Informal PUIs: No Recognition Required

James A. Landay, Jason Hong, Scott Klemmer, James Lin, Mark Newman
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley
Berkeley, CA 94720-1776
landay@cs.berkeley.edu

Abstract

The limitations of graphical user interfaces have slowed the spread of computer usage to the entire population. Perceptual user interfaces are one approach that can overcome many of these limitations. Adding perceptual capabilities, such as speech, sketching, and vision, is the key to making interfaces more effective. We argue that informal user interfaces, which do little or no up-front recognition of the perceptual input, have important applications and should not be forgotten by perceptual user interface researchers.

1. Introduction

Today's graphical user interfaces are limiting the use of computers to only a segment of our society and in only a few locations: the desktops of information workers and the homes of a large, though noninclusive, portion of our population. Current interfaces do not let users communicate in ways that they naturally do with other human beings (National Research Council 1997). Moreover, interfaces do not take advantage of many of our innate perceptual, motor, and cognitive abilities. As our society attempts to bring about universal access to a National Information Infrastructure and better-paying jobs to a larger percentage of the population, interfaces will need to rely more on these abilities.

Overcoming the limitations of GUIs is leading us towards a future of Perceptual User Interfaces (PUIs). These interfaces try to "perceive" what the user is doing using computer vision, speech recognition, and sketch recognition. Applications can then take advantage of a wide variety of human abilities, such as speech or gesture, to achieve a more natural human-computer communication. We maintain that informal user interfaces, those without much up-front recognition or interpretation of the input, will be valuable in future perceptual applications. In particular, applications in support of creative design or human-human communication may be better served by preserving the ambiguity and lack of rigid structure inherent in uninterpreted perceptual input.

2. WIMPy User Interfaces

Although easier to use than the cryptic command-line interfaces that preceded them, conventional WIMP (windows, icons, menus, and pointing) interfaces are still too hard to use and limited in when and where they can be used. These problems come mainly from attempts to make them overly general and from relying on only a few perceptual and motor skills.

A non-trivial percentage of the population is blind or has trouble seeing words and letters in ordinary newsprint [5% of those over age 15 (National Research Council 1997)] and this percentage will only increase as our population ages. The GUI has been far from a boon for these users. Many others have limited literacy skills [21% of Americans over age 16 (U.S. Department of Education 1992)] typing skills, or use of their hands. The later is often a result of using GUIs.

More importantly, the activities required by today's interfaces (e.g., sitting, looking at a screen, typing, and pointing with a mouse) are too awkward for many situations. Workers and consumers alike perform tasks in many different locations. For instance, doctors, delivery people, and salespeople are often mobile while performing their jobs. Likewise, in our homes, we often perform different tasks in different parts of the house (e.g., we might cook in the kitchen, entertain in the living room, and read in the bedroom). The standard GUI does not work well when users are standing up, are using their hands for something else, or are interacting with another person.

A computational infrastructure that *restricts* users to one device in one fixed location is simply that, restrictive. The problems stated above are a result of relying on an interface paradigm that was fine in its day, but has grown old, bloated, and incapable of meeting the needs of the future.

3. Perceptual UIs Can Breakout of the GUI Box

How can we move beyond the 25 year-old GUI legacy of the Xerox Star? During an invited talk at the CHI '97 conference, Bill Buxton urged researchers to quit "taking the GUI as a given" and instead push harder from other starting points in the design space. In particular, he encouraged researchers to move towards the goal of an

"invisible computer."

One such push has come from researchers working on *perceptual user interfaces* (PUIs) (Pentland and Darrell 1994, Turk 1997, Huang et al. 1995). These interfaces support natural modes of input by having the computer try to "perceive" what the user is doing. The perceptual processing has generally been restricted to speech recognition and computer vision. This definition of perceptual user interface is overly restrictive. Humans perceive sounds and light and then recognize them as speech and visual objects, respectively. Likewise, using our visual system we perceive handwritten words and sketches and then recognize what the symbols mean. People are comfortable using their sketching and writing skills everyday for many different tasks. We are missing an opportunity to bring natural interfaces to more people by neglecting these modes. Thus, the definition of perceptual user interface should also include interfaces using handwritten and sketched input.

