Multimodal User Interfaces 2019

[2] Multimodal Interaction Introduction

February 26th, 2019

Denis Lalanne

Outline

- **2.1.** Definitions and goals
- **2.2.** Uni-Modal systems
- **2.3.** History of HCI (towards multomodality)
- **2.4.** Designing Multimodal interfaces, learnings
- **2.5.** The 10 Myths of multimodal integration
- **2.6.** Perspectives
- **2.7.** CASE & CARE quick introduction
- **2.8.** What you shoud know and what’s next

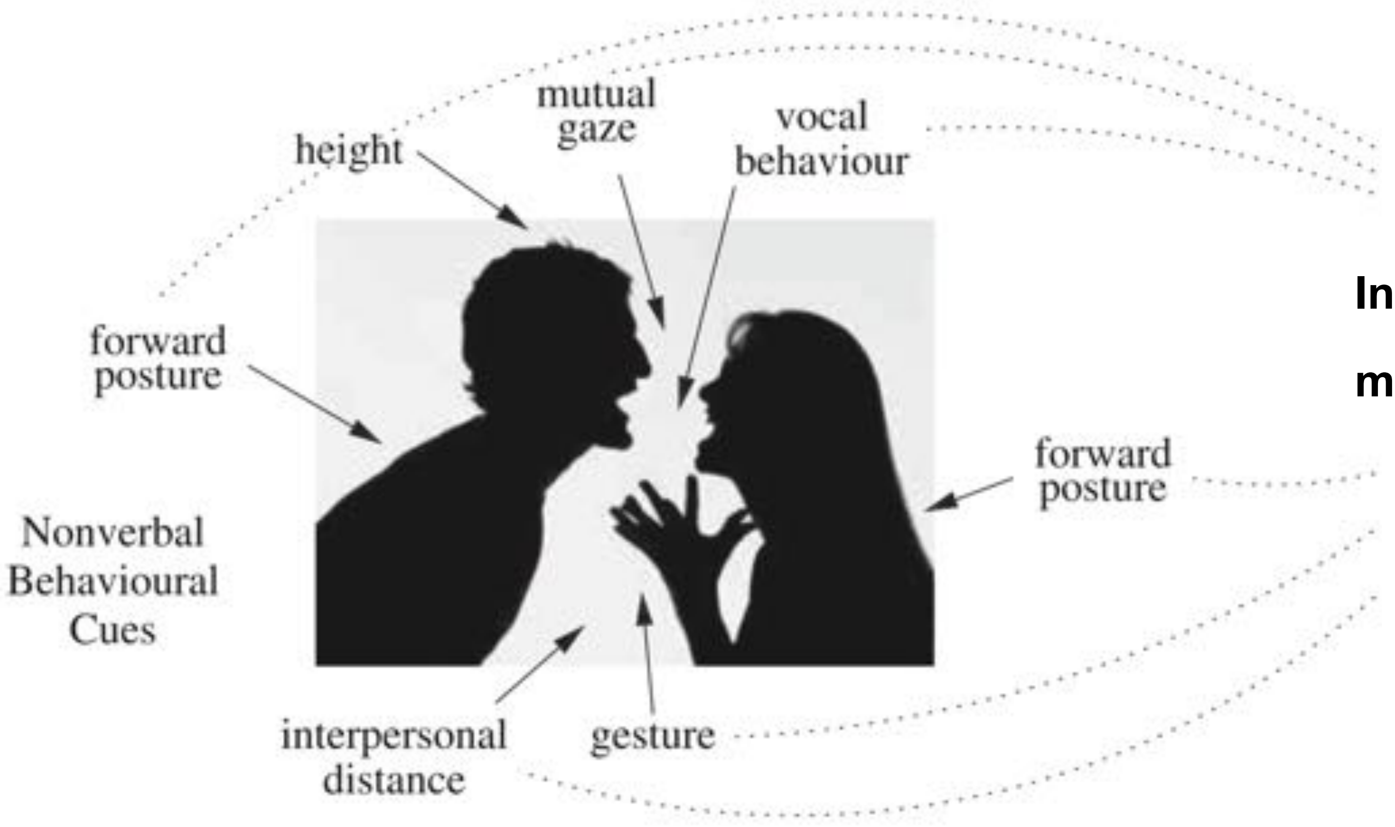
Multimodal Interfaces

[2.1] Definitions and goals

Denis Lalanne

Multimodal Interaction (MMI)

Combine in a natural manner complementary modalities



Interaction

modalities

Goals of multimodal interfaces

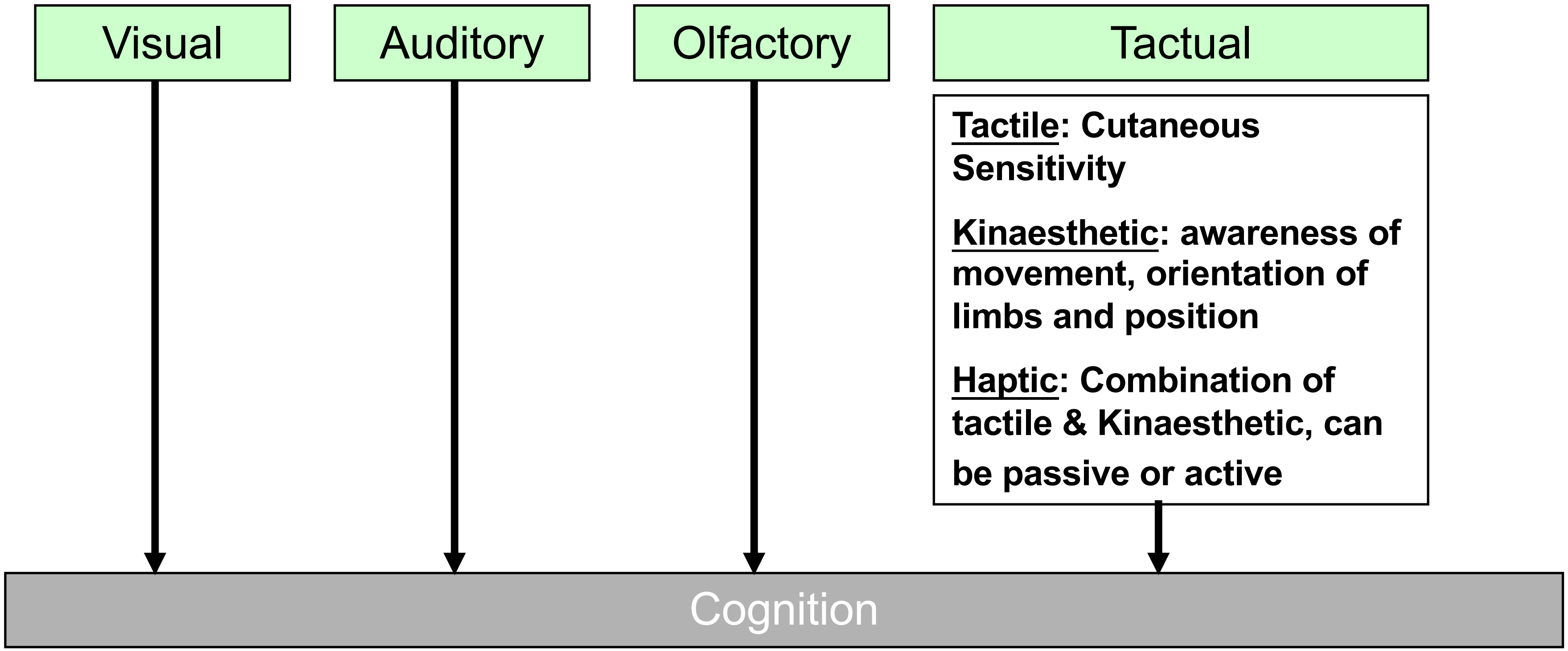
- Extend perceptual and cognitive human capabilities (AV)
- Integrate computational skills of computers in the real world (AR)

- Multimodality
  - A domain in constant evolution
  - Novel interaction techniques are appearing

Terminology (confusion of terms)

- Media
  - Material (signal on a channel)
  - The support of communication
  - Formats for presenting information (audio, video, graphics, etc.).
- Modality
  - A channel of communication e.g. (voice, gestures, facial expressions)
  - Sensorial (audition, vision, touch, smell, taste.)
- Multimedia:
  - The use of different media to convey information; e.g. text together with audio, graphics and animation
- Multimedia system:
  - Transport signals of different kinds
- Multimodal system:
  - Interpret signs belonging to various sensory and communication systems

