Building the Design Studio of the Future

Aaron Adler and **Jacob Eisenstein** and **Michael Oltmans** and **Lisa Guttentag** and **Randall Davis**
MIT Computer Science and Artificial Intelligence Laboratory
32 Vassar Street
Cambridge, MA 02139
{cadlerun, jacobe, moltmans, guttentag, davis}@csail.mit.edu

Abstract

The absence of software support for early-stage design suggests that conventional graphical user interfaces are inadequate for this task. Perceptual interaction – in particular, pen-based interaction – may afford more natural and intuitive user interfaces for design support. The development of a multimodal intelligent design studio poses several important research challenges. Natural multimodal interaction demands a better understanding of how designers use sketching in combination with other modalities to communicate ideas about design. Intelligent design feedback requires research on how designers want to interact with their tools, and how such interaction can be managed. This paper presents a research plan for developing an intelligent multimodal design studio. We will describe recently completed, ongoing, and planned empirical studies, as well as system components that we have built and plan to build.

Introduction

Much of the recent interest in sketching stems from its prevalence in early stage design (Davis 2002; Ullman, Wood, & Craig 1990). Sketch-based user interfaces could provide automated support for early stage design, bridging the divide between the "back of the envelope" and CAD software.

Intrigued by this possibility, our research group has worked on developing an intelligent multimodal design studio, based on natural human-human interaction. This paper presents a long-term research agenda for realizing this vision. We describe our ongoing work, including both the empirical studies that will help us define the proper specification for the studio, and the system components that will implement these requirements.

While most of our system development work has focused on sketch recognition (Alvarado, Oltmans, & Davis 2002; Hammond & Davis 2004; Sezgin, Stahovich, & Davis 2001), we have also conducted exploratory studies aimed at finding out more about how sketching is used in design situations. The conclusions were perhaps surprising: while sketching is indeed an ideal modality for describing imprecise spatial and topological features, sketching alone is almost never sufficient for accurately and fully describing a design. Speech and hand gestures are critical for a number of purposes:

- Expressing more precise spatial and structural relationships: "there are four identical, equally spaced pendulums."
- Relating sketches that are drawn from different perspectives: "[deictic gesture] this is the same thing from the top."
- Describing temporal events or paths of motion: "the wheel turns like this [iconic gesture]."

These are just a few of the ways in which speech and gesture communicate crucial information about a sketch. Unless these additional modalities are supported, this information will have to be communicated using either an artificial language of sketch-based "commands," or a traditional GUI. Neither method satisfies the original goal of uninhibited natural interaction.

The space of early-stage design toolkits is relatively unexplored; an exception is (Landay & Myers 1995). Consequently, we feel that human-human interaction is the best starting point for determining the desirable usability characteristics of such a system. Our efforts are aimed at building a dialogue-based design studio, simulating interactive design with a human partner as closely as possible. Once we have a working prototype, additional testing will help us locate places where divergence from the human-human interaction paradigm improves usability.

Building the Multimodal Intelligent Design Studio

Figure 1 illustrates our long-term vision for how the **Multimodal Intelligent Design Studio** should be designed, implemented, and evaluated. It includes empirical studies, represented as ovals, and system components, represented as rectangles. Arrows indicate dependency relationships showing, for example, that empirical studies inform the design of system components.

This section briefly describes the various components of Figure 1 and includes short scenarios illustrating the potential benefits of each component. This project is a work in progress; these scenarios illustrate our vision for the design studio, not current capabilities.

Exploratory Studies are initial, empirical research on early stage design; we describe the results of some completed exploratory studies in this paper.