4. Informal and Formal User Interfaces

An important but orthogonal question is how and when these natural modes should be interpreted. Immediate recognition imposes structure on the input that can often get in the way during creative or communications-oriented tasks, such as brainstorming (Moran et al. 1995) or design (Landay and Myers 2001). An alternative, informal approach has recently gained research interest. We use the term *informal user interfaces* to describe interfaces designed to support natural, ambiguous forms of human-computer interaction without much up-front recognition or transformation of the input. Informal interfaces are a reaction to the "overly restrictive computational formats" (National Research Council 1997) inherent in current applications.

These restrictive interfaces are often a result of tradition. Computers have found their greatest success in applications where they have *replaced* humans in tasks at which humans are poor, such as performing complex, redundant mathematical calculations, or storing and searching large amounts of data. We use the term *formal user interfaces* to describe these precision-oriented interfaces and perceptual interfaces that try to recognize human input immediately.

As computers grow more powerful, less expensive, and more widespread, we expect them to *assist* us in tasks that humans do well, such as writing, drawing, and designing. Unfortunately, the historical strengths of computers have led to a design bias towards precise computation, and away from the more human properties of creativity and communication. Consequently, interfaces are designed to facilitate structured input rather than natural human communication, which consists of the imprecise modes of speaking, writing, gesturing, and sketching.

Most of the work on perceptual user interfaces and sketch understanding has been biased towards precise computation and so tries to unambiguously interpret natural input as soon as possible and structure the data in a form that is best for the computer. There are several domains that are better served by a more informal approach.

4.1 Informal Sketching UIs

Design is one domain where an informal approach may be more appropriate. Current design applications generally require the user to drag objects from a "palette" to a "canvas" on which they are manipulated using a mouse. These objects, such as lines and circles, are often placed very precisely with an exact location and size. This precision makes it easy for the computer to represent the objects in the system. A precise representation is also important for some tasks, such as constructing the mechanical drawings for a part that is to be machined. However, this drawing interface is quite different from the way we construct drawings without computers: we often sketch rough, ambiguous drawings using pencil and paper.

Sketching is one mode of informal, perceptual interaction that has been shown to be especially valuable for creative design tasks (Gross and Do 1996, Wong 1992). For designers, the ability to sketch ambiguous objects (i.e., those with uncertain types, sizes, shapes, and positions) rapidly is very important to the creative process. Ambiguity encourages the designer to explore more ideas in the early design stages without being burdened by concern for inappropriate details such as colors, fonts, and precise alignment. Leaving a sketch uninterpreted, or at least in its rough state, is the key to preserving this fluidity (Goel 1995).