Human Input (Perceptual) Modalities



Visual

Auditory

Olfactory

Tactual

Tactile: Cutaneous Sensitivity

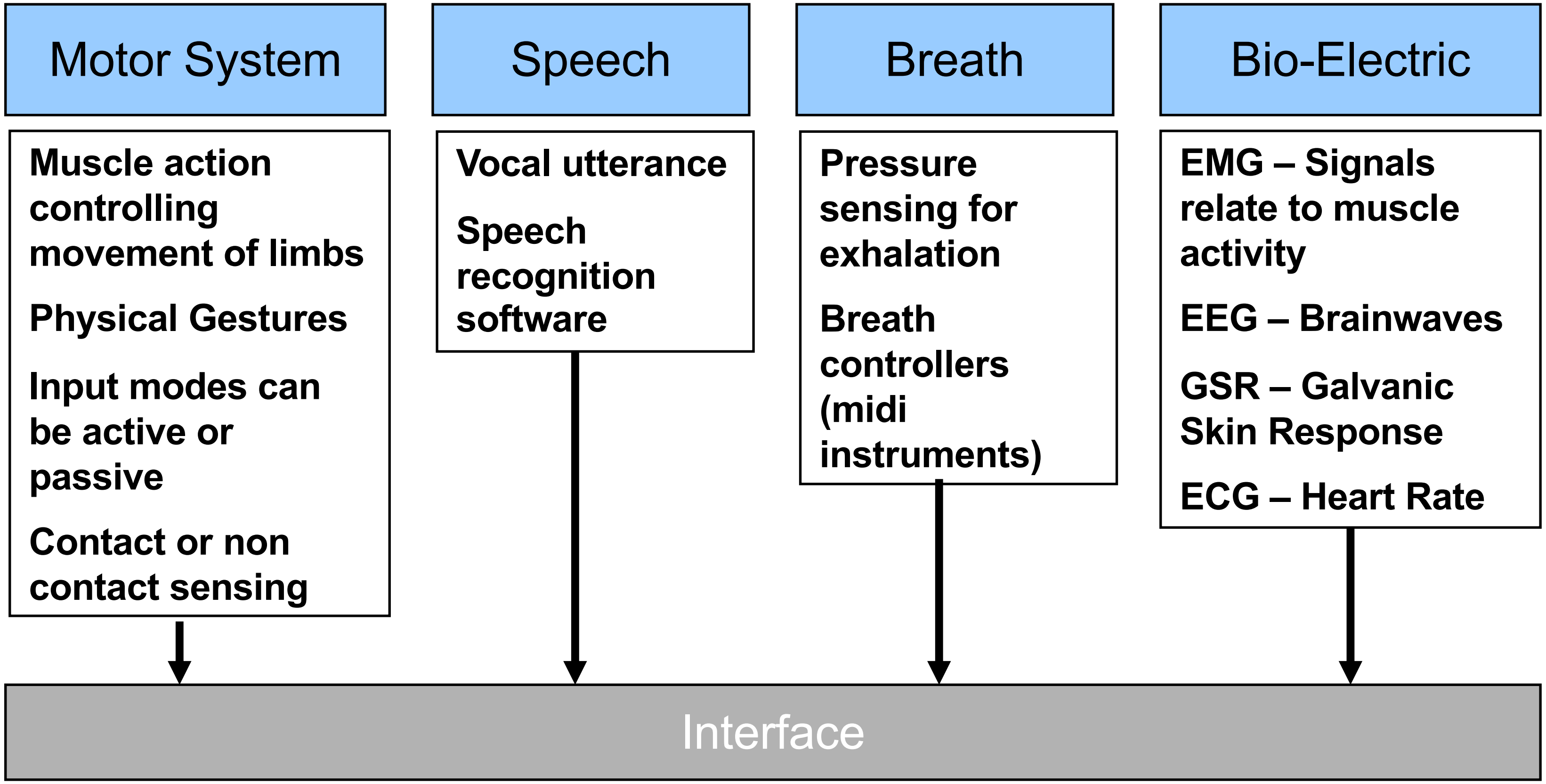
Kinaesthetic: awareness of movement, orientation of limbs and position

Haptic: Combination of tactile & Kinaesthetic, can be passive or active

Cognition

Multimodal interaction and multisensory feedback can exert significant cognitive demands on users. Attention is selective and has limited capacity. Short term memory also has limitations. Consider human ability when designing the interface, essential for control and learning affordances

Human Output Modalities



Motor System

Muscle action controlling movement of limbs

Physical Gestures

Input modes can be active or passive

Contact or non contact sensing

Speech

Vocal utterance

Speech recognition software

Breath

Pressure sensing for exhalation

Breath controllers (midi instruments)

Bio-Electric

EMG – Signals relate to muscle activity

EEG – Brainwaves

GSR – Galvanic Skin Response

ECG – Heart Rate

Interface

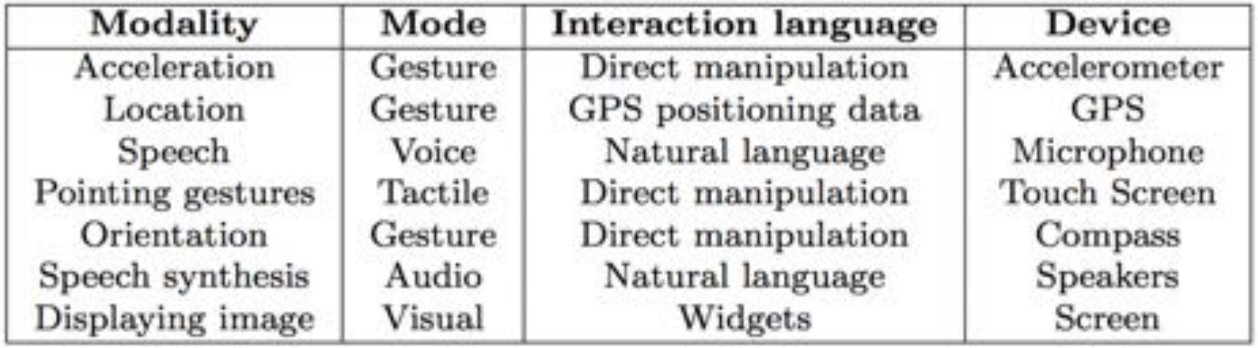
EMG = electromyography, EEG = Electroencephalography
GSR = Galvanic Skin Response, ECG = Electrocardiography
Warning: this list of modalities is not complete!

Basically, Sound is a modality… Right?

- Well, the real story is a bit more complicated than that.
- Sound can be processed at different levels of abstraction
  - Noise level
  - Pitch
  - Rhythm
  - Phonemes
  - Voice recognition (word commands)
  - Voice recognition (free speech)
- Each of these levels can be a modality
- *Every channel of communication or modality can be decomposed in different levels, resulting in different modalities*

(Stricter) Definition of a Modality

- In human-computer interaction, a modality is the general class of:
  - A device through which the computer receive inputs from humans
  - a path of communication between the human and the computer.
- A modality m is defined by m= <d, r> | <m, r> (Nigay and Coutaz)
  - “m” an interaction modality
  - “d” denotes the physical I/O device,
  - “r” an interaction language (representational system)
  - (m=<m,r>) means that m can be devised in “sub-modalities”



Modality	Mode	Interaction language	Device
Acceleration	Gesture	Direct manipulation	Accelerometer
Location	Gesture	GPS positioning data	GPS
Speech	Voice	Natural language	Microphone
Pointing gestures	Tactile	Direct manipulation	Touch Screen
Orientation	Gesture	Direct manipulation	Compass
Speech synthesis	Audio	Natural language	Speakers
Displaying image	Visual	Widgets	Screen

Table 1: Some examples of interaction modalities

Elouali, Nadia, et al. "Multimodal interaction: a survey from model driven engineering and mobile perspectives." Journal on Multimodal User Interfaces 7.4 (2013): 351-370.

Modalities: a Small Glimpse

- Modalities can be classified along many axes
  - *Input vs. output modalities*
  - *Atomic vs. continuous modalities*
  - *Raw vs. interpreted modalities*
  - *Active vs. passive modalities (conscious vs unconscious)*
- A complete review of all existing modalities would take too much time
- Study of a selection of modalities
  - Speech
  - Pen/stylus
  - Gaze

Multimodal Interfaces

[2.2] Uni-modal Natural Recognition-Based Interfaces

Denis Lalanne

Uni-modal Natural Recognition-Based Interfaces

- Speech- the “Portable Power Tool”

- Pen- the “Compact Multifunctional Chameleon”

- Human Gaze- the “Active Index of Interest”

The following slides are inspired from S. Oviatt

Speech- the "Portable Power Tool"

- Speedy input, strongly interactive, high-bandwidth
- Preferred for many communication tasks
- Wide-ranging expressivity, including description of:
  - *past & future events*
  - *out-of-view objects*
- Natural portability, eyes & hands free

When Speaking To Computers Is Useful

- During competing tasks, when the user's *hands and/or eyes are busy*
- During *mobile tasks*, or when only a limited keyboard and/or screen is available (telephony, PDAs)
- When the user is *temporarily or permanently "disabled"* (visually impaired)
- When *oral pronunciation is the subject matter* of computer use (reading, foreign language training)
- When the *users prefer speech* input for speed, ease, transparency, tight interactivity, or high-bandwidth features (during exchanges designed to persuade the listener, or to convey attitude, energy, personality, gender...)

Common Speech Applications

- Telephony (automated operator services, stock trading & service transactions- e.g., SpeechActs & Nuance systems)
- Dictation (word processing, medical records, etc.)
- Language training
- General information access for services
- Applications for disabled users (visually-impaired)
- Home for entertainment

Speech Interface Challenges and Design Techniques

- Designing speech interfaces *appropriately for the spoken modality* (i.e., NOT as a plug-in feature for an existing GUI!)
- Processing *spontaneous interactive* human speech in *realistic contexts*
- Managing *recognition errors*
- Designing *fluid prompts* & *clear confirmations* (especially for telephony!)
- *Giving users control* to avoid tediously lengthy interactions
- Designing for “mixed-initiative” dialogue (i.e. flexible interaction strategy)
- Supporting speech interfaces for *diverse users’* signal characteristics (accent, age, etc.)