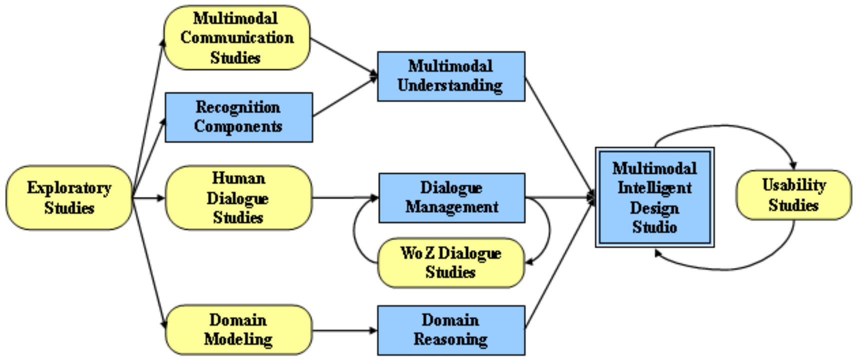

Figure 1: Building the Multimodal Intelligent Design Studio

Multimodal communication studies investigate how users combine sketch, speech, and gesture to express ideas about design. These studies, combined with robust sketch (Alvarado, Oltmans, & Davis 2002), speech (Seneff *et al.* 1998), and gesture **recognition components**, will help us ultimately build systems that perform **multimodal understanding** of natural communication.

Scenario 1: A designer sketches a circuit diagram, while verbally describing the function of the part that is being drawn. Later, the designer provides additional information by speaking and gesturing at that part in the drawing. In both cases, the multimodal understanding component resolves verbal references to referents in the non-verbal modality: either sketch or gesture. Non-verbal modalities can also provide other cues about the structure of the speech, the occurrence of speech repair events, etc. These cues can facilitate speech recognition through multimodal disambiguation.

Dialogue Studies are planned to determine how designers collaborate and interact. These studies will inform a **dialogue management** component, which is tasked with managing the interaction with the user. An iterative design cycle using **Wizard of Oz studies** is planned to refine the usability characteristics of this component.

Scenario 2: The designer sketches or says something that the system is unable to interpret; it is able to reason about how and whether to interrupt for clarification. Through this facility, the system is able to add new vocabulary terms and sketched symbols naturally, rather than through additional programming.

Domain Modeling is needed to learn the vocabulary, structure, and interaction properties of a design domain, such as mechanical design. This is necessary to build **domain reasoning** components that can provide intelligent feedback about design decisions.

Scenario 3: After having completed an initial conceptual design, the designer can ask the computer whether any important design guidelines have been violated, and then work interactively to correct them. If designers trust that essential guidelines and constraints will be managed by the computer, they will be more free to brainstorm creatively about design possibilities. Moreover, when the software explicitly manages the correction of design flaws, the underlying rationale will be documented transparently, allowing other designers and customers to understand the decisions that shaped the final design.

The remainder of this paper describes our ongoing and planned work on the components of Figure 1. While we address all parts of the diagram, we highlight the following areas of ongoing and recently completed research: exploratory studies of design scenarios; focused studies of multimodal communication, leading to a new corpus of sketch, speech, and gesture data; and the initial development of multimodal understanding components.

Exploratory Studies

Our exploratory studies investigating how sketching is used by designers have helped us identify the problems we are likely to face in supporting sketching in realistic design situations. These studies led us to pursue a more holistic, multimodal approach to design support.

In one such study, undergraduate students who participated in a robot design contest were asked to explain their designs. The sketches were recorded on paper with a special pen and clipboard (the Crosspad) that captured the stroke data digitally. Most of the sketches are very difficult to understand on their own, as Figure 2 demonstrates. Speech and gesture provide critical details and context; for example, a speaker might draw only a rectangle but say, "This gear train here…" Gesture was used to show behavior – for example, free-hand gestures were often used to simulate one of the moving parts, such as the robot's arm. Another common use of speech was to annotate the 3D relationships of parts when drawn in a 2D projection. For example, one speaker drew directly on top of an existing sketch, while saying, "this arm attaches on top like this," clarifying the relationship between the parts. Without this context, the bird's eye view gives no clues about the relative positions of the robots parts.