In this early phase, ambiguity also improves communication, both with collaborators and the target audience of the designed artifact. For example, an audience examining a sketched interface design will be inclined to focus on the important issues at this early stage, such as the overall structure and flow of the interaction, while not being distracted by the details of the look, such as, colors, fonts, and alignment. When the designer is ready to move past this stage and focus on the details, the interface can be recreated in a more formal and precise way. We have built several design tools that incorporate this style of interface. SILK (Landay and Myers 2001) is targeted at graphical user interface design and DENIM (Lin et al. 2000) is targeted at web site design (see Figure 1). DENIM has little built-in sketch recognition.
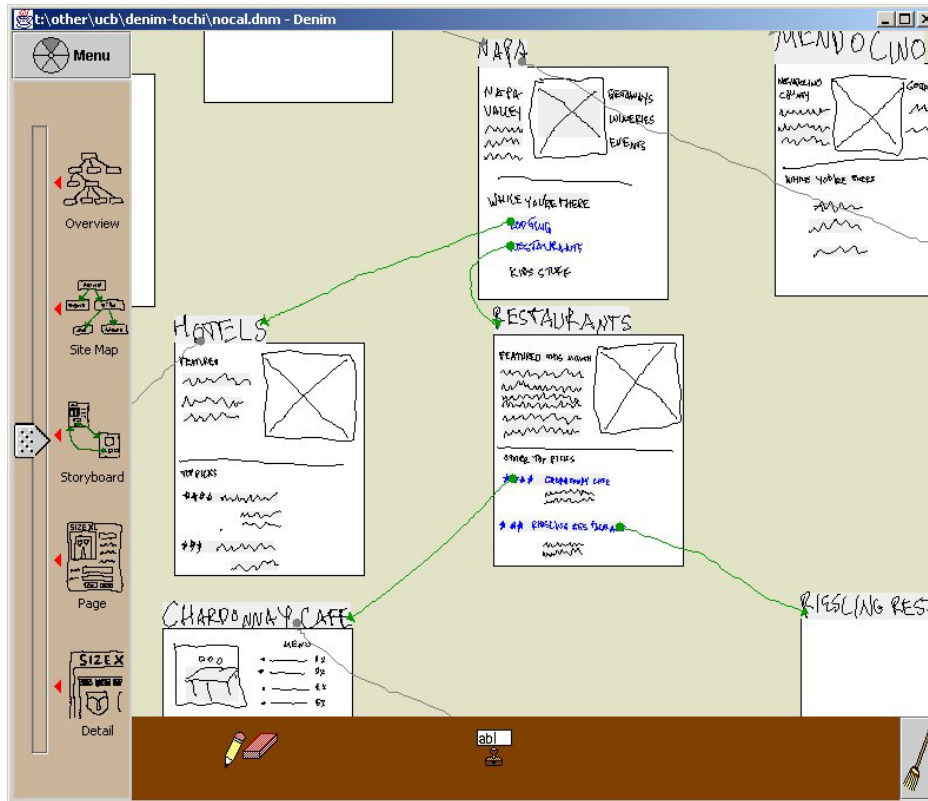
Figure 1. DENIM is a sketch-based web design tool that takes advantage of an informal user interface. In DENIM's storyboard view, multiple pages and the links between them are visible. DENIM tries to group ink on pages into links or graphical objects. Recognizing links between items and pages is the only other sketch recognition performed by DENIM.

4.2 Informal Speech UIs

Speech is another perceptual style of interaction that is starting to become popular. Although speech is natural and inherently human, current speech recognition systems do not have informal user interfaces. Consider that these systems generally fail when faced with the ambiguity inherent in human speech (e.g., "recognize speech" vs. "wreck a nice beach"). Also, current speech recognition systems try to recognize human speech and *translate* it into an accurate, internal representation that machines can deal with. These tools also often require or encourage the user to enter a dialog with the machine to correct mistakes as they occur[1]. Although this behavior might be appropriate for purchasing an airline ticket over the phone, it gets in the way when the user wants to write, design, or brainstorm.

Several systems from the MIT Media Lab Speech Group had interfaces that could be considered informal speech user interfaces. For example, this group produced a handheld speech notetaker that only uses recognition for

organizing the uninterpreted notes (Stifelman et al. 1993). They also built a paper-based notebook that synchronizes uninterpreted speech with handwritten notes (Stifelman 1996). This area warrants more exploration.

Our research group at Berkeley has developed SUEDE, a speech UI design tool that uses an informal interface (Klemmer et al. 2000). SUEDE allows designers to record their own voice for system prompts and then plays those prompts back when testing the interface with test participants (see Figure 2). During a test, the designer acts as a "wizard" by listening to the test participant's responses and then "recognizes" and chooses the next legal state of the interaction. By not using a speech recognizer and an associated grammar, the wizard can accommodate multiple responses from the participant and can use the log of those responses later in the design process to develop a robust grammar.