Pen- the "Compact Multifunctional Chameleon"

- Direct, constrained and precise spatial input
- Easy portability
- Multifunctionality *(text, digits, pointing, gestural marks, symbols, graphics, sketching & art, signatures, direct manipulation, etc.)*
- Visual feedback, permanent record, easy editing
- Preferred for spatial & graphic tasks, selection of objects, numeric & symbolic data, & signatures

Common Pen Applications

- Design applications *(flow charts, sketching, CAD),* drawing
- Document annotation
- Personal scheduling & information management *(e.g., Palm Pilots)*
- Mobile telephony (starting to use digit & character recognition, as well as selection)

- **Challenge**:
  - Providing *clear prompting and confirmation feedback* (i.e., without disrupting user’s task by echoing letter-by-letter recognition)

Human Gaze- the “Active Index of Interest”

- Promising for *passive* control involving *brief time intervals*
- Promising as early indicator for *monitoring user’s interest*
- Fast & highly sensitive, but often *difficult to interpret*
- Not under full conscious control- *intentional* looking mixed with periods of *blank staring*
- Easiest for some populations *(young children, neurologically impaired)*
- Good for hands busy tasks
- Still exploratory use in HCI tasks, although technology maturing rapidly

What Researchers Have Learned from Pupils: Myths about Gaze Patterns

- Myth #1: Users eyes *stop* to look at things
- Myth #2: Users look at things *intentionally*
- Myth #3: What users are looking at is an indication of *what they're thinking*
- Myth #4: The eyes and hands manipulate things *simultaneously*
- Myth #5: Eye trackers track eye movements *reliably*

Applications & Conclusion

Application

- Computer, communications, & self-care applications for severely-impaired users (e.g., quadriplegics, motor- impaired groups)
- Rendering visual details in VR environments (for computational efficiency)

Conclusion

- Eye gaze not suitable as an active pointing device, because it:
  - Is *hyperactive* & never stops moving to fixate precisely
  - *Fatigues easily* & isn't suitable for extended computer use
  - *Isn't always under conscious control* & often stares blankly

Therefore, gaze isn't a good mouse replacement!

It's more suitable for "passive" (non-command) multimodal interface designs

Multimodal Interfaces

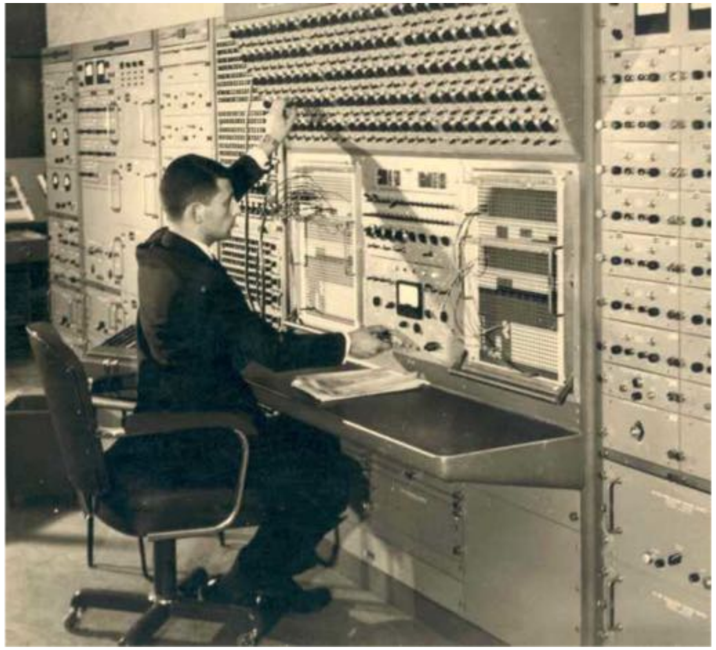
[2.3] History of HCI (towards multimodality)

Denis Lalanne

UNI
FR

Human-Machine Interaction

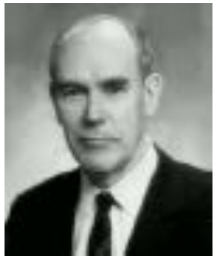
- "Bridging the gap" between the human and the machine



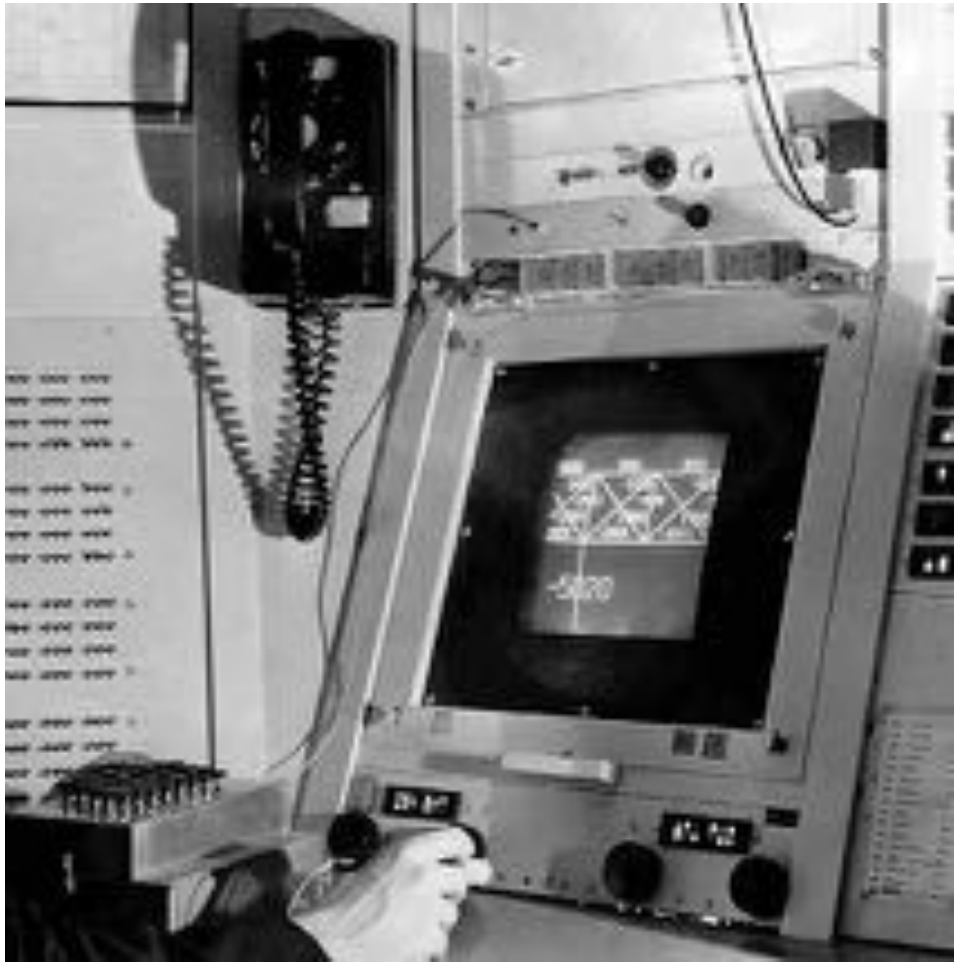
Analog computer of the late 50's

***Minority Report* (2002)**

Once upon a time HCI…

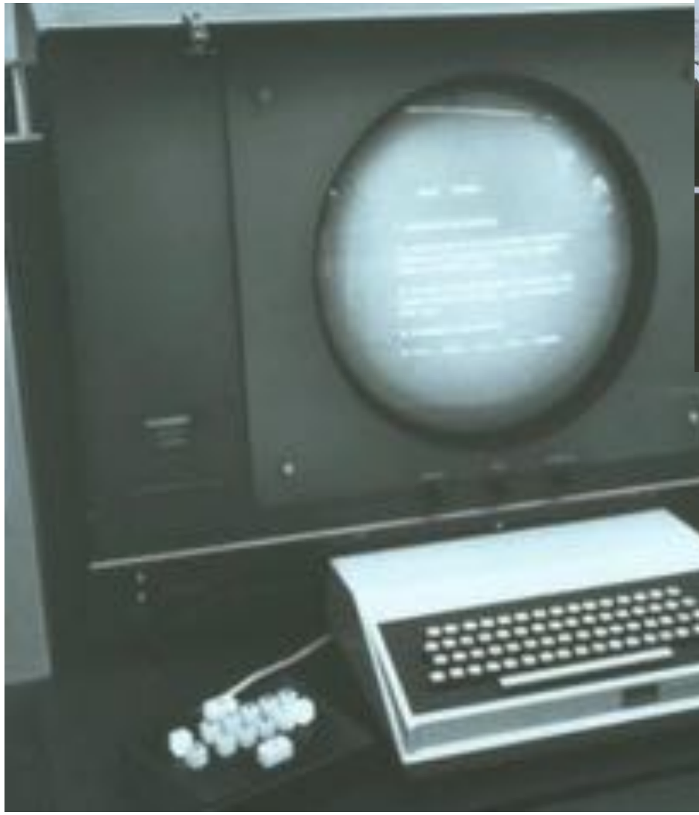


- Ivan Sutherland 1963 (PhD thesis MIT)
- Sketchpad
  - Drawing tool
  - Optical pen and buttons
  - Direct manipulation
  - Icons
  - Zoom
  - Copy/Paste



