Perceptual Interfaces

Adapted from
Matthew Turk’s (UCSB) and
George G. Robertson’s (Microsoft Research)
slides on perceptual interfaces

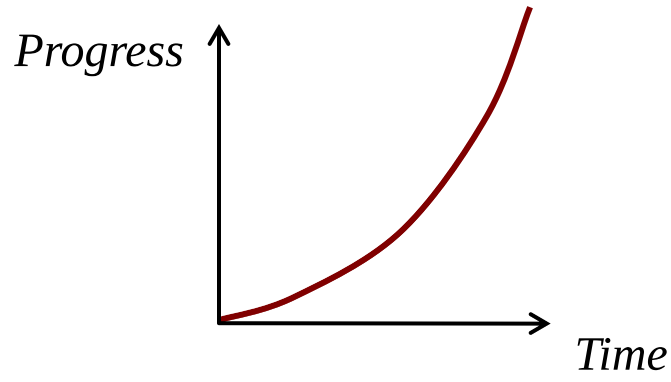
Outline

- ✓ Why Perceptual Interfaces?
- ✓ Multimodal interfaces
- ✓ Vision Based Interfaces (VBI)
- ✓ Examples

Observation

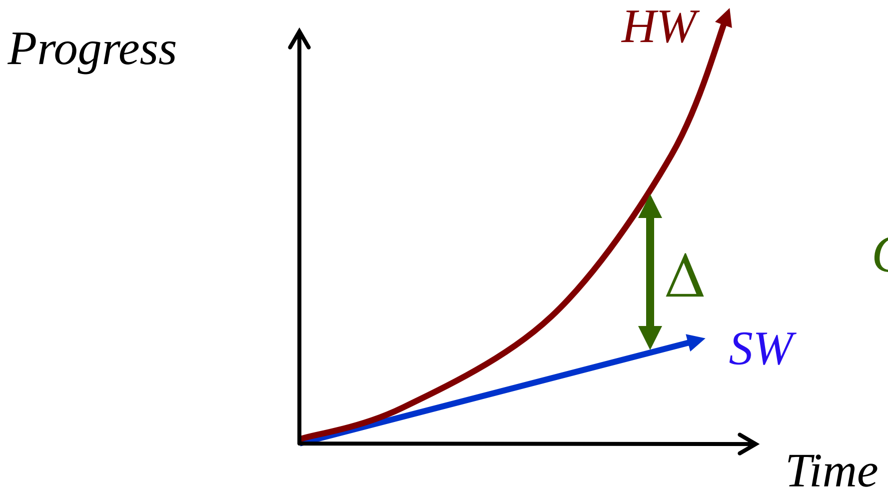
- **Moore's Law** has driven computer technology for decades

  Exponential improvement in HW
  - 5 years ~ 10x improvement
  - 10 years ~ 100x improvement
  - 20 years ~ 10,000x improvement



- But… there has been no Moore's Law for user interfaces!
  - The result?

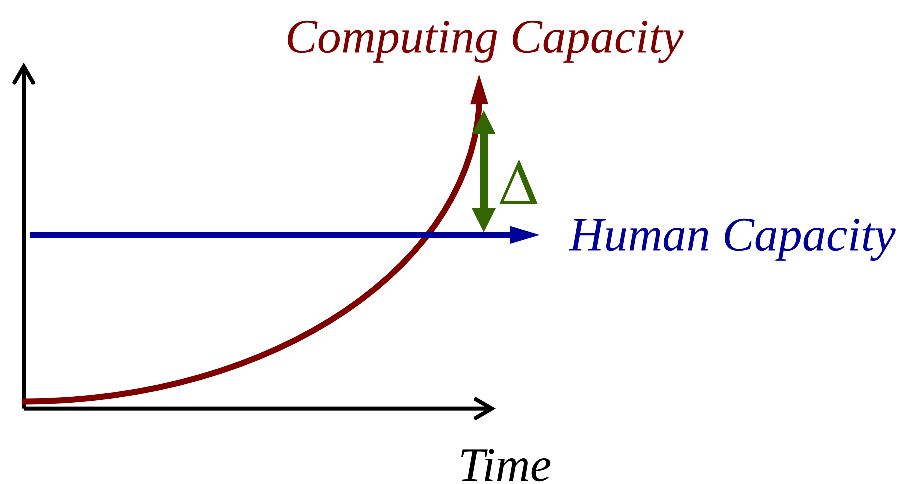
The result



Curse of the delta!

Another view:

There's no Moore's Law for people!



Curse of the delta

Evolution of user interfaces

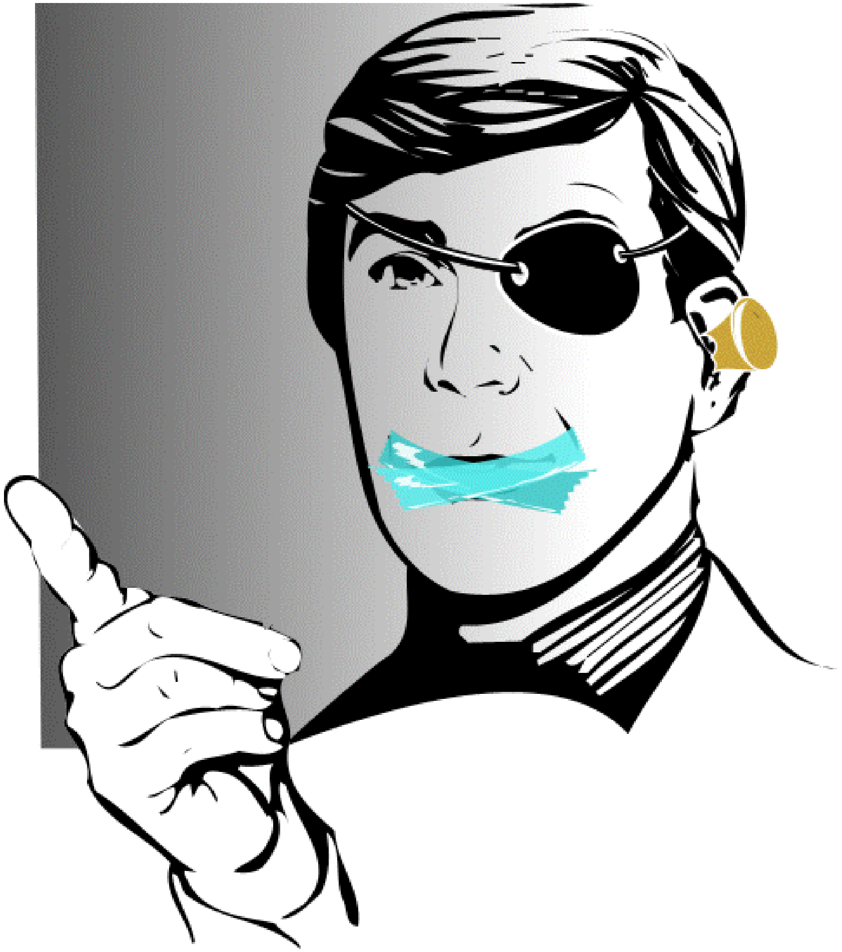
When	Implementation	Paradigm
1950s	Switches, punched cards	None
1970s	Command-line interface	Typewriter
1980s	Graphical UI (GUI)	Desktop
2000s	???	???

Current UI Limitations
Failure to use Human Abilities

Limited Vision (Flat, 2D)

No Speech

No Gestures



Limited Audio

One Hand Tied Behind Back

Limited Tactile

The Next Big Thing in UI?

- Immersive environments
  - Wearable computers, Virtual Reality, Augmented Reality…
- Ubiquitous Computing
  - Invisible, pervasive
- Tangible UI
  - Coupling of physical objects and digital data
- Multimodal UI
  - Sound, speech, gesture…
- Affective Computing
  - Computers that understand and express emotion

Evolution of user interfaces

When	Implementation	Paradigm
1950s	Switches, punched cards	None
1970s	Command-line interface	Typewriter
1980s	Graphical UI (GUI)	Desktop
2000s	Perceptual UI (PUI)	Natural interaction

Perceptual Interfaces

Highly interactive, multimodal interfaces modeled after natural human-to-human interaction

- Goal: For people to be able to interact with computers in a similar fashion to how they interact with each other and with the physical world

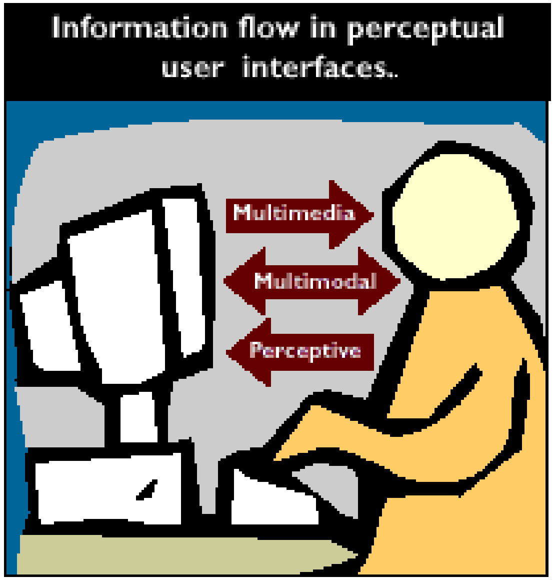
Not just passive

Multiple modalities, not just mouse, keyboard, monitor

“Perceptual” User Interfaces

- **Perceptive**
  - human-like perceptual capabilities (what is the user saying, who is the user, where is the user, what is he doing?)
- **Multimodal**
  - People use multiple modalities to communicate (speech, gestures, facial expressions, …)

