

A Prototype User Interface for a Mobile Multimedia Terminal¹

Allan Christian Long, Jr., Shankar Narayanaswamy, Andrew Burstein, Richard Han, Ken Lutz, Brian Richards, Samuel Sheng, Robert W. Brodersen, Jan Rabaey

Department of Electrical Engineering and Computer Sciences
The University of California at Berkeley, Cory Hall
Berkeley, CA 94720
allanl@cs.berkeley.edu²
+1-510-642-8814

KEYWORDS

Human computer interaction, mobile computing, speech recognition, handwriting recognition, pen-based computing, multimedia, multi-modal input.

ABSTRACT

We have shown a prototype user interface for the InfoPad, a portable terminal with multi-modal input and multimedia output. We believe that many of the people who could benefit from inexpensive, portable, networked terminals are not computer experts, and we are therefore designing the InfoPad and its user interface to be more like a notebook than a workstation. The InfoPad's main features are:

- Portability
- Continuous network connectivity using a high-bandwidth radio link
- Pen input with handwriting recognition
- Audio input with speech recognition
- Full-motion video playback with synchronized audio

The InfoPad's unique input and output characteristics offer challenges and opportunities for user interface design. We are prototyping applications and user interfaces to explore how handwriting and voice recognition may best be used together. We believe that the lessons we will learn can be applied to other multi-modal platforms.

INTRODUCTION

The Infopad project is a large, multi-disciplinary research effort involving a number of faculty and graduate students. Its goal is to build portable multimedia terminals connected to the network via high-bandwidth radio links (2 Mbits/sec

downlinks and 64 kBits/sec uplinks) in a picocellular environment. The research encompasses low-power integrated circuit design, high-frequency radio design, network design, handwriting and speech recognition, and user interface design.

Although we plan for the InfoPad to have a color display, current LCD technology is still expensive in both money and power. Therefore the present implementation uses a 9-inch monochrome text and graphics screen and a 4-inch color screen for full-motion video. A speaker is included for audio output. A keyboard is not included because it would add bulk, weight, and cost to the terminal.

InfoPad uniquely combines portability, network connectivity, and state-of-the-art interface technology. We believe the ideal user interfaces for InfoPad are fundamentally different from WIMP (window, icon, mouse, and pointer) interfaces.

The three technologies InfoPad brings together make it an ideal platform for many application areas not well supported by traditional workstations:

- Groupware (meeting support, collaborative design)
- Note taking and document annotation
- Mobile information browsing and retrieval
- Personal information management
- Personal communications

These applications are inherently mobile, are well-suited to the use of pointing devices in the user interface, and do not require mass text entry, which we feel is more efficiently done with a keyboard. For some of these applications, we can use existing workstation versions as starting points; for others, we must develop interfaces from the ground up.

We do not intend that the InfoPad attempt to surpass workstations in areas in which they already excel. For example, programming is more time-efficient on a desktop workstation than on an InfoPad since a proficient typist can type at least twice as fast as he or she can write. The InfoPad is a complement to workstations, not a competitor.

1. This work was supported by ARPA and the InfoPad Partners.

2. To reach other authors, use

<http://infopad.eecs.berkeley.edu/>

Since the InfoPad's input modalities are so different from those of traditional computers, we believe that even where the same application (e.g. e-mail, text editing) is implemented for the InfoPad, the user interface should be redesigned to take advantage of InfoPad's strengths. For example, although one could easily write a keyboard widget to enter text into one's favorite text editor, it would be better to implement a text editor that used the InfoPad's speech and handwriting capabilities.

Designing a user interface for the InfoPad's input modalities raises many questions. Some of the issues we intend to address are:

- How should a user interface deal with the uncertainty inherent in handwriting and speech recognition?
- To what extent can or should pen and voice be used as substitutes for mouse and keyboard?
- How can handwriting and speech recognition be used synergistically?

SYSTEM SOFTWARE ARCHITECTURE

Due to size and cost constraints on the InfoPad terminal, we moved as much processing as possible off the terminal and onto the network. The greatest advantage of this architecture is that the InfoPad has access to massive computational power, allowing the InfoPad to be "smarter" (e.g. with handwriting and speech recognition) than other portable devices'. The most significant disadvantage to the user interface is latency. We are optimizing our network software so that network latency does not cause our interface to be unresponsive.

PEN INPUT AND HANDWRITING RECOGNITION

Applications may use pen data from the InfoPad in three ways. First, they may treat the pen as a mouse, reading mouse events from the X Window System. Second, applications can bypass X to get higher resolution data from the pen. For example, a drawing program would likely want as much resolution as possible. Third, an application can use a handwriting recognizer, described below, to treat the pen almost like a keyboard.

An application programming interface (API) is provided for application programmers to easily access and control the recognizer. An application may use more than one recognizer to customize the grammar and vocabulary for different contexts. A handwriting recognition widget provides an easy-to-use abstraction for graphical user interface builders. The widget accepts handwriting and allows the user to correct it before returning the recognized text to the application. The application does not need to be aware that data is being entered by pen rather than on a keyboard.

AUDIO INPUT AND SPEECH RECOGNITION

Raw audio is available for applications such as telephony and voice annotation. Alternatively, applications may use the network based, continuous-word, speaker-independent speech recognizer. To increase its accuracy, the recognizer may be customized to individual speakers. Also, the recognizer can produce several estimates of the spoken sentence, increasing the likelihood of producing the correct sentence. The recognizer exports an API so programmers may easily incorporate recognition into applications.

The speech vocabulary and grammar are application- and context-specific. A word constructor program allows users to add words or customized pronunciations simply by speaking to the InfoPad. Another program allows programmers to construct vocabularies and grammars using a graphical interface.

FULL-MOTION VIDEO

Video is becoming pervasive in digital documents, so we think InfoPad users should have access to video as well as text and graphics. For example, we want to support Mosaic access to video documents.

The InfoPad's color display is 18 bits deep with a resolution of 128x240 (using wide pixels for a 4:3 aspect ratio). On this screen, the InfoPad can play full-motion digitized video at 30 frames per second. We use vector quantization (VQ) encoding for two reasons. One is to compress the video to our radio downlink bandwidth of 2 Mbits/s. The other reason is that VQ (unlike many other encoding schemes such as MPEG) is tolerant of data errors introduced by the radio link. We plan to further explore the trade-offs between different compression schemes.

APPLICATIONS

We are building applications to test our user interfaces and to demonstrate the usefulness of the InfoPad system. We are concentrating especially on applications that take advantage of both pen and speech input.

The first application is a Mosaic-like WWW client, which demonstrates the network access and retrieval, multimedia output, and recognition capabilities of the InfoPad. The second is a voice-driven command interface for the Magic integrated circuit layout editor. This demonstrates the use of voice commands in driving pre-written applications and can be used on regular workstations as well. The third application is a circuit schematic editor that will recognize text and schematic symbols drawn by the pen, as well as speech commands, to create and edit circuit schematics and simulate them in SPICE.

CONCLUSIONS AND FUTURE DIRECTIONS

Multimedia output, mobility, network connectivity, and recognized input make the InfoPad project unique. Pen and audio input, and in particular handwriting and speech recognition, make for a more natural user interface and will allow us to go beyond the traditional WIMP interface. We believe that the results we obtain from our exploration of pen and voice interaction techniques will be useful for designers of interfaces for many different applications on many different computers.