Initial Results from Speech and Sketching User Study

Aaron Adler CADLERUN@CSAIL.MIT.EDU
Randall Davis DAVIS@CSAIL.MIT.EDU
MIT Computer Science and Artificial Intelligence Laboratory, 32 Vassar Street, Cambridge MA, 02139 USA

1. Introduction

Sketching is commonly used in the early stages of design, however, some information is difficult to express using sketching alone. When designers talk while sketching, a considerable amount of information is also conveyed using speech. Consider for example, the robot and its sketch in Figure 1. It's impossible to make any sense of the sketch without the speech that went along it.

Our group has developed several sketching systems in a variety of domains (Alvarado & Davis, 2001; Hammond & Davis, 2002; She, 2006). For example, ASSIST lets the user sketch in a natural fashion and recognizes mechanical systems. It then interfaces with a simulation tool to allow users to view their sketch in action.

We aim to create a more natural user interface by adding speech recognition to the sketching system. There are existing systems that allow the user to make simple spoken commands to the system (Cohen et al., 1997; Demirdjian et al., 2005; Kaiser, 2005). Our goal is to move beyond simple commands to create a multimodal system where the user can have a natural conversation with the computer, similar to one they would have with another person. In order to better understand how such conversations happen between two people, we conducted a user study in which two people conversed about several circuit designs. This paper describes the study and some initial results.

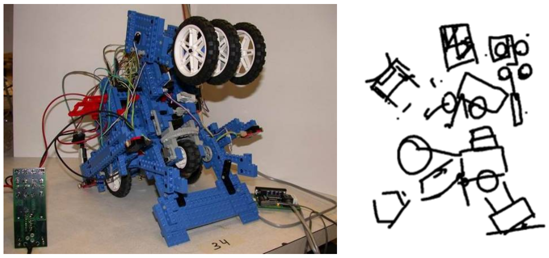


Figure 1. A robot and a sketch of it.

2. User Study

2.1 Study Motivation

Although there have been other systems that let users sketch and speak, they are limited in one or more of the following dimensions:

- Command-based speech – The user talks to the system using one or two words, not natural speech.
- Unidirectional communication – The system can't ask questions or add things to the sketch.
- Annotation instead of drawing – The user can only annotate an existing representation but not use free form drawing.
- Fixed set of symbols – The user has to know the fixed symbol vocabulary.

Ideally, we would have conducted a Wizard-of-Oz study in which responses to the participant would appear to be coming from a computer. However, we determined that this was too difficult given the open-ended nature of the speech and sketching in the study.

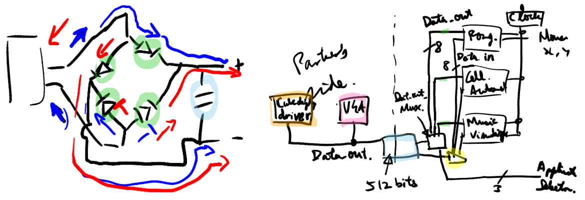
2.2 Study Setup

There were 21 participants in the study; all of them were students in the digital circuit design class at MIT. In the study, the experimenter and participant sat across a table from each other. Each person had a Tablet PC with software we designed that replicates on each tablet in real time whatever is drawn on the other tablet, in effect producing a single drawing surface usable by two people at once. The sketching software allowed the users to sketch using various colors of ink with a pen or a highlighter. The user could also select an eraser or start a new blank page. The software recorded the x and y positions, time, and pressure data for each point the user drew.

Two video cameras and headset microphones were used to record the study. The participants sketched and talked about several different items: a floor plan, the design for an AC/DC transformer, the design for a full adder, and the final project they built for their digital circuit design class. The experimenter added to the sketch and asked the participant questions about the different components of the sketch at various points during the study. The study software recorded the sketch, audio, and video data streams so that they were all synchronized and could be replayed later.