- Multimedia
  - Text, graphics, audio and video



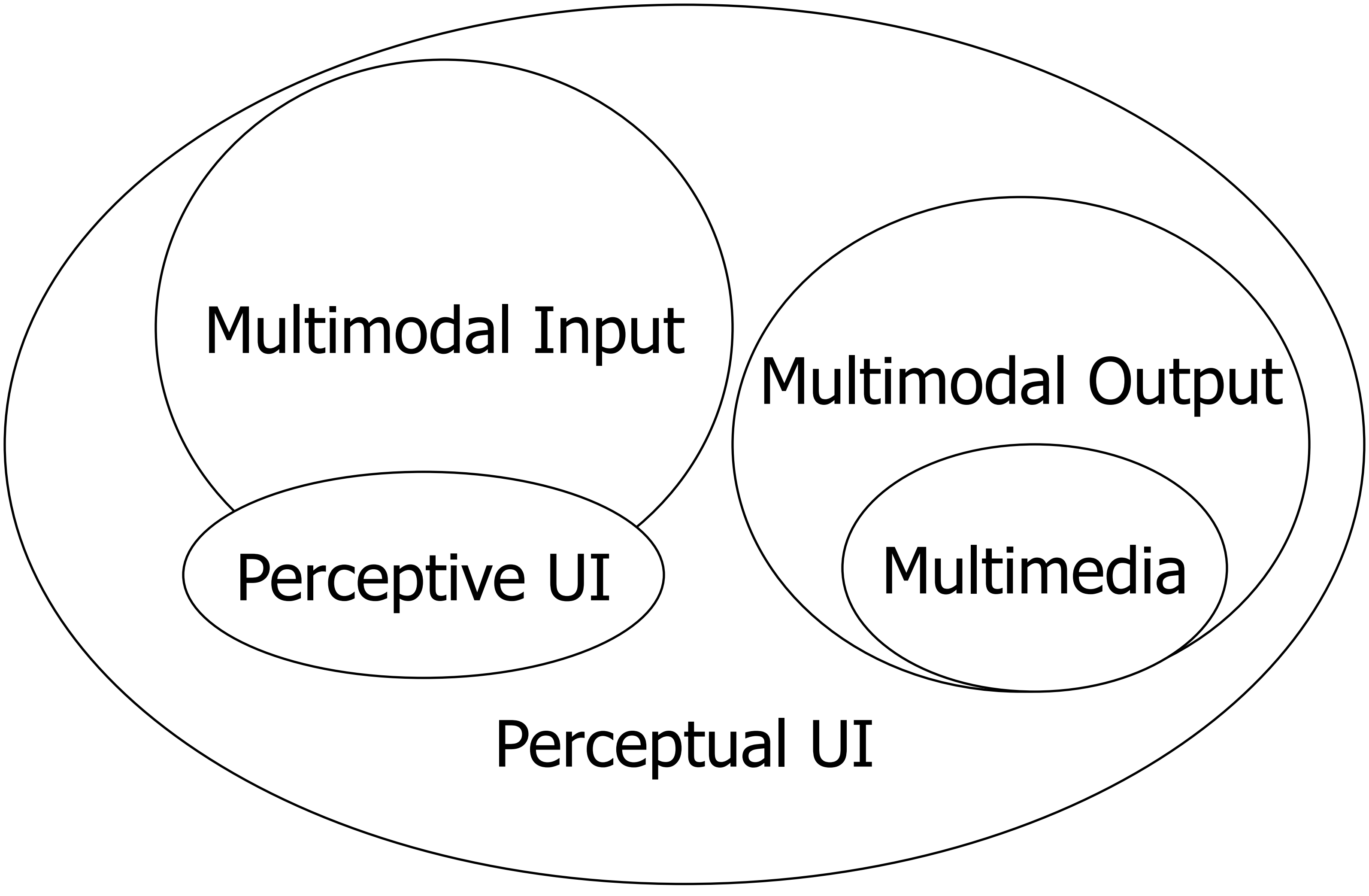
Perception

- In order to respond appropriately, objects/room need(s) to pay attention to
  - **People** and
  - **Context**

- Machines have to be *aware* of their environment:
  - **Who, What, When, Where and Why**?

- Interfaces must be **adaptive** to
  - Overall situation
  - Individual User

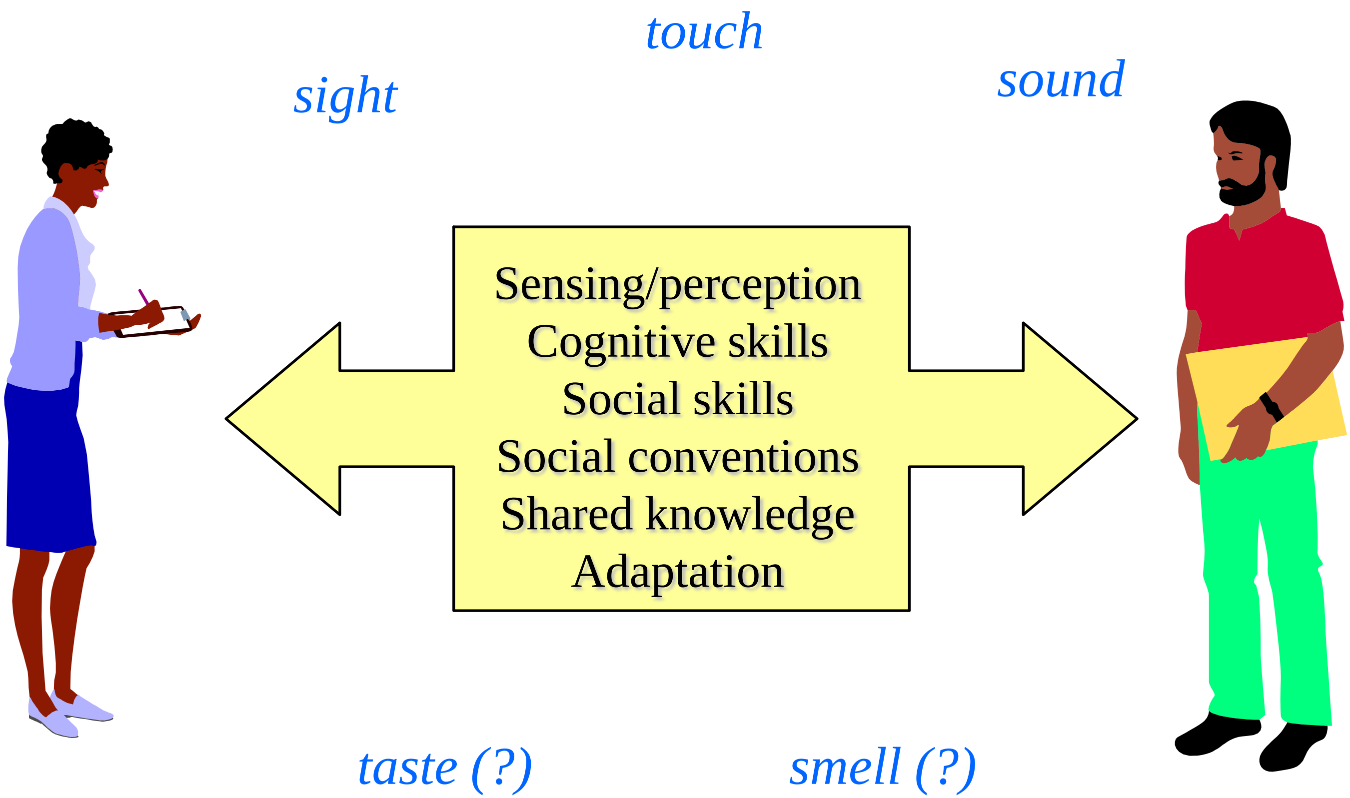
How Do The Pieces Fit?



