model. User interaction logs were recorded in real time.

5. Results

The study produced a wide range of results, only a few of which can be presented in this paper.

The session facilitator encouraged subjects during the initial learning session to repeat and exercise interactions until participants felt comfortable using them. It took participants an average of 10 minutes to feel competent. The guided task section took an average of 2,5 minutes to complete. After the training all participants were able to finish the given two unguided tasks without help in a timely manner. Participants with GIS experience did not have any noticeable advantage in using DAVE_G compared to participants with little or no GIS experience.

In the first unguided task, all participants were able to load data sets, search a specific highway intersection and place a fire marker. The average time for accomplishing this task was 2 minutes and 20 seconds. An average of 71% of the system requests were successful, while 29% of the system requests failed and had to be repeated by the participant.

In the second unguided task participants needed to select data sets, identify a specified railroad and retrieve railroad information. It took an average of 2 minutes and 4 seconds for participants to accomplish this task. In average, 70% of the system requests were successful, while 30% of the system requests failed and had to be repeated by the participant.

Speech-based system requests were not executed by DAVE_G when they included grammar errors, were incomplete or unclearly spoken (commands did not match the voice profile). Overall, the speech recognition can be described as being sophisticated, since many participants spoke with different accents. Participants had fewer problems with gesture than speech, with an average 90% of the gestures used during the study were successful.

The qualitative analysis of the interviews indicates that in general participants felt comfortable using the system, emphasizing its quick learnability and positively describing the experienced interaction with the system. Some participants stated that the speech-based dialog allowed them to visualize data sets easily, without knowing about GIS concepts and database searches. The free hand gesturing was described as being easy to learn but some subjects worried about "loosing the cursor", i.e. loosing the hand motion tracking.

6. Discussion

This initial study indicates that novice users are able to learn generic GIS interaction functions using a speech and gesture-based interface in a fairly short time and feel comfortable applying them.

In this first study we were able to collect information about general user performance, involving individuals who were knowledgeable about maps but were not emergency management domain experts. Subsequent user studies will involve local and national emergency management personal, focusing on solving domain specific tasks.

The results led to the following redesign issues:

- Natural language/system dialog: Most users did not feel as if they were having a dialogue with the system, since the current prototype provides predefined phrases as functions. A goal here is to achieve more natural dialog by increasing the system's ability to make use of problem, domain, and recent actions to take the initiative in the dialogue, e.g. providing relevant information about available data sets and better error responses.
- Gesture recognition: Sometimes hand tracking failed and users "lost" the cursor. A new camera system is being developed to provide stable gesture tracking.
- System latency: Since the current prototype is network-based, system/network latency interfered with system performance. This issue needs to be further investigated since crisis management is time critical.

Obviously, this experiment assessed only a first fraction of possible research questions. We will continue with our human-centered system approach and conduct multiple user studies. Upcoming studies will 1) replicate this user study for desktop GIS and compare performance with standard user interfaces, 2) involve more collaborative domain-specific tasks (multi-user interaction) and 3) investigate the use of multimodal mobile technology (compare tablet-based speech-pen interaction to large screen speech-freehand gesture interaction).

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grants No. BCS-0113030, EIA-0306845; PI: Alan M. MacEachren, CoPIs: Rajeev Sharma, Guoray Cai, Michael McNeese, Sven Fuhrmann. Grateful appreciation goes to the following individuals for their assistance and support: Rajeev Sharma, Guoray Cai, Ingmar Rauschert, Levent Bolelli, and Isaac Brewer. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding agency.

References

Brewer, I. (2002) Cognitive Systems Engineering and GIScience: Lessons learned from a work domain analysis. *Proc. GIScience 2002*, 22-25.

Brewer, I. (in press) Understanding use of geospatial information and technology in crisis management. Department of Geography. Penn State, University Park.

Cai, G., Wang, H. and MacEachren, A.M. (2003) Communicating Vague Spatial Concepts in Human-GIS Interactions: A Collaborative Dialogue Approach. *Proc. COSIT 2003*, Springer, 287-300.

Cohen, P., Johnston, M., McGee, D.R., Oviatt, S., Pittman, J., Smith, I., Chen L. and Clow, J. (1997) QuickSet: Multimodal interaction for distributed applications. *Proc. Multimedia '97*, ACM Press, 31-40.

Cohen, P., McGee, D.R. and Clow, J. (2000) The efficiency of multimodal interaction for a map-based task. *Proc. of the Applied Natural Language Processing Conference*, Morgan Kaufmann, 331-338.

ESRI (2003) ArcIMS 4 architecture and functionality. ESRI Redlands.

FEMA (2003) GIS Developments at FEMA, Mapping and Analysis Center, Federal Emergency Management Agency. http://www.gismaps.fema.gov/gis02.shtm.

Johnson, R. (2000) GIS technology for disasters and emergency management. ESRI, Redlands.

Kettebekov, S. and Sharma, R. (2001) Toward natural gesture/ speech control of a large display. *Proc. Engineering for Human-Computer Interaction*, Springer, 221-234.

Lamel, L., Bennacef, S., Gauvain, J.L., Dartiguest, H. and Temem, J.N. (1998) User Evaluation of the MASK Kiosk. *Proc. ICSLP 1998*.

Rasmussen, J., Pejtersen, A.M. and Goodstein, L. P. (1994) Cognitive engineering: Concepts and applications. Wiley & Sons, New York.

Schapira, E. and Sharma, R. (2001) Experimental evaluation of vision and speech-based multimodal interfaces. *Proc. Workshop on Perceptive User Interfaces*.

Schlaisich, I. and Egenhofer, M.J. (2001) Multimodal spatial querying: What people sketch and talk about. *Proc. 1st International Conference on Universal Access in Human-Computer Interaction*.

Sharma, R., Poddar, I., Ozyildiz, E., Kettebekov, S., Kim, H. and Huang, T.S. (1999) Toward interpretation of natural speech/gesture: Spatial planning on a virtual map. *Proc. of ARL Advanced Displays Annual Symposium*, 35-39.