Figure 3 illustrates a similar situation in another study. The speaker has created two sketches side by side. They

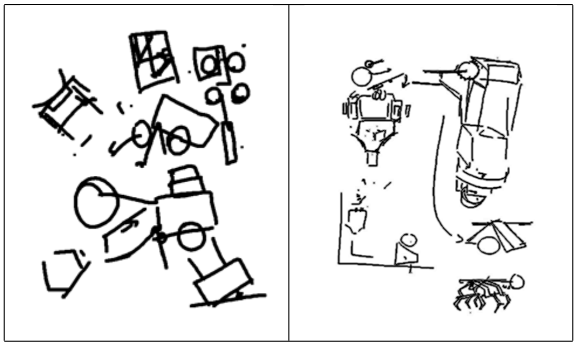

Figure 2: Without the context provided by speech and gesture, these sketches are hard to understand.

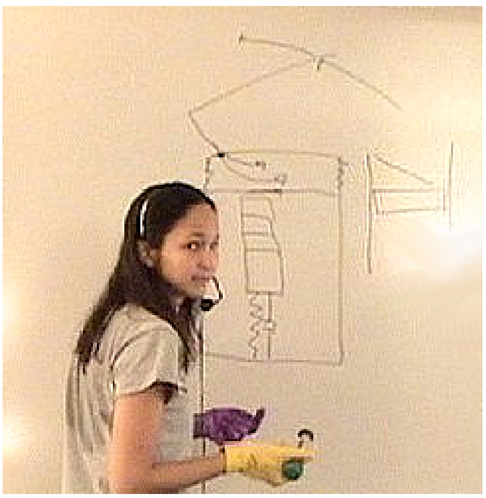

Figure 3: The sketch to the right is an aborted attempt at redrawing the main sketch to the left.

might be subcomponents of a single mechanical device, but in fact, the figure to the right is an aborted attempt to redraw the original figure, on the left. Without the surrounding conversational context, it is hard to see how one might determine this relationship.

In a third study, participants were asked to carry out four tasks:

1. sketch their current apartments,
2. brainstorm several apartments they would like to live in,
3. draw a cleaner version of their favorite sketch, and
4. redraw their favorite apartment while verbally describing the design to the moderator.

The purpose was to see whether sketching styles and neatness would differ in brainstorming versus tasks in which a more precise sketch was required.

The sketches were collected on a Tablet PC running Windows XP Tablet PC Edition. We collected a total of 127 sketches from 23 subjects for a combined total of over 11,000 strokes. Thus far, our data has not shown any significant difference in drawing style between the different sketching tasks. This is counter to our intuitions, but there are a few possible explanations. Since there were no design constraints to navigate (e.g., every bedroom must have a door to the living room), the brainstorming task was not particularly difficult. This allowed participants to focus on sketching rather than thinking about the design. Also, participants were not experts in architecture or interior design, so their "neat" sketches were still perhaps not as precise as those of a professional. We are planning a follow-up study in which architecture students are asked to design floorplans with respect to specific design constraints. This will help us understand whether brainstorming poses unique problems for sketch recognition.

Multimodal Communication Studies

Speech, sketching, and gesture are almost always most informative when used in combination. Building software that can process multiple modalities of communication requires a better understanding of how these modalities interact. We are conducting a series of studies to explore the interactions between sketch, speech, and gesture, and how they combine to express features of the design.

Speech and Sketching

We conducted an empirical investigation to collect informal and natural speech from users as they were sketching and verbally describing mechanical systems. The purpose of this study was to identify the vocabulary used to describe mechanical systems and find out which features of the system were described verbally and which were drawn. In addition, we wanted to identify relationships between the speech and the sketching inputs that would enable us to exploit the data and create a system that responds to the user's utterances.

Six users, drawn from the MIT community, were asked to draw and explain six mechanical devices (Adler & Davis 2004). The speakers reproduced small versions of the devices at a whiteboard and were videotaped while making their explanations. Their speech and sketching was manually transcribed with time stamp information. Topic shifts were also manually annotated.