By not requiring the designer to deploy real speech recognition and synthesis, the interface design process is expedited in the early stages of design. This allows designers to make several iterations on their early design ideas. Fast design iteration has been found to be one of the keys to creating high quality user interfaces.

---

[1] Some dictation systems have a mode that allows the user to ignore corrections until after the capture session.
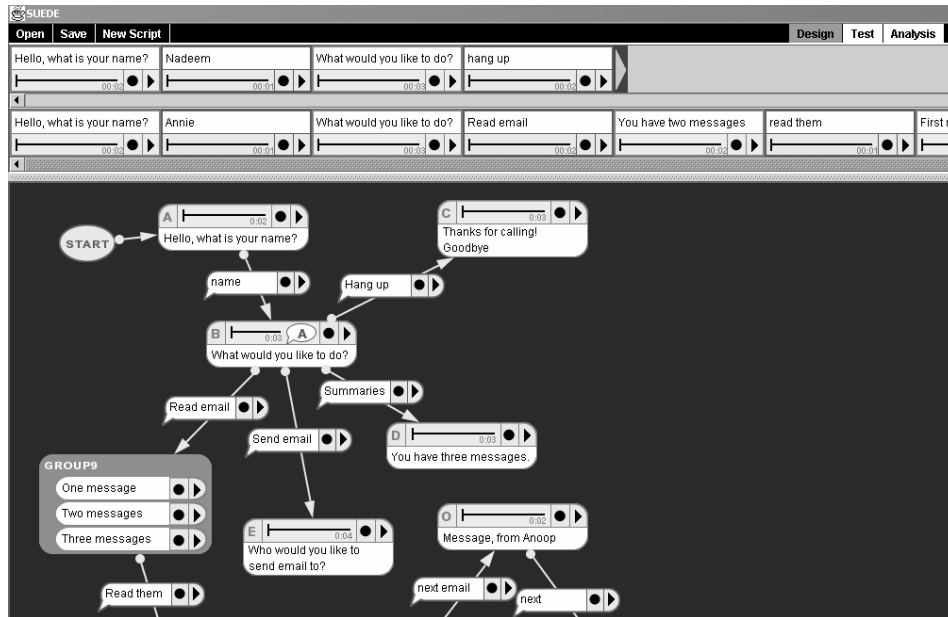
Figure 2. SUEDE is a speech user interface design tool that allows a designer to create, test, and analyze early interface design ideas. This tool does no speech recognition or synthesis. It instead relies on the designer's own voice for prompts and their own intelligence to recognize the end-user's utterances.

4.3 Informal Handwriting UIs

Handwriting recognition is another style of interaction that suffers from some of the same problems as speech recognition. Again, many systems try to recognize writing as the user writes. This might be useful for taking down someone's phone number, but begins to get in the way when the user is trying to focus on the task at hand (e.g., taking notes in a talk or writing up a new idea). The dialog with the recognition system requires users to focus their attention on the machine, rather than on their task. Deferred recognition is one way of dealing with this problem. We have built systems that treat electronic ink as a first-class type and let users write like they naturally do on paper (Davis et al. 1999). Others have built similar informal systems that allow searching of electronic ink (Poon, Weber, and Cass 1995).

5. Conclusions

The primary tenet of work in informal user interfaces is to bend computers to people's mode of interaction, not the other way around. For some domains it is clear that the formal approach imposed by conventional interfaces presents an obstacle to effective task performance. One such domain that we have investigated is user interface design itself, and this led to our work on SILK (Landay and Myers 2001) and DENIM (Lin et al. 2000), sketching tools for early stage interface and web design, respectively. We believe that informal speech and handwriting interfaces may also be useful. Any application that requires interaction between humans and computers, and in which absolute precision is unnecessary, could benefit from an informal approach.