Once upon a time HCI…

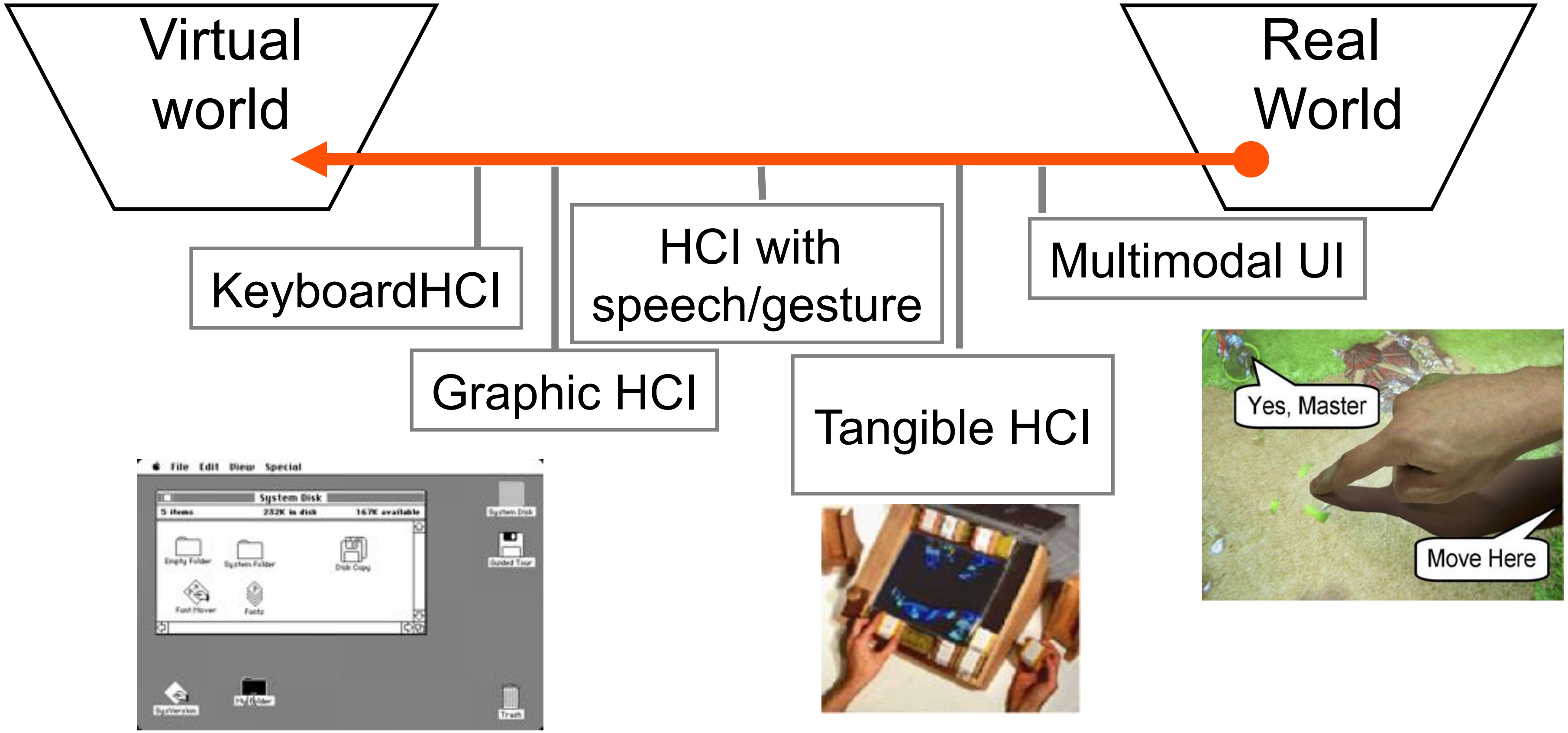
- Douglas Engelbart
- 1968 NSL oN Line System
- Augment/NSL
  - ➢ Text edition
  - ➢ Video conference
  - ➢ Two dimensional screen
  - ➢ Device on knee
  - ➢ Mouse
  - ➢ hypertext



=> 1984

Evolution of Interactions (1)

Augmented Virtuality (AV)



HCI using voice / speech (FUI_04)

- Put that there [1979]

- Direction Assistance [1988]

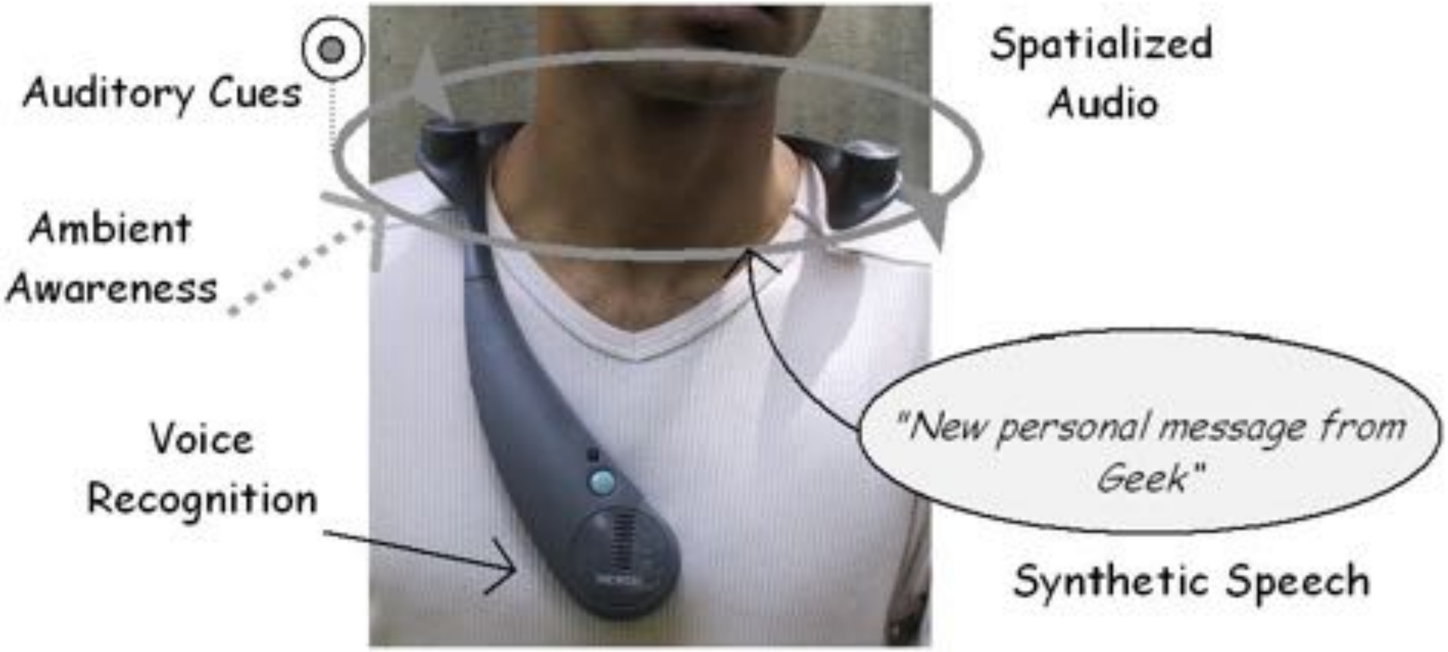
- Hyperspeech [1991]

- Talkback [2002]

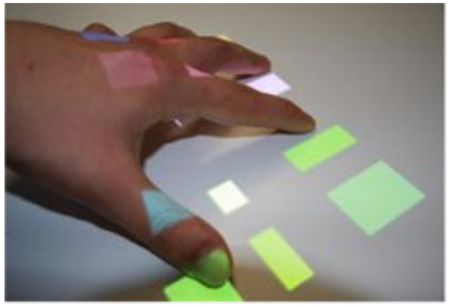
- Conversation Finder [2003]