Using these annotations and pen strokes, we tried to determine what sort of vocabulary was natural when describing the mechanical systems. Several general patterns emerged from the data. For example, disfluencies, such as "ahh" and "umm" were good indicators that the user was still talking about the same topic. Phrases such as "there are" and "and" were indicators that the user was starting a new topic. Ultimately, awareness of these patterns allowed us to build a multimodal topic segmentation system using sketch and speech data. This system enabled the integration of speech into the sketching framework, allowing the user's utterances to affect the sketch. The resulting system is described below in the section on Multimodal Understanding and Integration.

Speech and Gesture

Our initial exploratory studies also revealed that gesture played an important role in the explanation of sketched designs. Gesturing behavior can be extremely varied, thus pos-

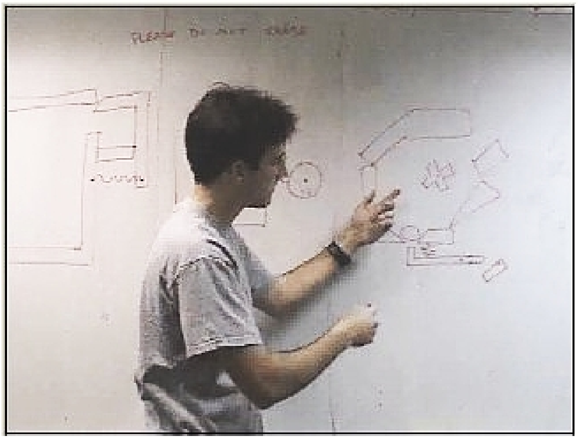

Figure 4: A speaker describes the behavior of a mechanical device.

Figure 5: The 3-D mechanical toy used in the study.

ing difficulties for machine recognition. Our goal was to determine whether there were common patterns to speech and gesture behavior, raising the possibility that these patterns could be exploited to enable the development of gesture understanding components.

In an initial study, nine speakers drawn from the MIT community were shown animated simulations of three simple mechanical devices (Eisenstein & Davis 2003). The speakers were then asked to describe the behavior of the device, with the aid of a pre-drawn diagram (Figure 4). Speech and gesture were both manually transcribed, using the notation described in (McNeill 1992).

Our experiments reveal that the presence of a sketched diagram actually simplifies speakers' gestures, significantly lowering the ratio of iconic gestures to deictic gestures. This is contrary to the existing psychology literature, which indicates that iconic gestures – in which the hand imitates an object or event – usually outnumber simple pointing (deictic) gestures by a rate of approximately ten to one. Iconic gestures are difficult to interpret because they are highly idiosyncratic; there are, after all, numerous ways in which a hand gesture could imitate some feature of an object or device. In contrast, deictic gestures are relatively easy to interpret; only the object of the deictic reference must be disambiguated. When presented with a sketched diagram, the ratio of iconics to deictics was only roughly three to two. Moreover, 80% of iconics simply described paths of motion, rather than more abstract features of the object.

The corpus generated by this study has also been used to investigate how gestures can help resolve verbal references (Eisenstein & Christoudias 2004) and is the basis for an ongoing study of sentence segmentation. We are currently annotating the data from a second, more rigorous iteration of this study, in which we compare the gestures that are observed across the following conditions: the speaker is given a pre-printed diagram; the speaker is not given a diagram; and a trimodal condition in which the speaker is given a tracked pen with which to sketch a diagram. We turn next to a discussion of some preliminary results from the trimodal condition of this study.

Trimodal Interaction

While the pairwise interactions between sketch, speech, and gesture make a good starting point for empirical research of multimodal interaction, a trimodal user interface ultimately requires studies of the interaction between all three modalities. For example, it seems likely that the presence of a pen alters the type of gestures and their relationship with the speech. To address this issues, we have recently conducted a study of the interactions between speech, gesture, and sketching – the first study of such trimodal interaction that we are aware of. In this study, speakers described mechanical devices to a friend or acquaintance. The use of acquaintances as dialogue partners is preferred because it is thought to reduce inhibition effects (McNeill 1992). Participants were given two minutes to view a simulation of a mechanical device, or to interact with the device itself. The simulated devices included a piston, a Pez dispenser, and a latchbox. The actual device was a mechanical toy, shown in Figure 5.