PUIs relying on a single input modality will likely only see success in limited domains. This is for the same reasons that GUIs are limited: using only a single input modality will be good for some tasks and not for others. On the other hand, a multimodal interface with multiple perceptual and GUI capabilities will often be more natural to use (Mignot, Valot and Carbonell 1993, Waibel et al. 1997). Successful multimodal user interfaces may support uninterpreted sketches and other informal styles, in addition to interpreted speech, recognized sketches and other more-studied modalities.

References

Davis, R.C., et al. 1999. NotePals: Lightweight Note Sharing by the Group, for the Group. In Proceedings of *Human Factors in Computing Systems: CHI '99*, 338-345. Pittsburgh, PA.

Goel, V. 1995. *Sketches of Thought*. Cambridge, MA: The MIT Press.

Gross, M.D. and Do, E.Y. 1996. Ambiguous Intentions: A Paper-like Interface for Creative Design. In Proceedings of *ACM Symposium on User Interface Software and Technology*, 183-192. Seattle, WA.

Huang, X.; Acero, A.; Alleva, F.; Hwang, M.Y.; Jiang, L.; and Mahajan, M. 1995. Microsoft Windows highly intelligent speech recognizer: Whisper, in *Proceedings of 1995 International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, 93-6

Klemmer, S.R.; Sinha, A.K.; Chen, J.; Landay, J.A.; Aboobaker, N.; and Wang, A., SUEDE: A Wizard of Oz Prototyping Tool for Speech User Interfaces. *CHI Letters*, 2000. 2(2): 1-10.

Landay, J.A. and Myers, B.A., Sketching Interfaces: Toward More Human Interface Design. *IEEE Computer*, 2001. 34(3): 56-64.

Lin, J.; Newman, M.W.; Hong, J.I.; and Landay, J.A., DENIM: Finding a tighter fit between tools and practice for web site design. *CHI Letters*, 2000. 2(1): 510-517.

Mignot, C.; Valot, C.; and Carbonell, N. 1993. An Experimental Study of Future 'Natural' Multimodal Human-Computer Interaction, in *Proceedings of ACM INTERCHI'93 Conference on Human Factors in Computing Systems -- Adjunct Proceedings*. 67-68.

Moran, T.P.; Chiu, P.; Melle, W.v.; and Kurtenbach, G. 1995. Implicit Structures for Pen-Based Systems Within a Freeform Interaction Paradigm. In Proceedings of *Human Factors in Computing Systems*, 487-494. Denver, CO.

National Research Council. 1997. *More than screen deep: toward every-citizen interfaces to the nation's information infrastructure*. Washington, D.C.: National Academy Press.

Pentland, A.P. and Darrell, T. 1994. Visual perception of human bodies and faces for multi-modal interfaces, in *1994 International Conference on Spoken Language Processing*, vol. 2, 543-6

Poon, A.; Weber, K.; and Cass, T. 1995. Scribbler: A Tool for Searching Digital Ink. In Proceedings of *Human Factors in Computing Systems*, 252-253. Denver, CO.

Stifelman, L.J. 1996. Augmenting Real-World Objects: A Paper-Based Audio Notebook. In Proceedings of *Human Factors in Computing Systems*, 199-200. Vancouver, Canada.

Stifelman, L.J.; Arons, B.; Schmandt, C.; and Hulteen, E.A. 1993. VoiceNotes: A Speech Interface for a Hand-Held Voice Notetaker, in *Proceedings of ACM INTERCHI'93 Conference on Human Factors in Computing Systems*. 179-186.

Turk, M. First Workshop on Perceptual User Interfaces, ed: Banff, Alberta, Canada, 1997.

U.S. Department of Education. 1992. *1992 National Adult Literacy Survey*. Washington, DC: U.S. Government Printing Office.

Waibel, A.; Suhm, B.; Vo, M.T.; and Yang, J. 1997. Multimodal interfaces for multimedia information agents, in *Proceedings of 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 1, 167-70.

Wong, Y.Y. 1992. Rough and Ready Prototypes: Lessons From Graphic Design. In Proceedings of *Human Factors in Computing Systems*, 83-84. Monterey, CA.