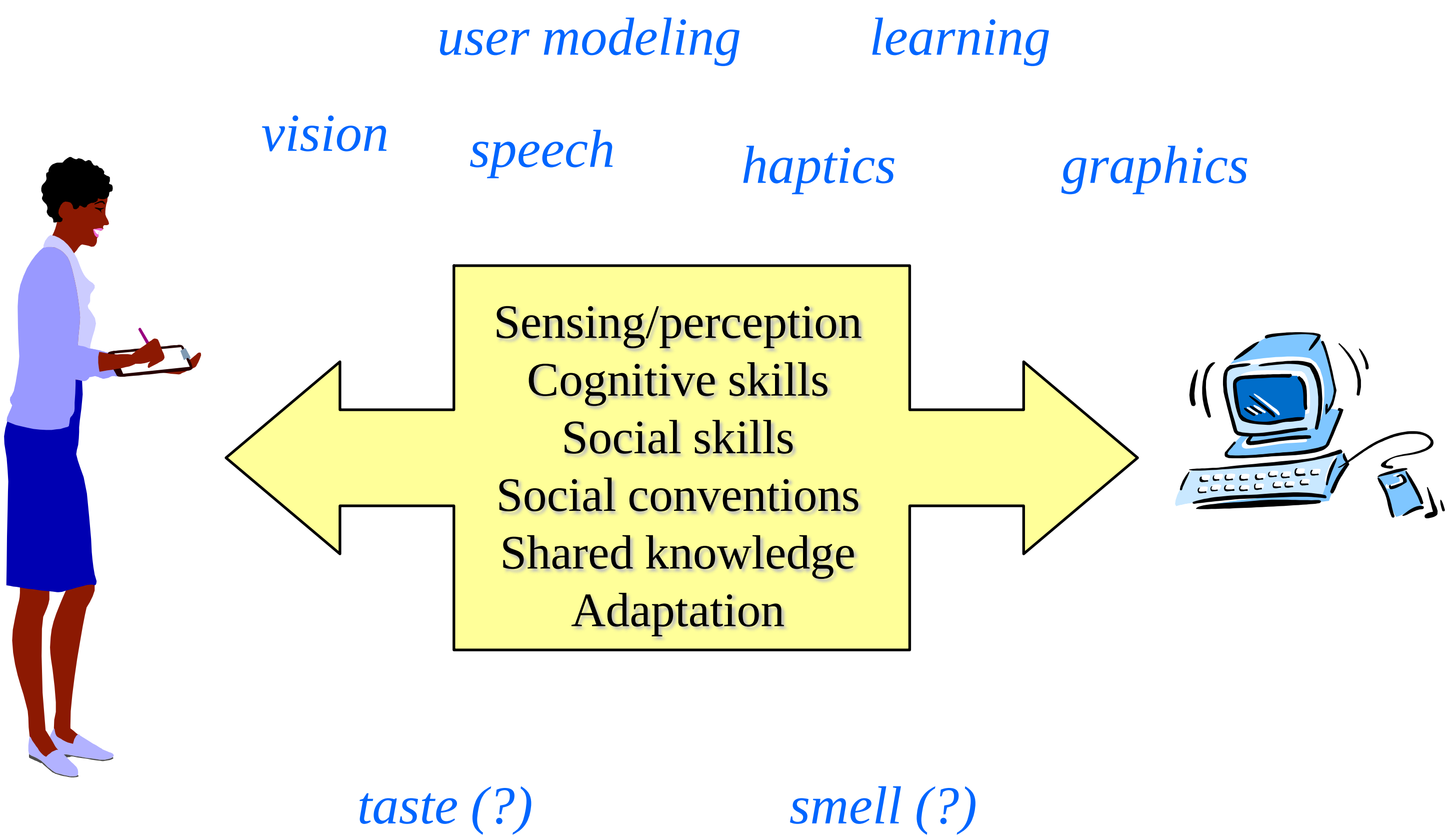
Perceptual User Interfaces (PUI)

- Special section on PUIs in the March 2000 issues of *Communications of the ACM*, edited by Matthew Turk and George Robertson.
- PUIs combine natural human capabilities of communication, motor, cognitive, and perceptual skills with computer I/O devices, machine perception, and reasoning.
- Integrate research results from different disciplines
  - vision, speech, graphics and visualization, user modeling, haptics, and cognitive psychology

Natural human interaction



Perceptual Interface

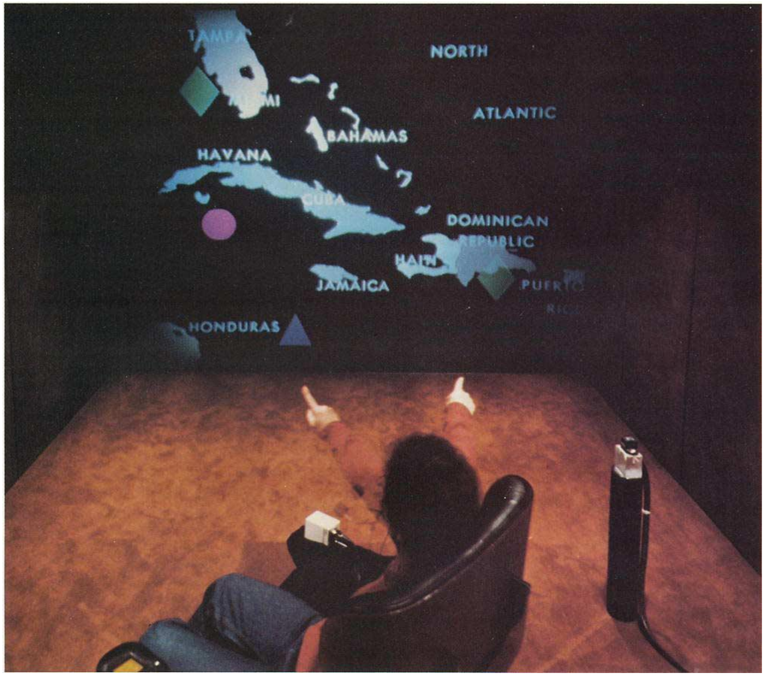


What are Multimodal Interfaces?

- Attempts to use human communication skills
- Provide user with multiple modalities
- May be simultaneous or not
- Fusion vs. Temporal Constraints
- Multiple styles of interaction

Early example

"Put That There" (Bolt 1980)…



Speech and gestures used simultaneously

Why Multimodal Interfaces?

- Today's interfaces fall far short of human capabilities
  - Higher bandwidth is possible
  - Different modalities excel at different tasks
  - Errors and disfluencies reduced
- Multimodal interfaces are more engaging
  - Users perceived multiple things at once
  - User do multiple things at once

Motivation: Why PUIs?

- Many reasons, including:
  - The "glorified typewriter" GUI model is too weak, too constraining, for the ways we will use computers in the future
  - One size doesn't fit all – diverse HCI requirements from small mobile devices to larger powerful embedded devices.
  - Transfer of natural, social skills – easy to learn
  - Simplicity: simple = natural, adaptive
  - Technology is coming: no longer deaf, dumb, and blind
  - To enable both *control* and *awareness*

How could we do this?

- Develop and integrate various relevant technologies, such as:

Speech recognition
Speech synthesis
Natural language processing
Vision (recognition and tracking)
Graphics, animation, visualization

Haptic I/O
Affective computing
Tangible interfaces
Sound recognition
Sound generation
User modeling
Conversational interfaces

Detecting gesture

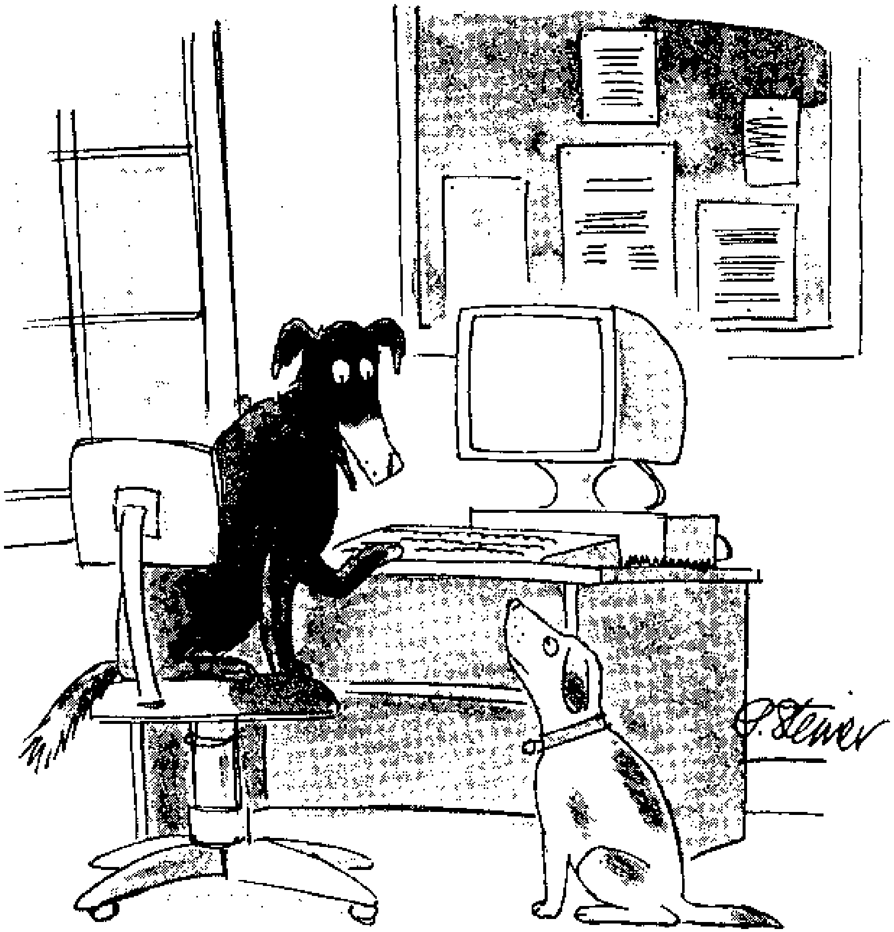
Being aware of the user

Natural navigation

There are many issues!

- What are the appropriate and most useful input/output modalities? (vision, speech, haptic, *taste*, *smell?)*
- Is the event-based model appropriate?
- What is a perceptual event?
- Is there a useful, reliable subset?
- Non-deterministic events
- Future progress (expanding the event set)
- Allocation of resources
- Multiple goal management
- Training, calibration
- Quality and control of sensors
- Environment restrictions
- Privacy

Issues (cont.)