While we have not yet completed the annotations from which to collect quantitative data, some potentially interesting phenomena were observed. In most cases, the speaker began with an uninterrupted monologue that lasted between thirty seconds and two minutes, followed by a question-answer period. Participants were not instructed to structure their interactions in this fashion, but as many of the participants were college students, this may mirror the typical mode of interaction with a teaching assistant.

The monologue phase can be described as having two distinct modes of explanation, corresponding roughly to the structure and function of the device. In the *structure* mode, participants explained the physical configuration of the device's parts, without describing the device behavior. In this mode, the sketching appeared to drive the monologue. Most of the speech involved naming the parts that were being drawn. A few speakers sketched silently and then named the parts, but most named the parts while drawing them. The act of drawing was used to ground anaphoric references, so there was very little gesture during this phase.

During the *function* phase, participants explained the purpose of the diagram, and how the parts moved and interacted to realize that purpose. Speech dominated in this phase; gesture and sketching were used to resolve references and indicate paths of motion. As noted above, in the Speech and Gesture section, complex iconic gestures – in which the hand represents a physical object through its shape or path of motion – were relatively infrequent, since the diagram could be used to establish all references. Some participants used sketching for deixis, by over-tracing the object of the reference. Sketching was also used in place of iconic gesture, e.g., by indicating a path of motion with an arrow. The choice of whether to use a hand gesture or a sketch appears to be idiosyncratic. Some participants never used sketching in place of gestures, while others did so frequently. However, in no case did sketching completely obviate the need for hand gestures.

The structure and function modes were combined in different ways, varying mainly in relation to the device that was being explained. When participants were asked to explain devices that they viewed via simulation, they almost always combined the modes sequentially, explaining the structure of the device in its entirety, and only then explaining the function. When explaining the mechanical toy, participants were more likely to interleave the structure and function modes. In such cases, participants would describe each subcomponent of the device – e.g., the sliders that keep score – starting with the structure and then describing the function.

There are a few possible explanations for this discrepancy. It is possible that when describing the simulations, participants were worried that they would forget the structure of the device if they didn't draw it immediately. They were then able to use their own drawing to help them remember the function that they observed when viewing the simulation. Another possible explanation is that, in the case of the 3-D mechanical toy, it was not immediately obvious what to draw, unlike the simulations. Participants were able to delay solving the difficult problem of how best to visually represent the toy by interleaving descriptions of the function of the device with descriptions of the structure. Moreover, by attending to the function of the device, speakers could determine what aspects of the structure were important to focus on.

corresponds to the speakers explanation and explanation phase.

We are considering what implications these findings have for the development of a multimodal design studio. Irrespective of the motivation for the varying combinations of the structure and function modes of explanation, it seems clear that the multimodal understanding componant could benefit by recognizing these two modes and adjusting its interpretation of the designer's explanation accordingly. For example, in the structure mode, anything that is drawn is probably a physical object; in the function mode, it might be an overtracing to establish reference, or a path of motion. Similarly, it seems likely that gesture frequencies vary substantially across the structure and function modes. In future work, we plan to investigate whether such a classification is valid by assessing the ability of human judges to reliably distinguish between these two modes. If this is successful, then we hope to use the evidence from the speech, sketch, and gesture to build a system that can guess which mode of explanation the speaker is in, and adjust its recognition and interpretation accordingly.

Multimodal Understanding and Integration

Adjunct to our empirical studies on speech-sketch and speech-gesture integration, we have built systems that reason about the integration patterns between these modalities. The information collected from the speech and sketching user study was used to create a set of approximately 50 integration rules, based on key words in the speech, the shape of sketched objects, and the timing between the speech and sketching events (Adler & Davis 2004). The rules segment and align the speech and sketching events, and a grammar framework recognizes certain adjectives and nouns. The grammar framework can recognize "pendulum" and adjectives such as "identical" and "touching." This enables users to say things like "there are three identical equally spaced pendulums" while sketching several pendulums. The system will then respond by making the pendulums identical and spacing them equally, as shown in Figure 6.

