Multimodal Spatial Querying: What People Sketch and Talk About

Isolde Schlaisich and Max J. Egenhofer

National Center for Geographic Information Analysis
Department of Spatial Information Science and Engineering
University of Maine, Orono, ME 04469-5711, USA
{isolde,max}@spatial.maine.edu

People want to use geographic information systems (GISs) while spending little time for learning how to use the system. This trend increases the need for easy-to-learn and intuitive user interfaces. In order to make GISs easier to use for untrained users, we investigate an interaction method that mimics the natural communication between people and allows users to formulate queries. In human-human interaction people often communicate about space by talking and simultaneously drawing a freehand sketch; therefore, this combination of modalities is attractive for describing spatial scenes when querying a GIS. To assess the usefulness of this approach, we have conducted a human-subject experiment in which subjects were asked to draw spatial queries and talk at the same time. We analyzed what modality subjects preferred when they had the choice of sketching and talking. We found that subjects gave complementary information through the two modalities, indicating that sketching cannot replace the verbal descriptions, and vice versa. We also found redundant information in both modalities, which provides a solid foundation for integrating elements from the two modalities.

1. INTRODUCTION

Future geographic information systems (GISs) must be easier to use for everyday users who do not have extensive training in the use of computer systems. GISs are widely used to query and analyze spatial data in such applications as car navigation, tourist information, and real estate management. While the community of interested GIS users is growing rapidly, the current systems are often too difficult to use or require too long a time to learn them. State-of-the-art GISs offer a WIMP (windows, icons, menus, pointers) interface where users often struggle to find their way through the available menus or type the commands they wish to execute. People are, however, more familiar with other approaches to communicate about space in everyday life, such as speech, drawings, gestures, and writing.

We are investigating a *Sketch-and-Talk* user interface, which combines the input of spoken language and drawn sketches to build an effective query method. This kind of interface is expected to lead to a broader acceptance of GIS and suit the demand for a wide-ranging access to computer systems (Stephanidis 2001). For this goal we conducted an experiment to investigate how people communicate about space, simulating the interaction with a *Sketch-and-Talk* GIS. We recorded the responses of twelve subjects while they were asked to respond to four spatial queries. The analysis of the drawings and the corresponding text showed that all subjects used both interaction modalities. The spatial objects in the queries are at times stated in only one modality and at times in both, generating complementary and redundant information. More than half of the spatial objects were mentioned in both modalities, giving further evidence that one modality alone would be insufficient.

The remainder of the paper is structured as follows: Section 2 briefly discusses interaction methods with GISs. Section 3 describes the experiment to determine what people sketch and to what they refer with voice annotations. Section 4 analyzes the experiment, followed by our conclusions in Section 5.

2. INTERACTION WITH GEOGRAPHIC INFORMATION SYSTEMS

GISs have a large audience of users due to the attractive use of spatial access and spatial analysis in a variety of application areas (Longley *et al*. 1999). Since many potential GIS users often lack specific training in the use of GISs, the need for more intuitive user interfaces arises. One viable approach to the design of GISs and their user interfaces is the consideration of people's common-sense knowledge of the geographic world (Egenhofer and Mark 1995).

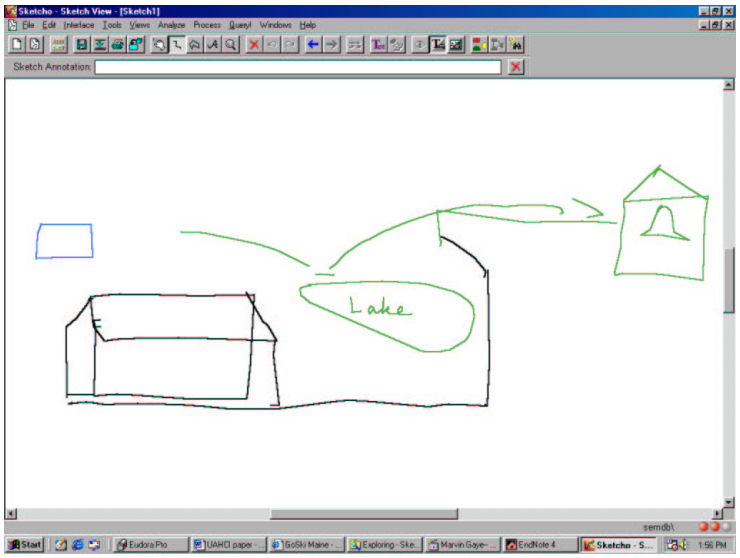
New interaction methods that deviate from the traditional ones are asked for in order to make GISs more usable (Egenhofer and Kuhn 1999; Bass 2001). These interaction methods include talking, drawing, and gesturing. Sketching as a spatial query language offers a fairly natural way of expressing spatial information (Egenhofer 1996a), as was demonstrated with a recent implementation of Spatial-Query-by-Sketch (Blaser and Egenhofer 2000). It enables the users to draw a spatial configuration of the scene they are looking for and search for the most similar settings in a spatial database. Sketching alone, however, is limiting and a more efficient interaction could be achieved by offering multiple modalities (Oviatt and Cohen 2000).