"On the Internet, nobody knows you're a dog."

New Yorker, 5-Jul-1993, p. 61

Some PUI objections

- Arguments against intelligent, adaptive, agent-based, and anthropomorphic interfaces
- HCI should be characterized by:
  - Direct manipulation
  - <u>Predictable</u> interactions
  - Giving responsibility and a sense of accomplishment to users
- Won’t work – “AI hard”
  - Is 50% of HAL good enough?

Two major obstacles

- Technology (the easy one)
  - Lots of researchers worldwide
  - Increasing interest
  - Consistent progress

- The Marketplace (the hard one)
  - But there’s growing convergence: hw/sw advances, commercial interest in biometrics, accessibility, recognition technologies, virtual reality, entertainment….

but still... not quite there yet...

versus

Figure 4. The author wearing a variety of new devices. The glasses (built by Microoptical, Boston) contain a computer display nearly invisible to others. The jacket has a keyboard literally embroidered into the cloth. The lapel has a context sensor that classifies the user's surroundings. And, of course, there's a computer (not visible in this photograph).

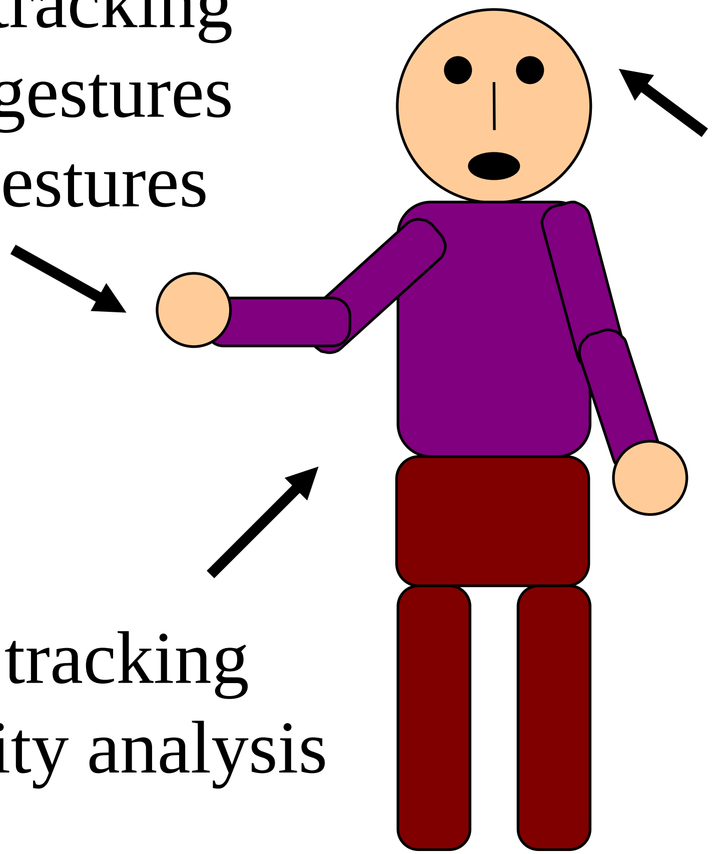
Vision Based Interfaces (VBI)

- Visual cues are important in communication!
- Useful visual cues
  - Presence
  - Location
  - Identity (and age, sex, nationality, etc.)
  - Facial expression
  - Body language
  - Attention (gaze direction)
  - Gestures for control and communication
  - Lip movement
  - Activity

VBI – using computer vision to perceive these cues

Elements of VBI

Hand tracking
Hand gestures
Arm gestures

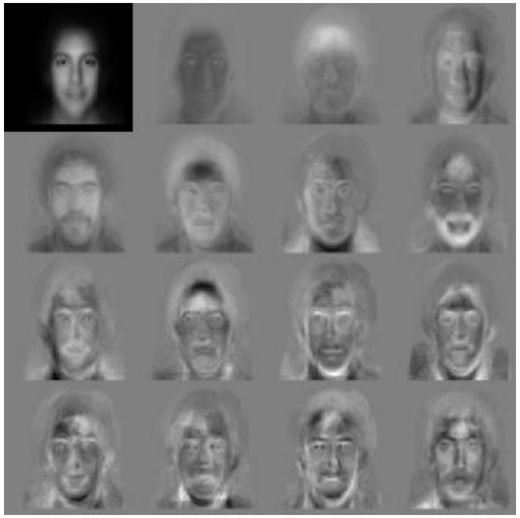


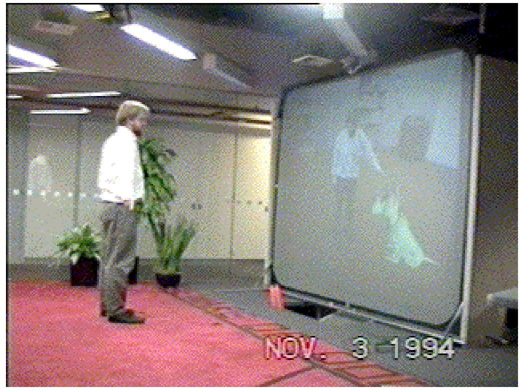
Head tracking
Gaze tracking
Lip reading
Face recognition
Facial expression

Body tracking
Activity analysis

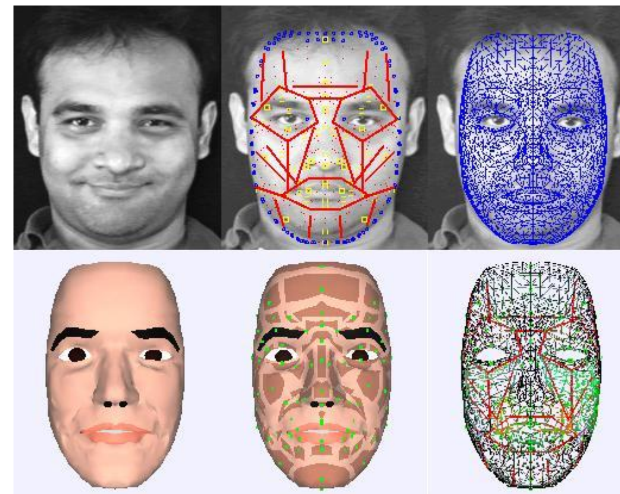
Some VBI application areas

- Accessibility, hands-free computing
- Game input
- Social interfaces
- Teleconferencing
- Improved speech recognition (speechreading)
- User-aware applications
- Intelligent environments
- Biometrics
- Movement analysis (medicine, sports)





MIT Media Lab 1990s



Perceptual Window

- Hand and mouse form the dominant stream
- Head is used as non-dominant stream
- Better than eye tracking
  - Fixation and saccades

Figure 2. The Perceptual Window uses small head motions as a second input stream to navigate within a document.

KidsRoom (Bobick et al 2000)

(a) A view of the KidsRoom showing the two projection screens and the movable bed.

(b) A child and mother rowing the boat together. Rowing was detected using story context and motion energy.

The technology

- Tracking faces
  - tracking the whole face, lips, gaze, or focus of attention
- Tracking bodies
  - person tracking
- Combining audio info with lip tracking info

Tracking of Human Faces

- A face provides different functions:
  - identification
  - perception of emotional expressions

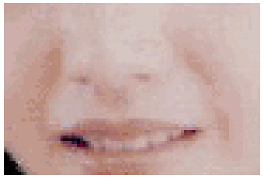
- Tracking of faces:
  - lip-reading
  - eye/gaze tracking
  - facial action analysis / synthesis

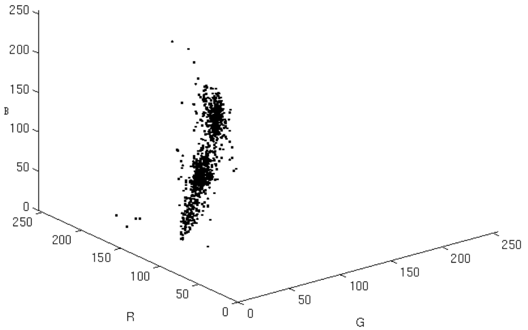
Color Based Face Tracking

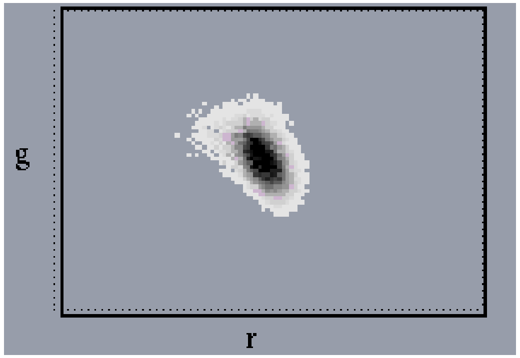
Human skin-colors:

- cluster in a small area of a color space
- skin-colors of different people mainly differ in intensity!
- variance can be reduced by color normalization
- distribution can be characterized by a Gaussian model

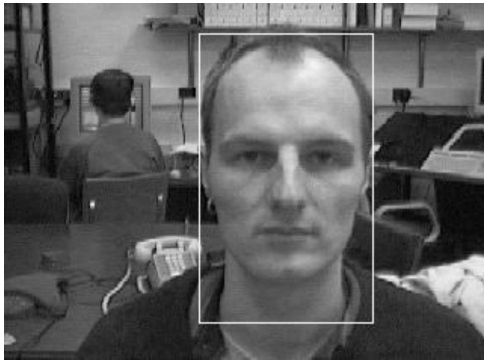
Chromatic colors: $r = \frac{R}{R+G+B}$ $g = \frac{G}{R+G+B}$