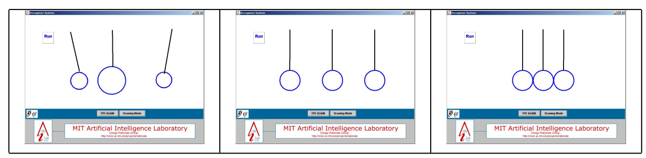

Figure 6: Three successive steps in our multimodal system. The first image shows the sketch before the user says anything. The second image shows the sketch after the user says "there are three identical equally spaced pendulums." The third image shows the sketch after the user says that the pendulums are touching.

We want to improve the system so that it can learn new speech and sketching vocabulary from the users and by making the communication between the system and the user bidirectional. This will allow the system to enter into a conversation with the user when the system does not understand what the user is doing or when the user uses a new vocabulary word. This will create a more natural interaction.

For gesture-speech integration, an optimization-based approach was applied to determine the proper alignment between referential pronouns (e.g., this, here) and gestures to the diagram (Eisenstein & Christoudias 2004). We used a salience function whose form was derived from the psycholinguistics literature on gesture-speech integration, but whose parameters were learned from a labeled corpus. This system achieved a 10% improvement in performance over previous approaches, finding the proper alignment 95% of the time. The resolution of such anaphoric references is crucial to the interpretation of multimodal utterances such as "[pointing gesture] this piece is made of wood" or "it moves [motion gesture] back and forth like this."

Future work on multimodal integration will extend these results to a more general, trimodal system for extracting semantics from speech, sketching, and gesture.

Dialogue Management

We are currently in the planning stage of a study of how to manage a spoken dialogue with a user in a design situation. Unlike previous studies, which focused on human-human interaction, here we want to focus directly on human-computer interaction. The obvious choice is to use Wizard of Oz methodology, simulating various possible dialogue strategies, with a human "wizard" who controls the interface. However, we are not aware of any Wizard of Oz studies that allow free form sketching and speech input, allowing the wizard to manipulate the sketch and control the speech output. As a first step, we will conduct the study using two humans, one acting like a wizard, but interacting directly with the other human, instead of through a computer. This will allow us to learn about the interaction between the users without developing complicated controls for the wizard.

The intent will be to determine things like: how to learn new, out-of-vocabulary terms; how to handle disfluency; how prosody reveals cues about the speakers intentions; how conversations are structured; and how often and when it is okay to interrupt the user. We will also learn about what the user sketches and what will be required to set up a true Wizard of Oz study. By studying how two users converse about new vocabulary terms, we hope to learn how to create a natural conversational interface that will allow users to add new vocabulary to our system. Our previous study examined natural communication in one direction, from the user, while this study aims to determine how to create a natural bi-directional conversation between the user and the computer.

There has been ample previous research on how to manage spoken dialogues (Rich & Sidner 1998; Allen *et al.* 2001; Glass *et al.* 2004), some of which is relevant to managing a multimodal dialogue about design. Ideas such as turn taking and determining which information you still need from the user, are still applicable. The situation is made more complicated by the additional modalities and our desire to make the interaction as natural as possible.

Domain Modeling and Reasoning

Domain modeling and reasoning about the function of objects in a domain have already been shown to be effective techniques for providing designers with intelligent decision support and design critiques. For example, in VLSI design, there are many examples of domain knowldge aiding the design of circuits (Taylor & Russell 1992). Some simple knowledge about the requirements for the width and spacing of the wires in a circuit can reveal critical errors the design. For example, if wires are too close together, a short circuit can occur. The development of perceptual user interfaces for such design support systems is a major area of our future work.

Domain knowledge can also be used improve recognition. Just as the performance of speech recognition improves in the presence of grammars that use context to limit the space of possible recognition results, ambiguities in the sketch can be resolved using knowledge about the visual grammar of the domain. (Hammond & Davis 2004) describes LADDER, a language for specifying domain-specific shape vocabularies and grammars. The use of this type of context has been shown to improve sketch recognition (Alvarado 2004).