In human-human interaction people typically use spoken language or written comments to augment sketches (Egenhofer 1996b). The two interaction modalities are especially useful for GISs, because some types of spatial information are better expressed by drawing a sketch, while others are better conveyed with words. Particularly spatial relations and metric information are drawn faster and with less ambiguity than if they are given verbally. For example, it is more precise to describe the spatial configuration of a house close to the northwest shore of a lake by drawing a sketch than by explaining the same details with words. On the other hand, attributes of an entity, such as the address of a building, are better stated in spoken or written language. Sketching and talking are most powerful in a spatial query language if they are used together. For instance, a person who describes the way from her workplace to her house by drawing a sketch that shows her house, the streets surrounding it, and crucial landmarks, could also simultaneously use spoken language to add information about the name of the streets. If such an intuitive interaction style were available with a GIS, it would enable more casual users to access and analyze more spatial information, because this combination mimics the way people often communicate about space. Better use of spatial information will enable people to make better-informed decisions in their everyday lives.

3. EXPERIMENT

To explore how people use sketches and verbal annotation, we simulated a *Sketch-and-Talk* environment and asked subjects to state various queries by drawing on a touch-sensitive screen and talking at the same time. The sketch was entered through the Spatial-Query-by-Sketch interface (Blaser and Egenhofer 2000), while the spoken words were recorded with a tape recorder.

Six male and six female subjects in the age range of 20-29 years participated in the experiment. The instruction for the first question (Query 0) was to interact with an information system by sketching the way from the subject's house to the post office. Subjects were not told that they could use voice input as well. The reason for the first query was to allow subjects to familiarize themselves with the sketch system. After that, they were told to imagine that the system could process spoken language as well. Their task was now to formulate another three queries by sketching and talking. The particular instructions were: "Communicate with the system to retrieve the place where you grew up and went to school" (Query 1), "Look for your dream vacation resort" (Query 2), and "Look for a campus that has a setting that is similar to the University of Maine" (Query 3). Figure 1 shows an example of one subject's sketch, describing her way from home to school and the accompanying text.



"I lived with my grandparents on a farm. It took me approximately 10 to 15 minutes to school, because we lived on a hill. I lived off of a major route and then I would have to change over to another route. I would have to go around a lake, which would be here, and go up another road, ascend another hill, and there was my school. It was situated so that, if you looked out across the road from the farm, you could actually overlook the lake and the school."

Figure 1. Sketch of subject 3, describing her way to school and the corresponding text.

4. ANALYSIS

For the analysis of the experiment, the spoken text was typed into a text file and cleaned up from noise. To compare the content of the sketch with the text, we counted in each sketch the number of recognizable objects drawn and in each text the words as well as the number of objects (i.e., nouns) used in the text. Examples of spatial objects that can be found in Figure 1 are buildings, streets, and a lake. The twelve subjects sketched 428 objects in queries 1-3, while they referred to 314 objects in the spoken parts of queries 1-3.

	Query 0	Query1	Query 2	Query 3	Average
Number of objects in sketch	6	10	9	17	11
Number of words in text	–	156	120	188	155
Number of objects in text	–	10	7	11	9

Table 1. Average number of objects per sketch, and words and objects per annotation.

Table 1 shows the average number (arithmetic mean) of the objects that were found in the four sketches and the average number of objects in all sketches of the experiment in the second row. In each of the 48 sketches, subjects drew on average eleven objects. This number is relatively small, confirming Blaser's (1997) finding that a person draws a sketch of average complexity with 5 to 25 objects.

The number of words in the text is approximately 16 times higher than the number of objects. The comparison of drawn objects versus verbally mentioned objects shows that on average up to 50% more objects were drawn than talked about.

4.1 Choice of Modality

The experiment showed that people embrace the offer to sketch and talk as a method of interaction with a GIS. Some subjects were talking more, while others were sketching more, but none did completely ignore one of the modalities. All subjects used both modalities in at least one of their queries. Ten of the twelve

subjects used both modalities for each query. One subject only talked to formulate query 2, while another subject used only sketching in queries 2 and 3.

Five out of the twelve subjects (42%) already talked during the first query, although they had no information about the possibility of voice input. They did not expect the system to process the words. This voluntary use of sketching and verbal annotations is another indicator that the simultaneous use of sketching and talking is a natural interaction style for spatial querying.

4.2 Complementarity

Users have a choice to formulate different parts of the query with the modality that best fits their mental model. This type of interaction may lead to simple requests, with fewer mistakes and without complicated formulations. At the same time, some parts of a query may be found only in the sketch, while others may be expressed only verbally; therefore, the two modalities promote a complementary query formulation. Through a single interaction modality, this information would be either lost or users would have to reformulate the query using the only available modality.

Table 2 shows that 46% of the 428 objects that appeared in the sketches only exist in the sketched part of the queries. On the other hand, the 83 objects that are only mentioned while talking comprise only 26% of the 314 objects in the spoken part. These results indicate that objects mentioned in the verbal description are more likely to have a graphical counterpart, while almost half of the drawn objects lack a verbal annotation.