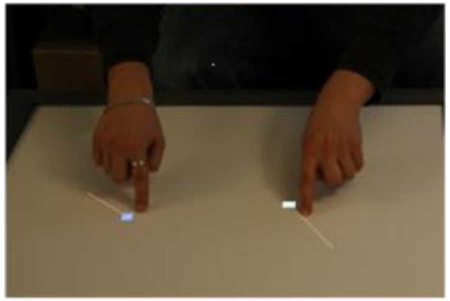
Sound interfaces

Nomadic Radio



2D gestures & interactive tables



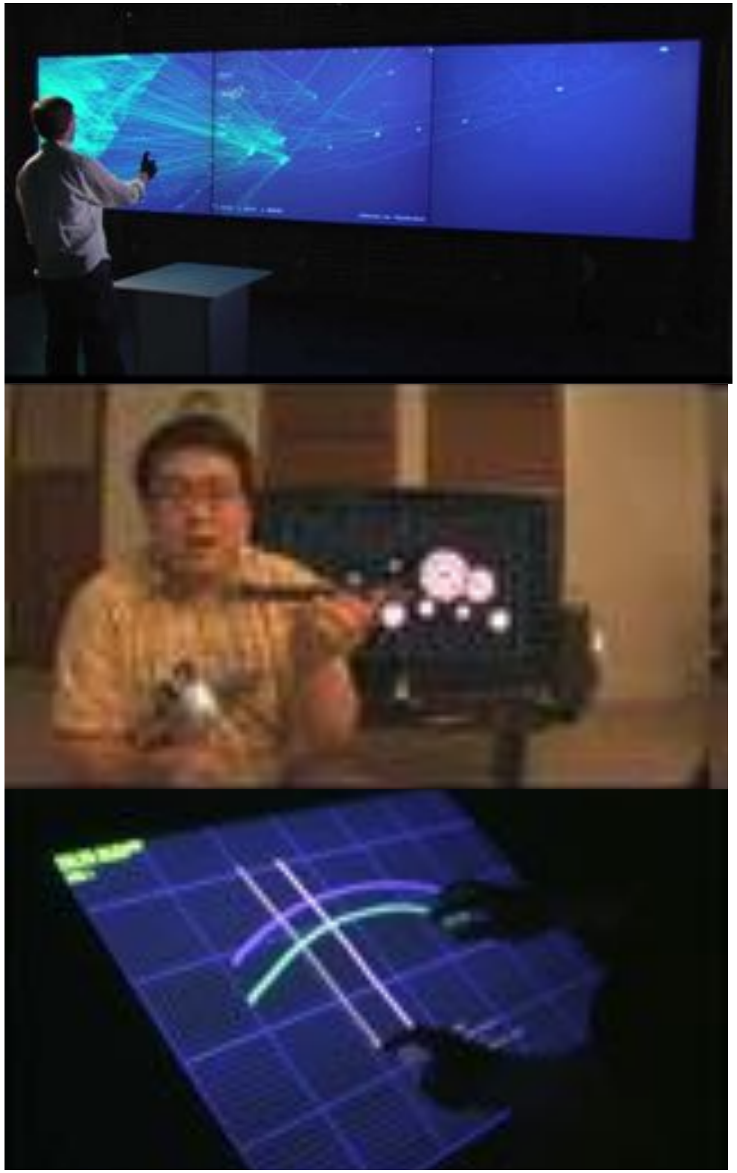


Mid-air gestures (FUI_04)

- Two handed interaction [Buxton]

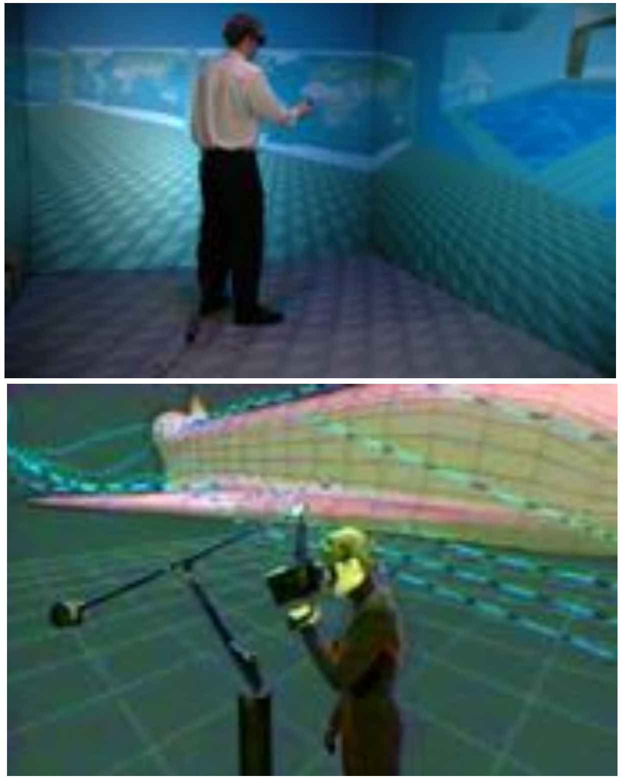
- … and the winner is *Jeff Han* [2006]

  (Bi-manual, multi-point, and multi-user interactions)



Virtual Reality (FUI_06)

- The goals of the virtual reality (VR) community is to **immerse** humans in a virtual world

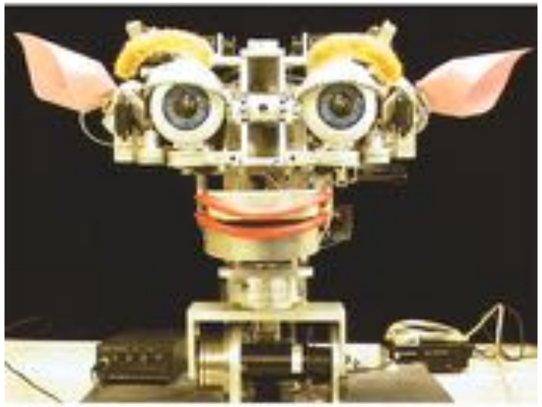


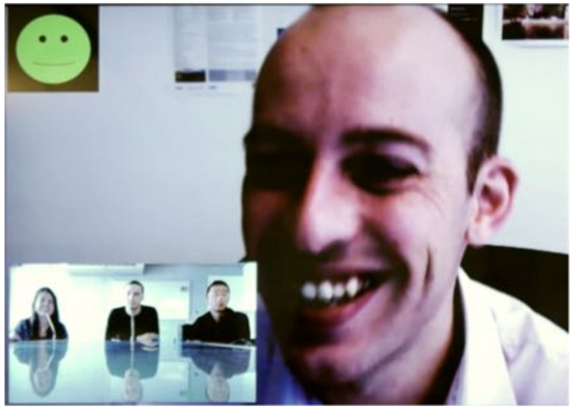
CAVE automated virtual environment at the National Center for Supercomputing Applications (NCSA).

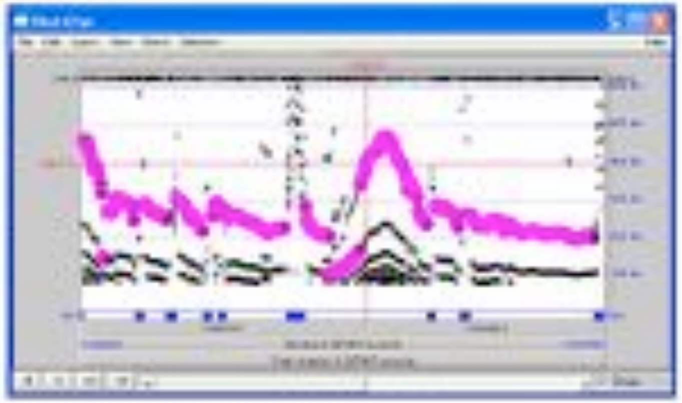
Brain machine interfaces (FUI_09)

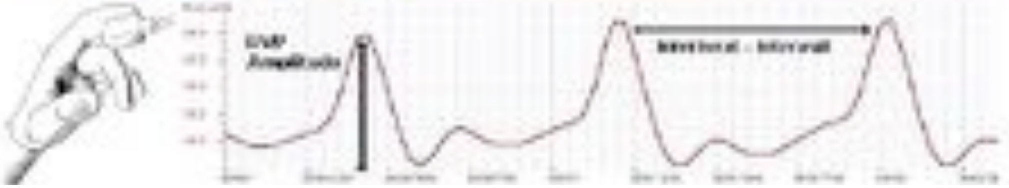
Emotions (FUI_09)

Synthesis & recognition



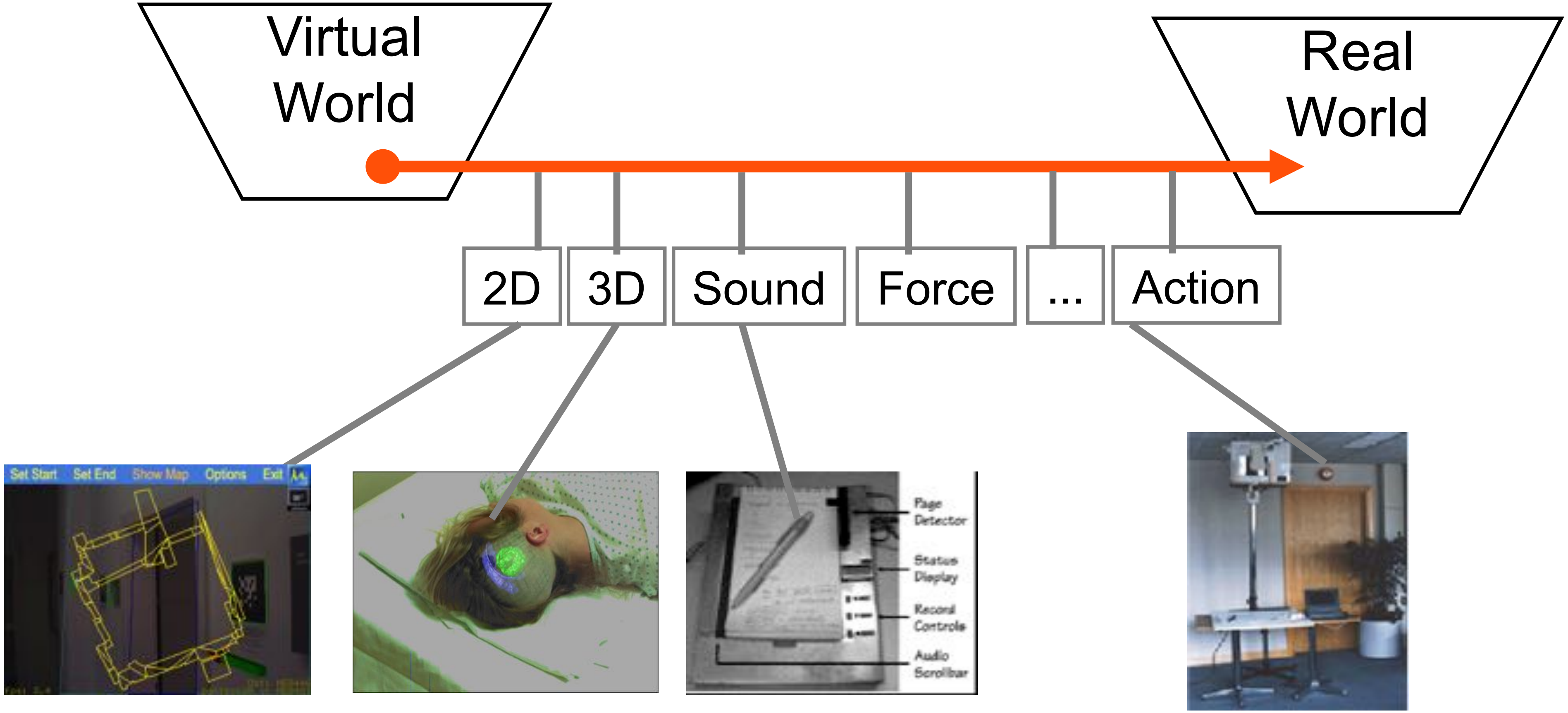






Evolution of Interactions (2)

Augmented Reality (AR)



Virtual World

Real World

2D

3D

Sound

Force

...