Higher-level functional domain knowledge is also applicable. Consider a circuit diagram in which a given component might be either a battery or a bearing. The knowledge that a wire connects to a battery and the fact that a wire is indeed adjacent to the ambiguous object can resolve the uncertainty(Kurtoglu & Stahovich 2002). The incorporation of this type of high-level domain knowledge is another area of future work.

Related Work

Bolt's "Put-That-There" is the original multimodal user interface. Working in the domain of rescue mission planning, Bolt's system used pointing gestures to resolve designated keywords in the speech (Bolt 1980). The field has gradually grown to include more interesting and complex non-verbal input. Quickset (Cohen *et al.* 1997) is a multimodal interface that recognizes sketched icons, and offers a more general architecture for multimodal fusion. Quickset is usually described in applications involving course-of-action diagrams. ASSISTANCE – earlier work by the third and fifth authors of this paper – can understand multimodal explanations of device behavior including speech, sketches of device components, and sketched annotations (Oltmans & Davis 2001). The iMap system handles free-hand gestures in a map-control user interface, using prosody cues to improve gesture recognition (Krahnstoever *et al.* 2002).

Focusing explicitly on managing multimodal dialogues, Johnston et al. describe MATCH in (Johnston *et al.* 2002). MATCH includes an FST-based component for combining multimodal inputs including speech, sketch, and handwriting in the domain of map-based information retrieval. MATCH's dialogue manager enables a goal-directed conversation, using a speech-act dialogue model similar to (Rich & Sidner 1998).

All of these systems have benefitted from a series of empirical studies of multimodal communication. Oviatt et al. document users' multimodal integration patterns across speech and pen gestures in (Oviatt, DeAngeli, & Kuhn 1997). The game of Myst is used as a testbed in a wizard-of-oz study investigating modality choice in (Corradini & Cohen 2002). Cassell was among the first to argue that natural, free-hand gestures can be relevant to HCI, and presents a helpful framework for gestural interaction in (Cassell 1998). The relationship between the discourse structure and a "catchment" model of gesture is described in (Quek *et al.* 2002). A very early study on the relevance of sketch to the design process is given in (Ullman, Wood, & Craig 1990).

Conclusion

Sketching is a promising modality around which to build the design studio of the future. However, our studies suggest that sketching alone is not enough; the supporting modalities of speech and gesture must also be included. Initial work on multimodal fusion shows that this is possible, but the path from the various communicative modalities of speech, sketch, and gesture, to the semantics of a given design domain is still far from understood. We have presented some of our recent and ongoing work on this topic, and have laid out a long-term research plan, one which we hope will lead to a new and radically different role for computers in the design process.

Acknowledgements

The authors would like to thank Christine Alvarado, Sonya Cates, Tracy Hammond, Metin Sezgin, and Ron Wiken for all their help with our research and our paper. Also, thank you to Max Van Kleek for giving us helpful feedback and comments about our paper.

References

Adler, A., and Davis, R. 2004. Speech and sketching for multimodal design. In *Proceedings of the 9th International Conference on Intelligent User Interfaces*, 214–216. ACM Press.

Allen, J. F.; Byron, D. K.; Dzikovska, M.; Ferguson, G.; Galescu, L.; and Stent, A. 2001. Towards conversational human-computer interaction. *AI Magazine* 22(4):27–38.

Alvarado, C.; Oltmans, M.; and Davis, R. 2002. A framework for multi-domain sketch recognition. *AAAI Spring Symposium on Sketch Understanding* 1–8.

Alvarado, C. 2004. *Multi-Domain Sketch Understanding*. Ph.D. Dissertation, Massachusetts Institute of Technology. To be published, August 2004.

Bolt, R. A. 1980. Put-that-there: Voice and gesture at the graphics interface. In *Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques*, 262–270.

Cassell, J. 1998. A framework for gesture generation and interpretation. In *Computer Vision in Human-Machine Interaction*, 191–215. Cambridge University Press.