Objects in sketch only	Objects in sketch and talk	Objects in talk only
197	231	83

Table 2. Number of objects that appear only in a sketch, only in the verbal part, and in both.

4.3 Redundancy

In the same way the two modalities enable complementarity, they offer users to describe the same content redundantly through either sketching or talking. Redundancy may be an indication for the importance of a piece of a query. Redundancy also provides an important glue when integrating the sketch and the spoken request into a single query.

Subjects often referred redundantly to an object in the sketch and in the verbal annotation. For example, in Figure 1 the objects farm, lake, school, and routes where drawn in the sketch as well as stated in the text. Table 2 shows that 231 of the objects are referred to in both modalities, which is 54% of all sketched objects and 74% of all objects in the verbal part. This indicates that the subjects took advantage of the possibility of simultaneous input of sketches and voice, and produced a remarkable quantity of redundant information.

5. CONCLUSIONS

Simpler and more intuitive user interfaces are needed to give inexperienced users access to GISs. The combined use of sketching and talking mimics the way people typically communicate about space, therefore, these two modalities is a natural way to formulate questions about space. We expected that inexperienced GIS users are more at ease to combine speech and sketches to convey their knowledge of space than to use a single mode. Also the use of the two modalities—speech and drawing—is especially suitable for communication about spatial topics (Egenhofer 1996b; Oviatt 1996). Although the emphasis of this paper is on providing easy-to-use interfaces for inexperienced users, the approach of *Sketch-and-Talk* can also facilitate the routine work of experienced GIS users. The natural way of interaction will make them at ease with the use of the system and might improve their performance as well.

We conducted an experiment to demonstrate that people make use of both modalities when formulating spatial queries. We found significant complementary and redundant information in the spoken and sketched parts of the queries. From a user-interface point of view, these results demonstrate that both modalities are needed for a natural formulation of spatial queries. From a query processing point of view the information about redundancy forms a solid basis for query processing, as it enables to integrate the sketched and the spoken query parts.

ACKNOWLEDGMENTS

This work was partially supported by a grant from the National Science Foundation under grant number IRI-9613646. Max Egenhofer's research is further supported by NSF grants IRI-9613646, IIS-9970123, and EPS-9983432; the National Institute of Environmental Health Sciences, NIH, under grant number 1 R 01 ES09816-01; the National Imagery and Mapping Agency under grant number NMA202-97-1-1023 and NMA201-00-1-2009, and by Lockheed Martin M&DS.

REFERENCES

L. Bass (2001) Interaction Technologies: Beyond the Desktop. In: C. Stephanidis (Ed.), *User Interfaces for All: Concepts, Methods, and Tools*, pp. 81-95. Mahwah, NJ, Lawrence Erlbaum Associates.

A. Blaser (1997) *User Interaction in a Sketch-Based GIS User Interface*. Technical Report, National Center for Geographic Information and Analysis, University of Maine, Orono, ME.

A. Blaser and M. Egenhofer (2000) A Visual Tool for Querying Geographic Databases. In: V. Di Gesù, S. Levialdi, and L. Tarantini (Eds), *AVI2000—Advanced Visual Databases*, Salerno, Italy, pp. 211-216.

M. Egenhofer (1996a) Spatial-Query-by-Sketch. In: M. Burnett and W. Citrin (Eds), *VL '96: IEEE Symposium on Visual Languages*, Boulder, CO, pp. 60-67, IEEE Press.

M. Egenhofer (1996b) Multi-Modal Spatial Querying. In: M.-J. Kraak and M. Molenaar (Eds), *Advances in GIS Research II, Seventh International Symposium on Spatial Data Handling*, pp. 12B.1-15. Delft, The Netherlands.

M. Egenhofer and W. Kuhn (1999) Interacting with GIS. In: P. Longley, M. Goodchild, D. Maguire, and D. Rhind (Eds), *Geographical Information Systems: Principles, Techniques, Applications, and Management*, Vol. 1, pp. 401-412. New York, John Wiley & Sons.

M. Egenhofer and D. Mark (1995) Naive Geography. In: A. Frank and W. Kuhn (Eds). *Spatial Information Theory*. Lecture Notes in Computer Science 988, pp. 1-15. Berlin, Springer-Verlag.

P. Longley, M. Goodchild, D. Maguire, and D. Rhind, Eds. (1999) *Geographical Information Systems: Principles, Techniques, Applications, and Management*. New York, John Wiley & Sons.

S. Oviatt (1996) Multimodal Interfaces for Dynamic Interactive Maps, *Conference on Human Factors in Computing Systems - CHI'96*, pp. 95-102, ACM Press, New York.

S. Oviatt and P. Cohen (2000) Multimodal Interfaces That Process What Comes Naturally. *Communications of the ACM* 43(3): 45-53.

C. Stephanidis (2001) User Interfaces for All: New Perspectives into Human-Computer Interaction. In: C. Stephanidis (Ed.), *User Interfaces for All: Concepts, Methods, and Tools*, pp. 3-17. Mahwah, NJ, Lawrence Erlbaum Associates.