Action

Augmented Reality (FUI_06)

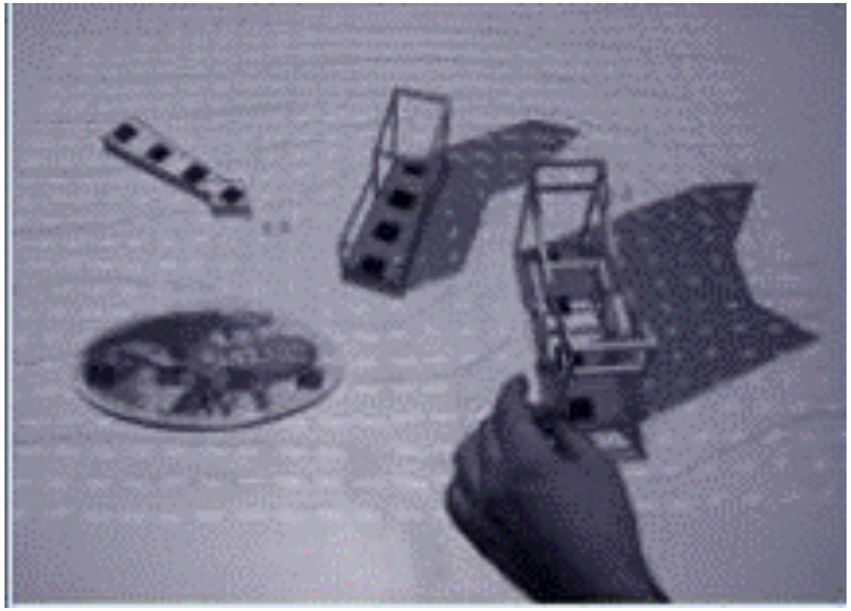
- The goal of AR is to create a seamless integration between real and virtual objects in a way that augments the user's perception and experience.
- Criteria for AR environments
  - The virtual information must be relevant to and in sync with the real-world environment.

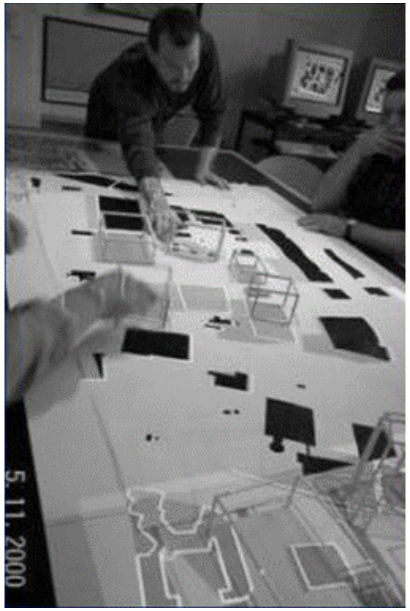
Tangible Interfaces (FUI_05)

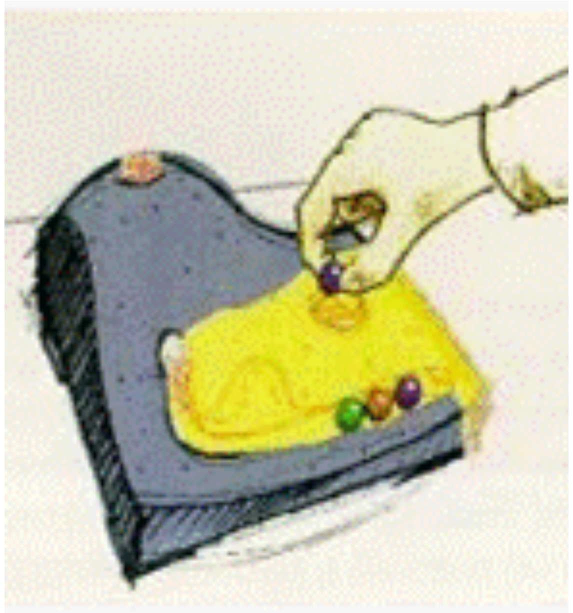
« giving physical form to digital information »