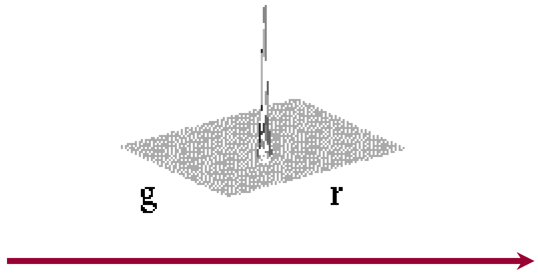






Color Model





Advantages:
- very fast
- orientation invariant
- stable object representation
- not person-dependent
- model parameters can be quickly adapted

Disadvantages:
- environment dependent
- (light-sources heavily affect color distribution)

Tracking Gaze and Focus of Attention

- In meetings:
  - to determine the addressee of a speech act
  - to track the participants attention
  - to analyze, who was in the center of focus
  - for meeting indexing / retrieval
- Interactive rooms
  - to guide the environments focus to the right application
  - to suppress unwanted responses
- Virtual collaborative workspaces (CSCW)
- Human-Robot Cooperation
- Cars (Driver monitoring)

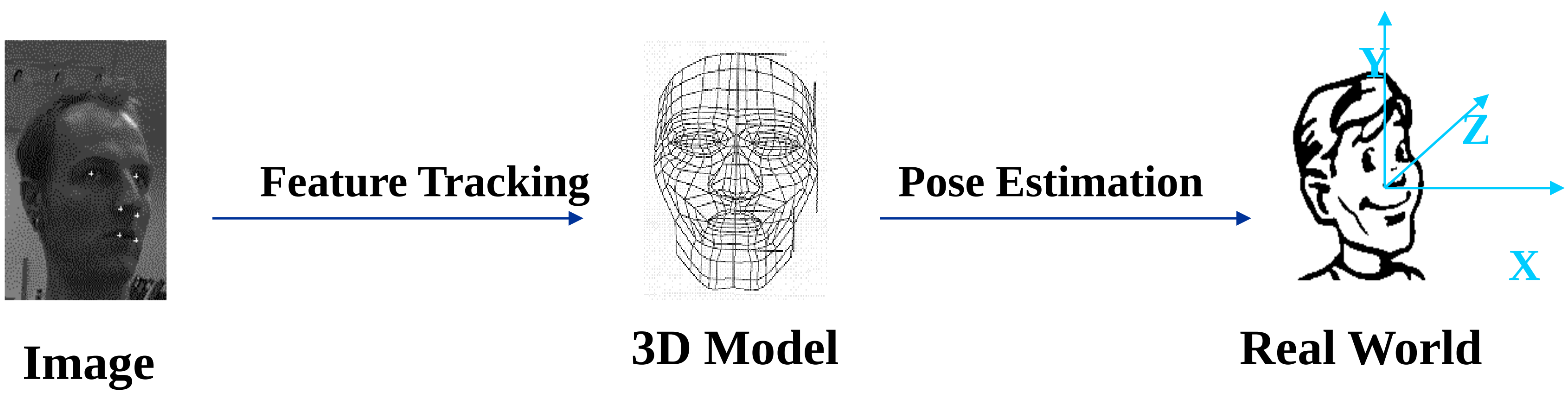
Head Pose Estimation

- Model-based approaches:
  - Locate and track a number of facial features
  - Compute head pose from 2D to 3D correspondences (Gee & Cipolla '94, Stiefelhagen et.al '96, Jebara & Pentland '97,Toyama '98)
- Example-based approaches:
  - estimate new pose with function approximator
  - use face database to encode images (Pentland et.al. '94)

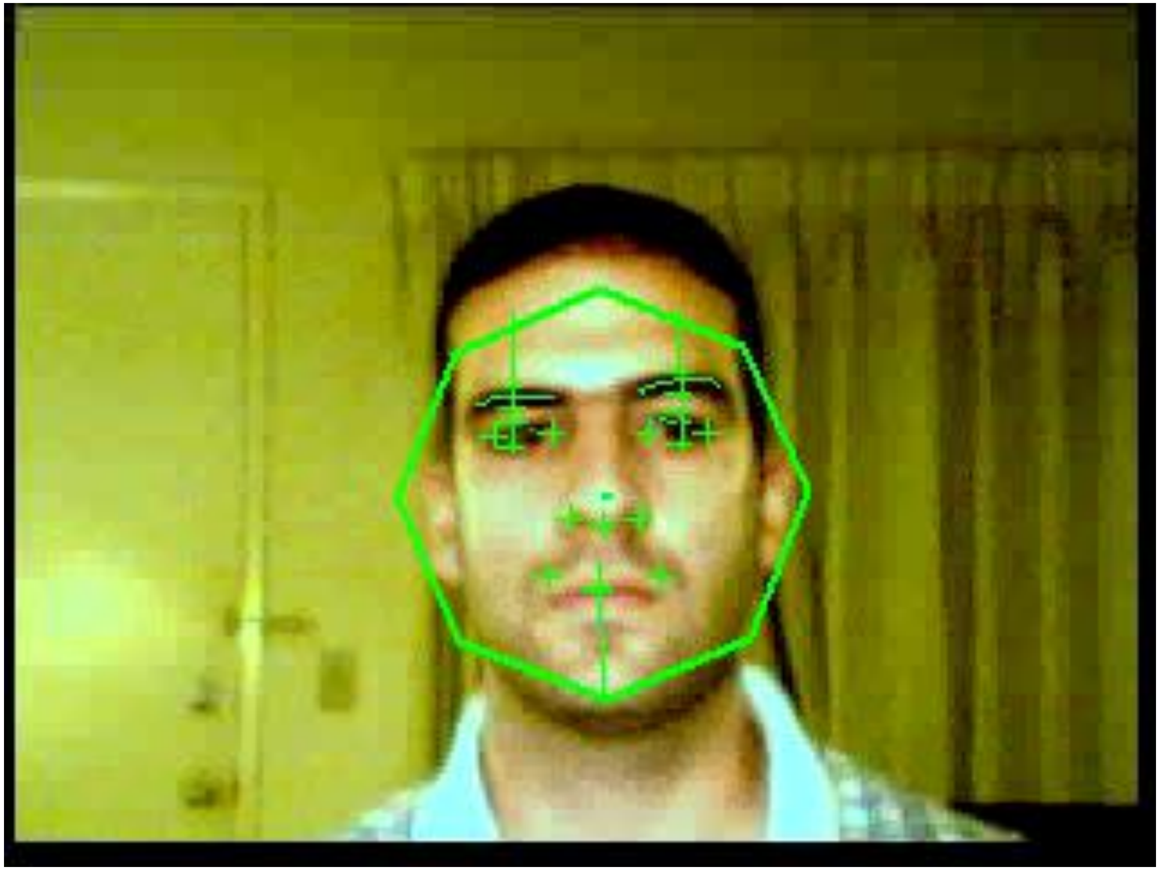
Model-based Head Pose estimation

- Find correspondences between points in a 3D model and points in the image

- Iteratively solve linear equation system to find pose parameters (r_x, r_y, r_z, t_x, t_y, t_z)



Head tracking demo

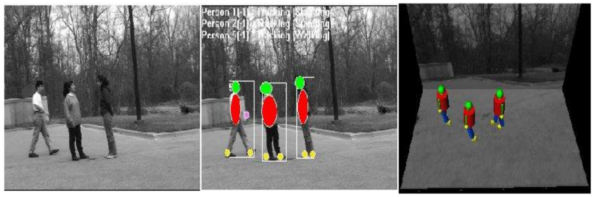


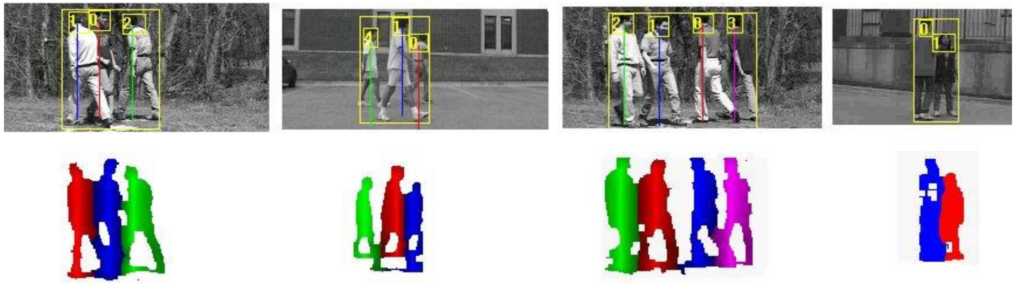
Person Tracking

Vision based localization of people/objects:

• Single Perspective:

•Multiple Perspective:





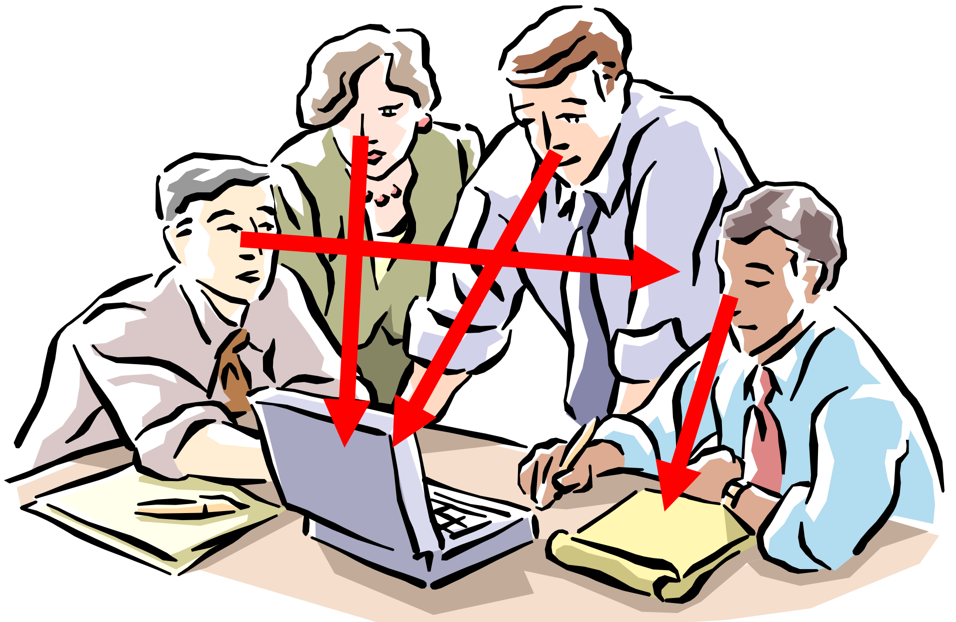
More examples

- Some applications from UCSB Four Eyes lab
- 4 I's: **I**maging, **I**nteraction, and **I**nnovative **I**nterfaces

- Research in computer vision and human-computer interaction
  - Vision based and multimodal interfaces
  - Augmented reality and virtual environments
  - Multimodal biometrics
  - Wearable and mobile computing
  - 3D graphics
  - ….

1. Coarse face direction

- Problem: Coarsely track multiple, possibly low-resolution face images in a scene
- Goal: Capture group behavior (attention); real-time
  - Estimate the "Focus of Intention" (attention + semantics)

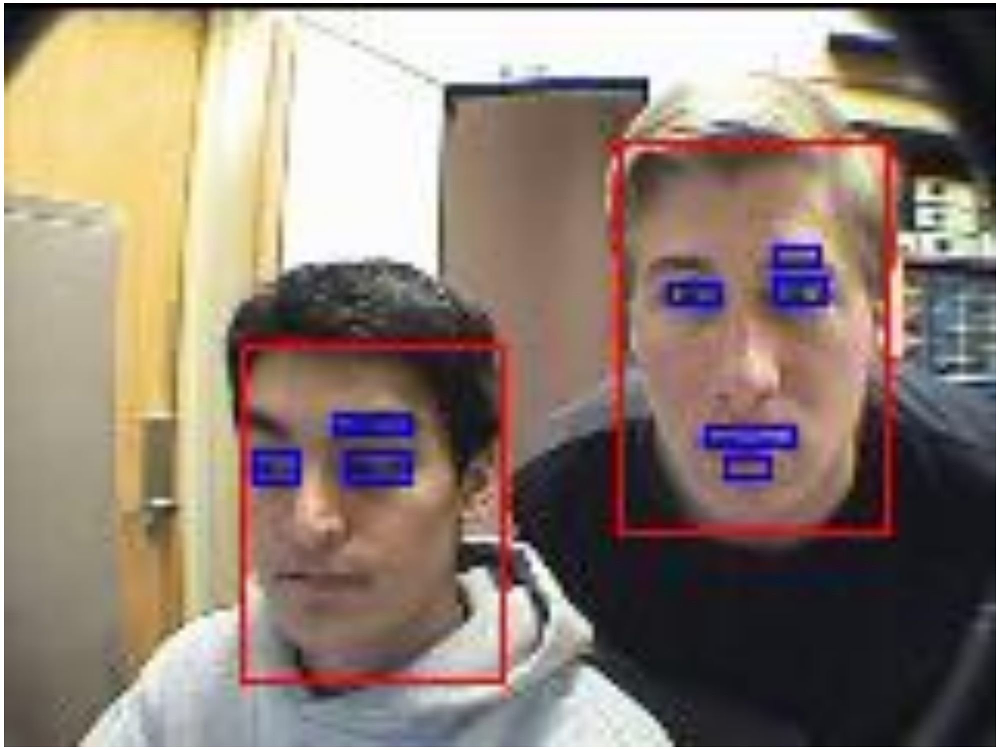


Action understanding
Meeting annotation
Audience feedback
Videoconferencing
Etc.

Coarse face direction (cont.)

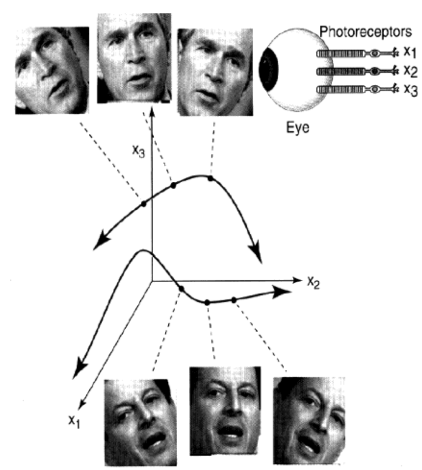
- Strategy:
  - Fast color-based skin tracking
  - Simple feature location
    - Non-skin areas
  - Simple statistics
  - Look for correlation with head direction (relative to camera)
  - $f\,(statistical\ measures) = direction$

Example results



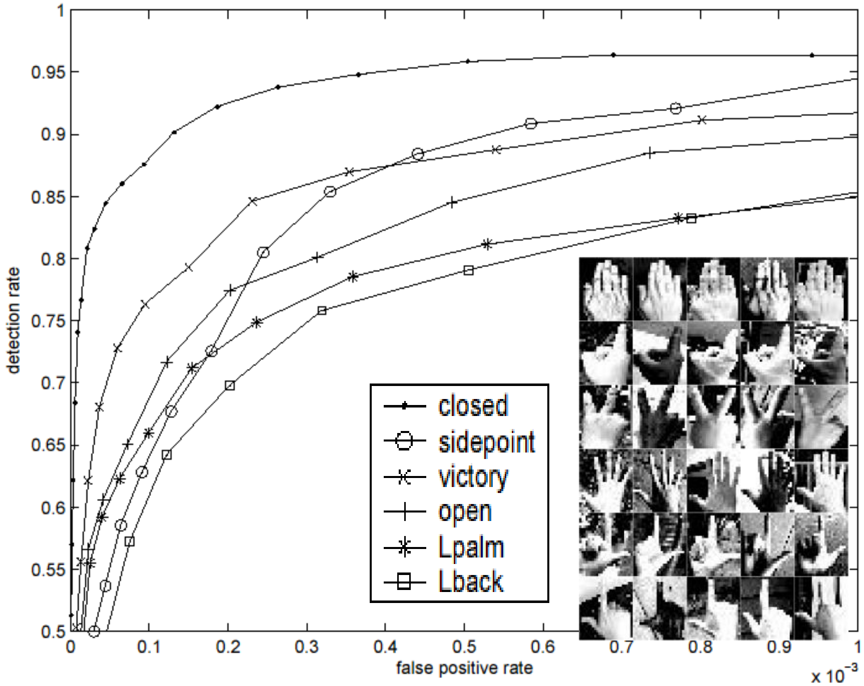
2. Facial expression analysis

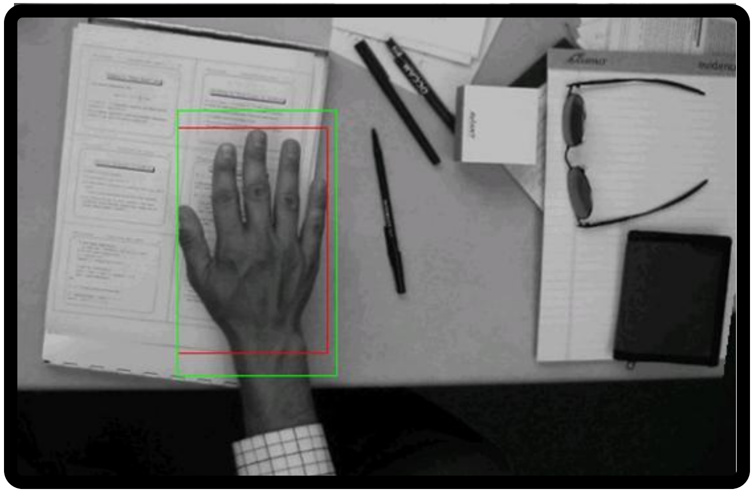
- Facial expression representation and visualization
- Use non-linear manifolds to represent dynamic facial expressions
- Intuition:
  - The images of all facial expressions by a person makes a smooth manifold in (high-dimensional) image space, with the "neutral" face as the central reference point.