Cohen, P. R.; Johnston, M.; McGee, D.; Oviatt, S.; Pittman, J.; Smith, I.; Chen, L.; and Clow, J. 1997. Quickset: Multimodal interaction for distributed applications. In *ACM Multimedia'97*, 31–40. ACM Press.

Corradini, A., and Cohen, P. R. 2002. On the relationships among speech, gestures, and object manipulation in virtual environments: Initial evidence. In *2002 CLASS Workshop on Natural, Intelligent, and Effective Interaction in Multimodal Dialog Systems*.

Davis, R. 2002. Sketch understanding in design: Overview of work at the MIT AI lab. *Sketch Understanding, Papers from the 2002 AAAI Spring Symposium* 24–31.

Eisenstein, J., and Christoudias, C. M. 2004. A salience-based approach to gesture-speech alignment. In *HLT-NAACL'04*, 25–32. ACL Press.

Eisenstein, J., and Davis, R. 2003. Natural gesture in descriptive monologues. In *UIST'03 Supplemental Proceedings*, 69–70. ACM Press.

Glass, J.; Weinstein, E.; Cyphers, S.; Polifroni, J.; Chung, G.; and Nakano, M. 2004. A framework for developing conversational user interface. In *Proceedings of the 4th International Conference on Computer-Aided Design of User Interfaces*, 354–365.

Hammond, T., and Davis, R. 2004. Automatically transforming symbolic shape descriptions for use in sketch recognition. *Proceedings of the Nineteenth National Conference on Artificial Intelligence (AAAI-04)*.

Johnston, M.; Ehlen, P.; Walker, M.; Whittaker, S.; and Maloor, P. 2002. Match: An architecture for multimodal dialogue systems. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 276–383.

Krahnstoever, N.; Kettebekov, S.; Yeasin, M.; and Sharma, R. 2002. A real-time framework for natural multimodal interaction with large screen displays. In *Proc. of Fourth Intl. Conference on Multimodal Interfaces (ICMI 2002), Pittsburgh, PA, USA*.

Kurtoglu, T., and Stahovich, T. F. 2002. Interpreting schematic sketches using physical reasoning. *Sketch Understanding, Papers from the 2002 AAAI Spring Symposium* 78–85. interpreting meaning, what happens, mechanical.

Landay, J. A., and Myers, B. A. 1995. Interactive sketching for the early stages of user interface design. In *Proceedings of CHI '95: Human Factors in Computing Systems*, 43–50.

McNeill, D. 1992. *Hand and Mind*. The University of Chicago Press.

Oltmans, M., and Davis, R. 2001. Naturally conveyed explanations of device behavior. In *Workshop on Perceptive User Interfaces*.

Oviatt, S.; DeAngeli, A.; and Kuhn, K. 1997. Integration and synchronization of input modes during multimodal human-computer interaction. In *Conference Proceedings on Human Factors in Computing Systems*, 415–422.

Quek, F.; McNeill, D.; Bryll, R.; Duncan, S.; Ma, X.-F.; Kirbas, C.; McCullough, K. E.; and Ansari, R. 2002. Multimodal human discourse: gesture and speech. *Transactions on Computer-Human Interaction* 9(3):171–193.

Rich, C., and Sidner, C. 1998. COLLAGEN: A collaboration manager for software interface agents. *User Modeling and User-Adapter Interaction* 8(3–4):315–350.

Seneff, S.; Hurley, E.; Lau, R.; Pao, C.; Schmid, P.; and Zue, V. 1998. Galaxy-II: A reference architecture for conversational system development. In *Proceedings of ICSLP'98*, volume 3, 931–934.

Sezgin, T. M.; Stahovich, T.; and Davis, R. 2001. Sketch based interfaces: Early processing for sketch understanding. In *The Proceedings of 2001 Perceptive User Interfaces Workshop (PUI'01)*.

Taylor, G. E., and Russell, G., eds. 1992. *Algorithmic and Knowledge-Based CAD for VLSI*. Institution of Electrical Engineers.

Ullman, D. G.; Wood, S.; and Craig, D. 1990. The importance of drawing in the mechanical design process. *Computers and Graphics* 14(2):263–274.