+

interaction capabilities

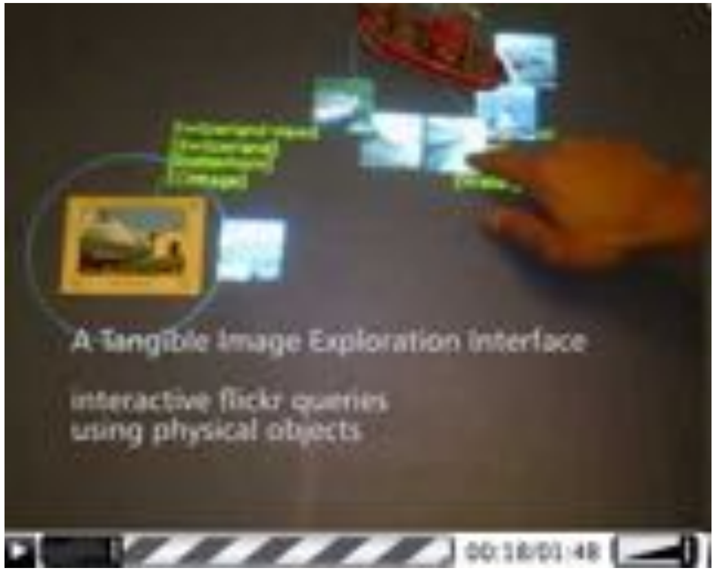






Tangible User Interfaces

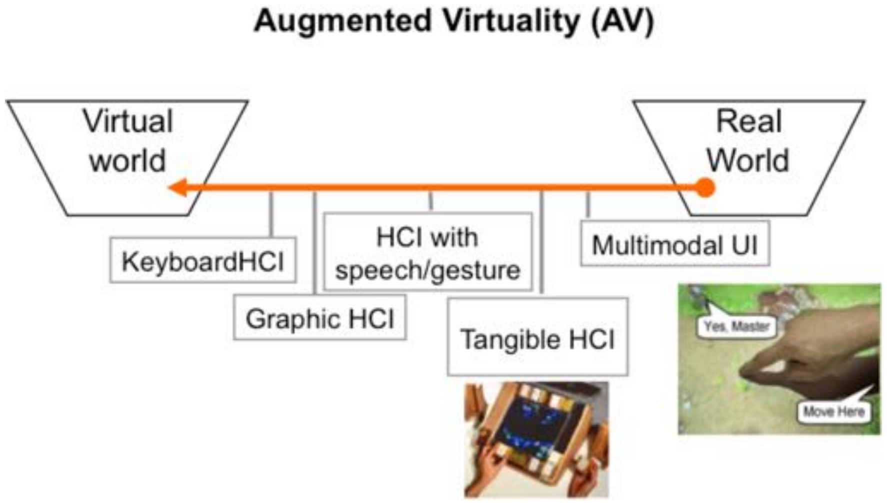
MeModules

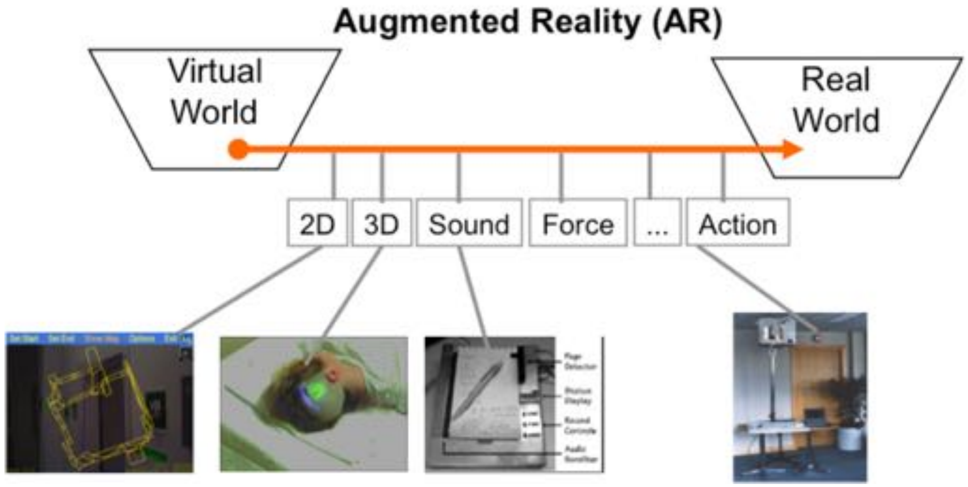


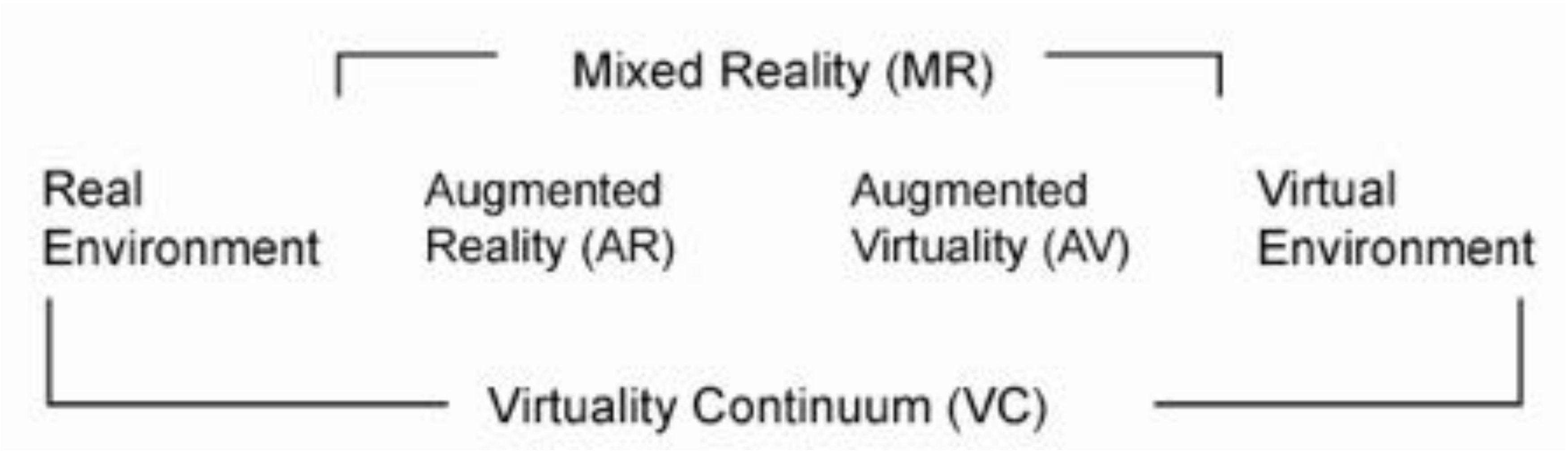
A Tangible Image Exploration Interface

interactive flickr queries
using physical objects

Virtuality Continuum





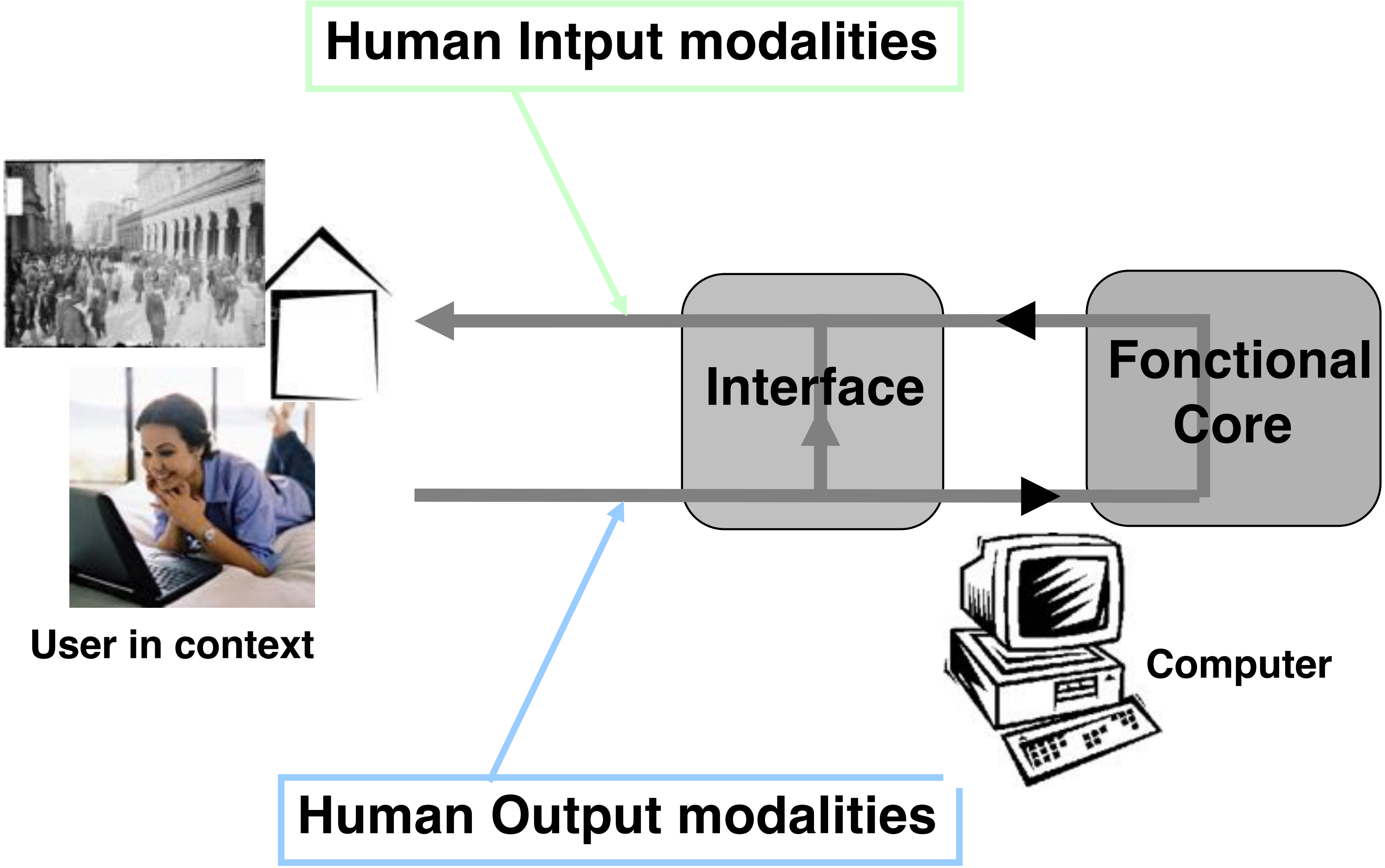


Multimodal Interfaces

[2.4] Designing Multimodal Interfaces

Denis Lalanne

A multimodal interactive system



Human Intput modalities

Interface

Fonctional Core

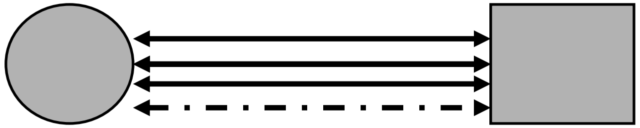
User in context

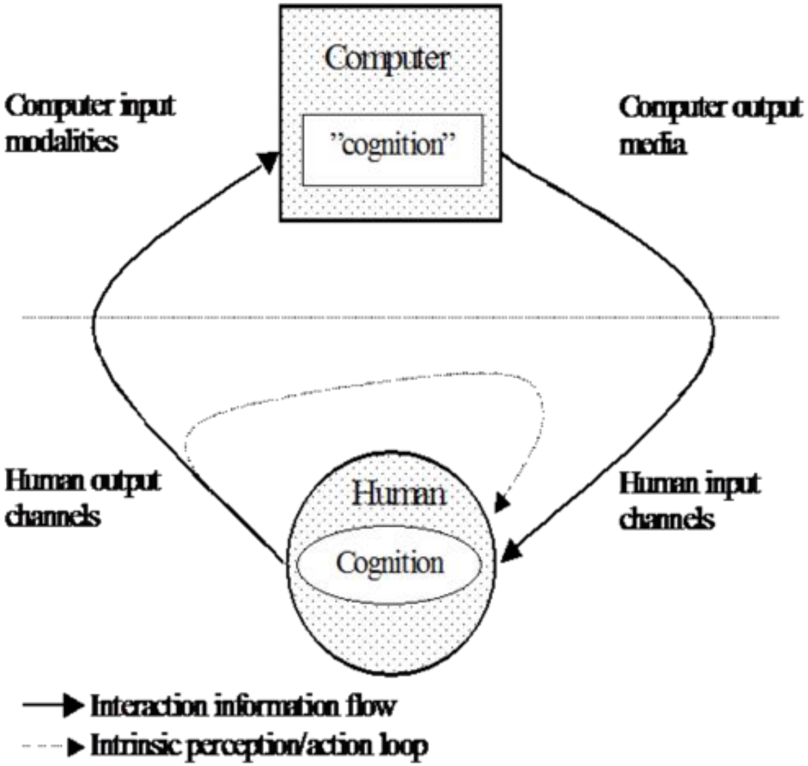
Computer

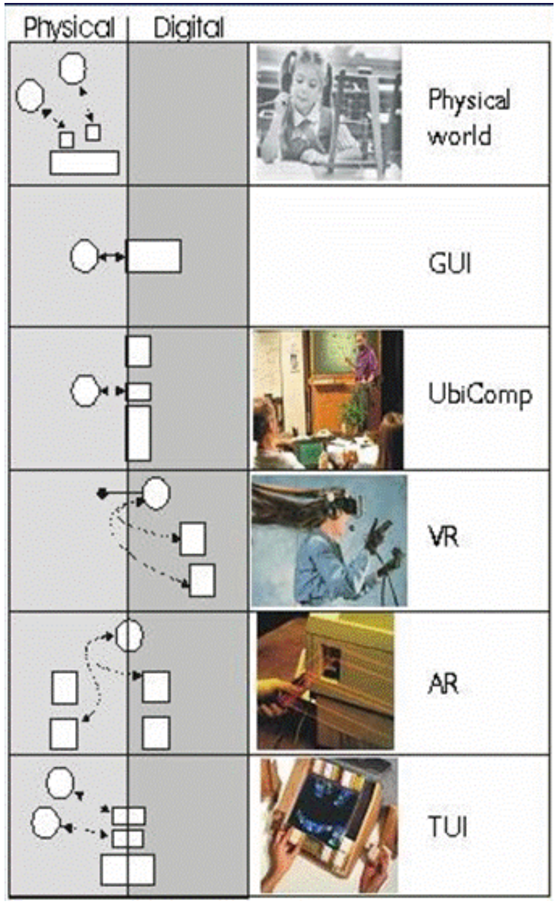
Human Output modalities

Multimodal interfaces ?