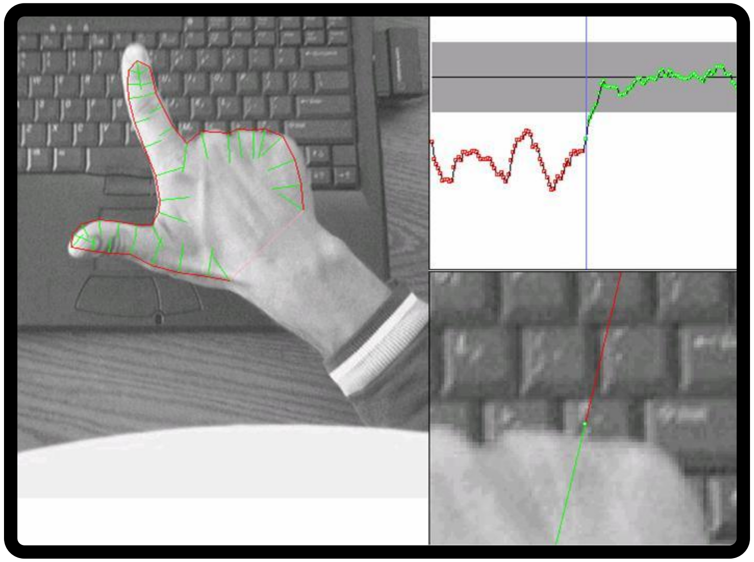


3. Hand detection, tracking, and recognition

Robust single-view detection







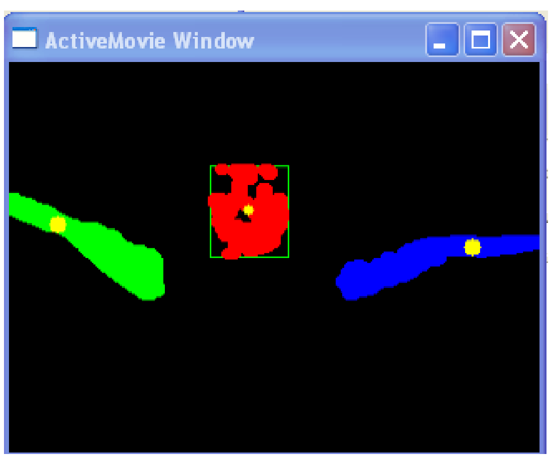
View-dependent posture recognition

Hand tracking demo

4. Recognizing body gestures and activity

- Current: Real-time tracking for
  - Interactive digital art applications
  - Autonomous aircraft on carrier flight deck

Restricted EM algorithm for skin classification
Head and hand/arm tracking



UBIQUITOUS COMPUTING

Summary

- Introduction to Ubiquitous Computing
- History of Ubiquitous Computing
- Challenges and Requirements

Introduction to Ubiquitous Computing

- What is
- Characteristics
- Goals

What is (1/3)

- the method of enhancing computing use by making many devices (services) available throughout the physical environment, but making them effectively invisible to the user (Mark Weiser)

Computing Everywhere

Ubiquitous means:

- present everywhere
- simultaneously encountered in numerous different instances
- computers become a useful but invisible force, assisting the user in meeting his needs without getting lost in the way

What is (2/3)

- tries to construct a universal computing environment (UCE) that conceals (hides):
  - computing instruments
  - devices
  - resources
  - technology

  from applications or customers

- invisible to users

What is (3/3)

- *computing everywhere*

- many embedded, wearable, handheld devices communicate transparently to provide different services to the users

- devices mostly have low power and short-range wireless communication capabilities

- devices utilize multiple on-board sensors to gather information about surrounding environments

Characteristics of Ubicomp Applications

- context-awareness (also a key-characteristic of perceptual interfaces)

- improvised and dynamic interaction

- interactions among applications are based on specific context

Goals

- the promise of ubiquitous computing:
  a life in which our tasks are powerfully,
  though invisibly, assisted by computers

Summary

- Introduction to Ubiquitous Computing
- History of Ubiquitous Computing
- Challenges and Requirements

History of Ubiquitous Computing

- History
- Mark Weiser
- Experiments

History

- Active Badge
  - Andy Hopper

- Xerox PARC 1991-2000
  - Mark Weiser (until, sadly,April 1999)

- Calm Technology

Mark Weiser:
the father of ubiquitous computing

- researcher in the Computer Science Lab at Xerox's PARC (Palo Alto Research Center)

- first articulated the idea of ubiquitous computing in 1988

- has called UC "…highest ideal is to make a computer so embedded, so fitting, so natural, that we use it without even thinking about it."

Ubiquitous Computing

- During one of his talks, Weiser outlined a set of principles describing ubiquitous computing:
  - The purpose of a computer is to help you do something else.
  - The best computer is a quiet, invisible servant.
  - The more you can do by intuition the smarter you are; the computer should extend your *unconscious*.
  - Technology should create calm.
- In Designing Calm Technology, Weiser and John Seeley Brown describe *calm technology* as "that which informs but doesn't demand our focus or attention".

Xerox PARC 1991-2000

- PARC = Palo Alto Research Center
  - 41 people immersed in ubiquitous computing environment
- virtual UCE with several interconnected devices such as notepads, blackboards and electronic scrap papers
- difference from a standard PC:

  people using these devices do not perceive them as computers anymore and can therefore focus on the actual tasks

Active Badge 1988

- smart telephone networks

- problem of automatically routing telephone calls to the correct place in a building

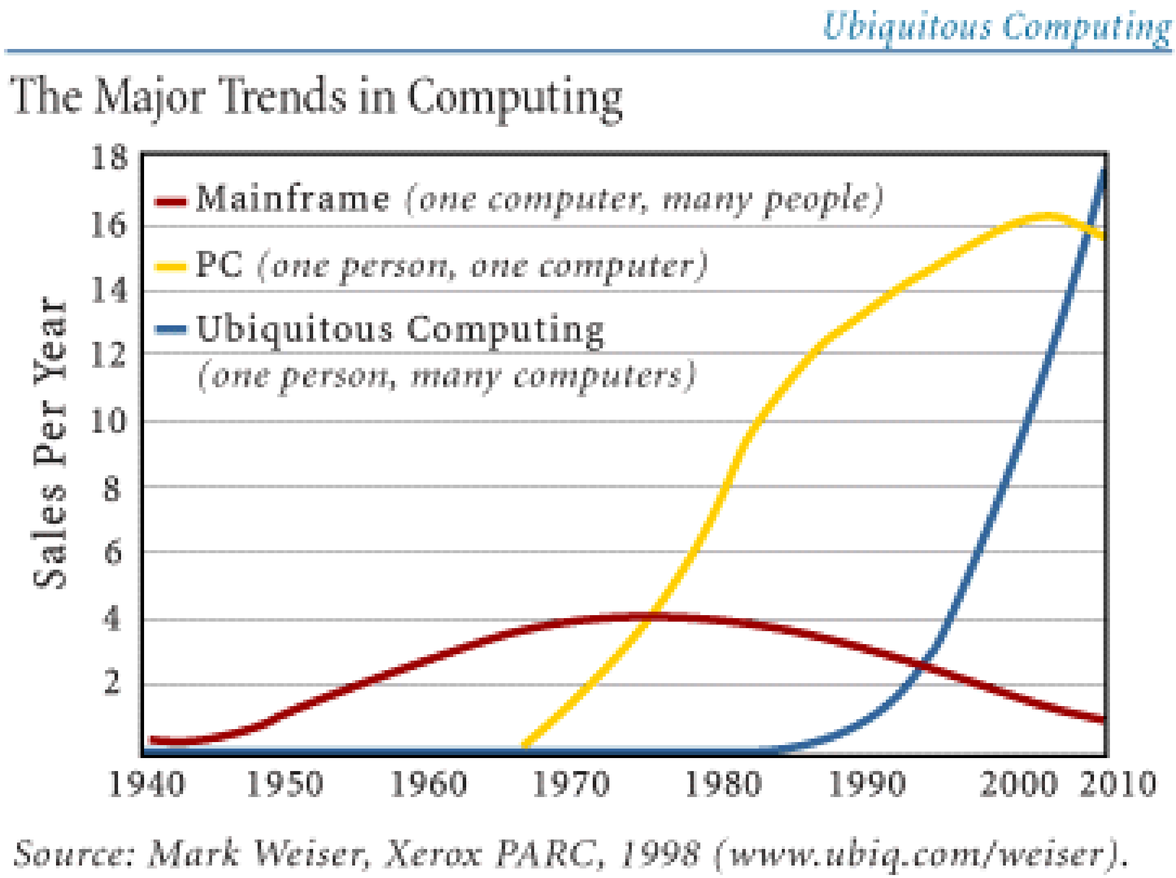
- opened up a whole new area of research and helped to realize a new opportunity for context based computing

Calm Technology (1/3)

The Major Trends in Computing

Mainframe many people share a computer
Personal Computer one computer, one person
Internet - Widespread Distributed Computing . . . transition to . . .
Ubiquitous Computing many computers share each of us

Calm Technology (2/3)



Calm Technology (3/3)

Today Internet is carrying us through an era of widespread *distributed computing* towards the relationship of *ubiquitous computing*, characterized by deeply embedding computation in the world.

Ubiquitous computing will require a new approach to fitting technology to our life, an approach called "calm technology".