3. Initial Results

Speech from both participants was transcribed and time stamped using the Sphinx speech recognizer forced-alignment function. To date detailed transcripts have been produced for six subjects. Figures 2(a) and 2(b) show sketches from the study. Figure 3 shows several parts of the conversation about Figure 2(b).



(a) A transformer. (b) A participant's project.

Figure 2. Two sketches from the user study.

Experimenter: so then what's what's um this piece what's that

Participant: that would be the mux for the data input actually

---

Participant: that was a uh uh yeah a memory bank with five hundred and twelve um yep five hundred and twelve bits this ah I could that I had read and write access to

Figure 3. Two fragments of the conversation about Figure 2(b). Notice the disfluencies and repeated words.

Our analysis of the study has focused on how speech and sketching work together when people are interacting with each other. We want to use this as a guideline for the development of a digital whiteboard that could understand the same sorts of speech and sketching, ultimately entering into the same sort of dialog with the user.

A qualitative analysis of the recorded and transcribed data has led to a series of initial observations that we are studying further:

- Sketching
  - Identifying regions – Color was used to refer to existing parts of sketches and establish correspondences between different objects in that sketch.
  - Differentiating objects – Participants switched colors to indicate new objects.
  - Being artistic – Colors were chosen to reflect the real-world color of objects.
- Language
  - Disfluent, repetitious speech – This type of speech appears to occur frequently when designers are thinking about what to say.
  - Question responses – Responses tend to reuse vocabulary from the question.
  - Speech relates to current sketching – Speech is related to what is currently being sketched.
- Multimodal
  - Referencing lists of items – Lists of items are spoken and sketched in the same order.
  - Referencing written words – Written words or their abbreviations tend to occur concurrently with their spoken utterances.
  - Coordination between input modalities – If a designer's speech gets too far ahead of their sketching, they slow down or pause their speech.
- Questions
  - Revision – Questions can cause revisions to the sketch to make it more accurate.
  - Broader explanation – Questions about one part of the sketch can spur explanations about other, unrelated parts of the sketch.
- Comments
  - Uncertainty – Uncertainty is indicated by using phrases like "I believe."
  - High-level comments – For example, comments about switching ink color.
  - Forgotten vocabulary – Both people are expected to be able to fill in words that their partner forgot.

4. Analysis to be done

The above initial qualitative results are intriguing, but additional data analysis work remains. Some of the areas we are looking at include: how sketched objects relate to the words and phrases that reference them in the speech, and how pauses between words in the speech relate to the concurrent sketching.

5. Conclusion

We conducted a user study to gather data about natural conversations about designs. There are many interesting initial observations that we are investigating in more depth. These results will help us build a system that can interact and converse more naturally with its users.

6. Acknowledgements

This work is funded by Microsoft iCampus, Project Oxygen, and Intel Corporation.

References

Alvarado, C., & Davis, R. (2001). Resolving ambiguities to create a natural sketch based interface. *Proceedings. of IJCAI-2001*.

Cohen, P. R., Johnston, M., McGee, D. R., Oviatt, S. L., Pittman, J., Smith, I., Chen, L., & Clowi, J. (1997). Quickset: Multimodal interaction for distributed applications. *Proceedings of Mutimedia '97* (pp. 31–40). ACM Press.

Demirdjian, D., Ko, T., & Darrell, T. (2005). Untethered gesture acquisition and recognition for virtual world manipulation. *Virtual Reality*, *8*, 222–230.

Hammond, T., & Davis, R. (2002). Tahuti: A geometrical sketch recognition system for UML class diagrams. *AAAI Spring Symposium on Sketch Understanding*, 59–68.

Kaiser, E. C. (2005). Multimodal new vocabulary recognition through speech and handwriting in a whiteboard scheduling application. *IUI '05: Proceedings of the 10th international conference on intelligent user interfaces* (pp. 51–58). New York, NY, USA: ACM Press.

She, C. (2006). A natural interaction reasoning system for electronic circuit analysis in an educational setting. Master's thesis, MIT.