Experiments

- Tabs
- Pads
- Boards

1988 – 1994 at PARC Xerox

- SAAMPad (Software Architecture Analysis Method Pad)
- The Conference Assistant

Experiment at PARC - TAB

TAB

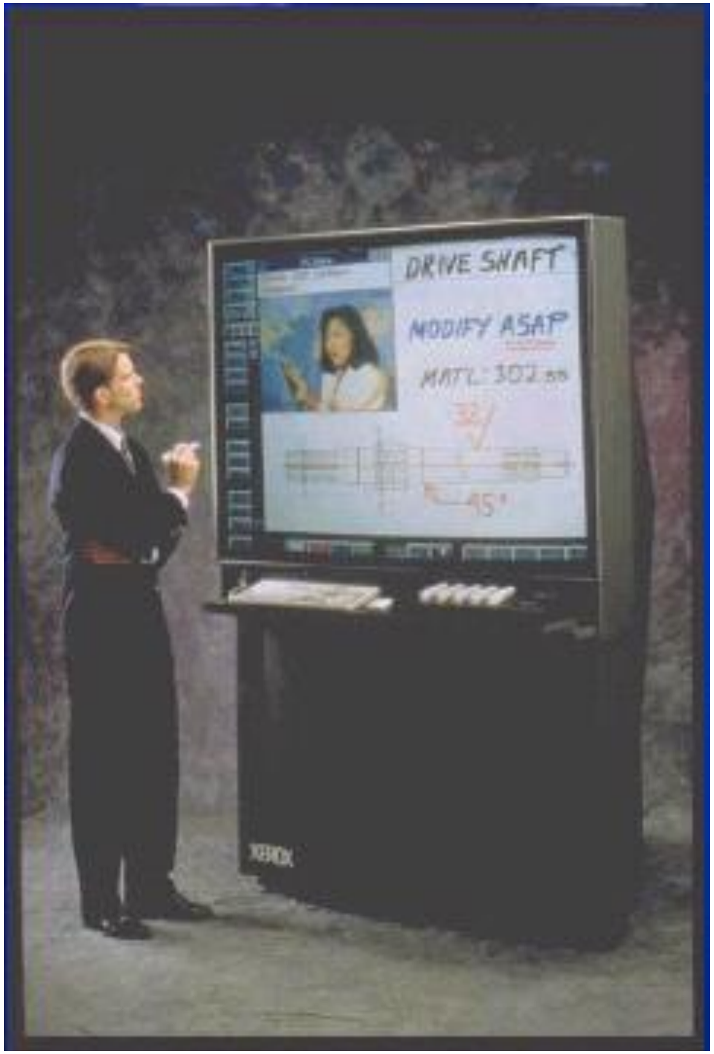
- prototype handheld computer
- was 2x3x0.5", had a 2 week battery life on rechargeable batteries, and weighed 7 oz
- used a Phillips 8051 processor with 128k NVRAM
- featured an external I^2C external bus, a custom resistive touch screen, and a 128x64 mono display
- included an infrared base station in the ceiling for LAN connectivity

The Tab project is considered by many to be the most significant of the three prototyping efforts

Experiment at PARC - PAD

Experiment at PARC – BOARD

Liveboard



Summary

- Introduction to Ubiquitous Computing
- History of Ubiquitous Computing
- Challenges and Requirements

Challenges and Requirements

- Hardware
- Applications
- User Interfaces
- Networking
- Mobility
- Scalability
- Reliability
- Interoperability
- Resource Discovery
- Privacy and Security

Nanotechnology (1/2)

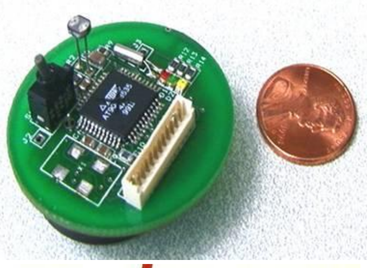
The trend toward miniaturization of computer components down to an atomic scale is known as nanotechnology

Nanotechnology (2/2)

- Mobile data technology
  - GSM, GPRS, UMTS, CDMA, WAP, Imode
- Wireless data technology
  - Bluetooth, 802.11b
- Internet data technology
  - IP over optical, Broadband
- Content services
  - Web & WAP
- Applications
  - Multimedia, Internet messaging

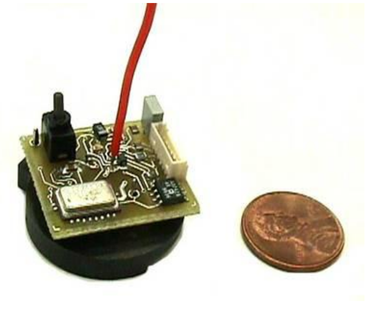
Smaller sensors

weC
codesigned by
James McClurkin



RF 916.5 MHz OOK
10kbps 20 meter range
Sensors: light, temperature

Mini Mote
codesigned by
Christina Adela



RF 916.5MHz OOK
10kbps 20 meter range
Sensors: temperature

New Technologies:
Light Emitting Polymers

- Plastic displays (~ 1 mm thick)
- Applications are emerging (e.g., curved or flexible displays)

Applications

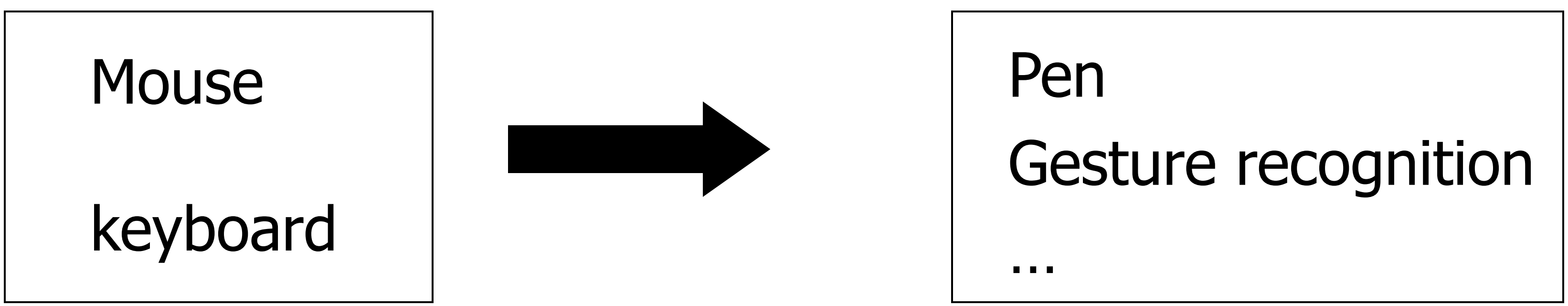
- main motivation of ubiquitous computing (Weiser 1993)

- need to have an awareness of their context:

  a combination of several factors, including the current location, the current user or if there are any other Ubicomp devices present in the near surroundings

Users Interface

The multitude of different Ubicomp devices with their different sizes of displays and interaction capabilities represents another challenge



Networking

Another key driver for the final transition will be the use of short-range wireless as well as traditional wired technologies

Mobility

Mobility is made possible through wireless communication technologies

Problem of disconnectivity!!!

This behaviour is an inherent property of the ubicomp concept and it should not be treated as a failure

Scalability

In a ubiquitous computing environment where possibly thousands and thousands of devices are part of scalability of the whole system is a key requirement

All the devices are autonomous and must be able to operate independently a decentralized management will most likely be most suitable

Reliability

Thus the reliability of ubiquitous services and devices is a crucial requirement

In order to construct reliable systems self-monitoring, self-regulating and self-healing features like they are found in biology might be a solution

Interoperability

This will probably be one of the major factors for the success or failure of the Ubicomp vision

This diversity will make it impossible that there is only one agreed standard

Resource Discovery

The ability of devices to describe their behaviour to the network is a key requirement.

On the other hand, it can not be assumed that devices in a ubiquitous environment have prior knowledge of the capabilites of other occupants.

Privacy and Security

In a fully networked world with ubiquitous, sensor-equipped devices several privacy and security issues arise

- the people in this environment will be worried about their privacy since there is the potential of total monitoring

- must be understandable by the user and it must be modelled into the system architecture

Examples

- Ambient Devices
  - Ambient orb
  - Ambient dashboard
  - Ambient weather beacon

Mobile Interface Design Guidelines

- iPhone design guidelines:
  - http://developer.apple.com/iphone/library/documentation/userexperience/conceptual/mobilehig/Introduction/Introduction.html
- Small Surfaces
  - http://www.smallsurfaces.com/
- Nokia design guidelines:
  - http://wiki.forum.nokia.com/index.php/Guidelines_for_Mobile_Interface_Design
- Cxpartners Mobile interface design:
  - http://www.cxpartners.co.uk/services/mobile_interface_design

Presentations next week

- The presentation schedule are posted on the web page (at the Schedule/Lecture Notes section)
- The presentations will be about 5-10 minutes, describing what you have done briefly.
- Any group member may make the presentation. It is OK if all the group members are not present during the presentation.

Project Reports

- For the final phase of your project, you are going to write a project report containing:
  - A description of the prototype or completed interface proposed in phase 1.
    - Textual description, snapshots, walkthrough of the system
  - Which design guidelines did you employ?
    - Visibility, mapping, user feedback, error-handling, etc.
  - Evaluation results
    - Which evaluation strategy did you use?
    - How many users?
    - What were the results?
    - Did you re-design your interface based on feedback from user evaluations?
- Final project reports are due on the last day of finals (send your reports by e-mail).
- The report should also contain which group member did which task.