Provides various means for humans and machine to interact.





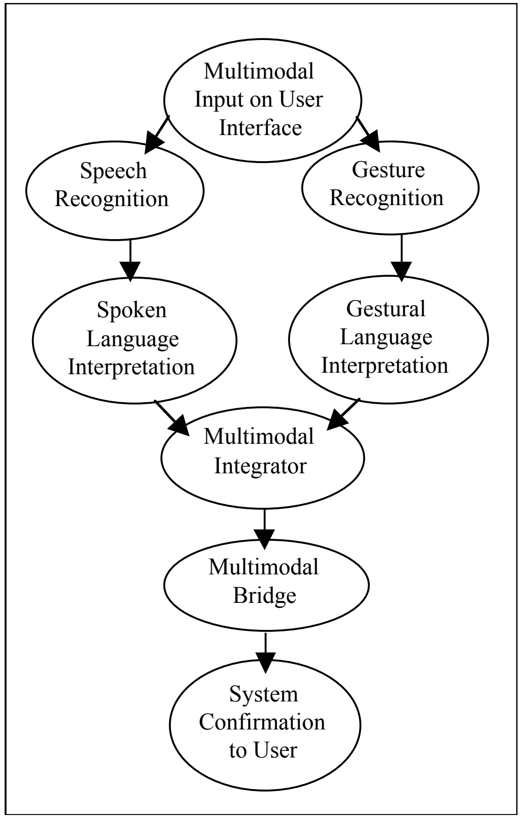


Introduction- Philosophy of Designing Multimodal Systems

- Combine naturally complementary modalities in a manner that optimizes the individual strengths of each, while simultaneously overcoming their weaknesses
- Advantages:
  - 36% fewer errors
  - 10% faster task completion
  - Greater expressive power
  - Greater precision in visual-spatial tasks
  - Support for users' preferred interaction style
  - Accommodation of diverse users, tasks & usage environments (higher recognition rates for "challenging" cases- e.g., accented speakers & mobile environments)

Different Types of Multimodal Systems

- *Alternative* modes vs. *parallel processing* of input signals
- *Feature fusion* vs. *semantic fusion* (e.g., speech & lip movements vs. speech and pen input)


Multimodal Input on User Interface
Speech Recognition
Gesture Recognition
Spoken Language Interpretation
Gestural Language Interpretation
Multimodal Integrator
Multimodal Bridge
System Confirmation to User

Multimodal Processing Flow

Multimodal Interfaces

[2.5] The 10 myths of multimodal integration
(S. Oviatt)

10 Myths of Multimodal Integration

1. *If you build a multimodal system, users will interact multimodally*
2. *Users' multimodal commands are integrated in a uniform way*
3. *Speech & pointing is the dominant multimodal integration pattern*
4. *Multimodal input involves simultaneous signals*
5. *Multimodal integration involves redundancy of content between modes*
6. *Speech is the primary input mode in any multimodal system that includes it*
7. *Different input modes are capable of transmitting comparable content*
8. *Multimodal language is the same as other forms of unimodal language*
9. *Error-prone recognition technologies combine multimodally to produce even greater unreliability*
10. *Enhanced efficiency is the main advantage of multimodal systems*

Myth 1

- *If you build a multimodal system, users will interact multimodally*
- Why is it a myth?
  - Users show indeed a preference to interact multimodally, at least with geospatial tasks…
  - … However this does not mean that users will always interact multimodally.
  - In practice, unimodal and multimodal commands are intermixed.

Myth 2

- *Users' multimodal commands are integrated in a uniform way*
- Why is it a myth?
  - Big differences in the way users will use a multimodal interface
  - Some users might use a quasi-simultaneous integration pattern while other users will input commands sequentially
  - However a user's specific integration pattern remains consistent
  - Interest in multimodal systems able to detect a user's specific integration pattern and adapt to it

Myth 3

- *Speech & pointing is the dominant multimodal integration pattern*
- Why is it a myth?
  - Heritage from “put-that-there”, where gesture and speech were used to select – just as you would with a mouse
  - Modalities can offer a lot more
    - ✓ Written input
    - ✓ Symbolic gestures
    - ✓ Facial expressions
    - ✓ Etc.

Myth 4

- *Multimodal input involves simultaneous signals*
- Why is it a myth?
  - Overlapping signals occur rarely
  - In fact, users frequently introduce (consciously or not) a small delay (1-4 seconds, typically) between two modal inputs
  - Also, difference between overlapping signals for the user and for the machine

Myth 5

- *Multimodal integration involves redundancy of content between modes*
- Why is it a myth?
  - This comes from the early days of research in multimodal interaction, where it was assumed that different modalities could help improve recognition of machine learning-based input modes
  - However, users do not naturally “help” the machine
  - They do so only if the first commands failed, and they want to be sure that the machine understands them
  - Complementarity and equivalence (see CARE properties) are much more prevalent

Myth 6

- *Speech is the primary input mode in any multimodal system that includes it*
- Why is it a myth?
    - In human-to-human communication, speech is considered as a primary input mode
    - However, human-to-machine communication uses very different channels
    - Try indicating a position on the screen by speech only!

Myth 7

- *Different input modes are capable of transmitting comparable content*
- Why is it a myth?
  - The original predicate behind this myth is that two different modalities would share a common language allowing to translate commands from one to the other
  - Some modalities can indeed be compared (e.g. speech and writing)
  - However, as we have seen, each modality has its own very distinctive features
  - E.g. gaze vs. a pointing device like the Wiimote

Myth 8

- *Multimodal language is the same as other forms of unimodal language*
- Why is it a myth?
  - The myth comes by assuming that any language is a language…
  - But is a modality a complete language?
  - Different modalities will complement each other
  - Typical exemple: speech and gesture

Myth 9

- *Error-prone recognition technologies combine multimodally to produce even greater unreliability*
- Why is it a myth?
  - This myth assumes that using two error-prone input modes, such as speech and handwriting recognition, will result in even greater unreliability
  - In fact, increased robustness can be observed
  - Mutual disambiguation between the modalities

Myth 10

- *Enhanced efficiency is the main advantage of multimodal systems*
- Why is it a myth?
  - ➢ Oviatt, in some of her experiments with pen/speech multimodal interfaces, observed only a 10% increase in speed compared to unimodal commands
  - ➢ The bigger advantage is elsewhere:
    - ✓ Fewer errors
    - ✓ Enhanced usability
    - ✓ Flexibility
    - ✓ Mutual disambiguation
    - ✓ Adaptation

What Do These Myths Tell Us?
(see Oviatt et al, 1997 for data)

- People organize their use of modalities in a functionally contrastive manner (i.e. depends on task)
- Multimodal interfaces have the largest performance advantage in visual-spatial domains (e.g., maps)
- Users intermix unimodal & multimodal constructions
- Users ‘integration patterns:
  - Simultaneous or sequential integrators (bimodal user groups)
  - Individual users highly consistent in their pattern
- Complementary linguistic content carried in spoken and written modes

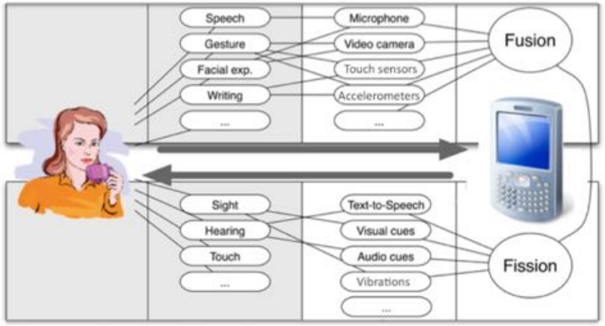
References

- Bolt, R. A. “Put-that-there”: Voice and Gesture at the Graphics Interface. In Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 1980) (Seattle, USA, July 1980), 262–270.
- Coutaz, J., Nigay, L., Salber, D., Blandford, A., May, J., and Young, R. M. Four Easy Pieces for Assessing the Usability of Multimodal Interaction: The CARE Properties. In Proceedings of the 5th International Conference on Human-Computer Interaction (Interact 1995) (Lillehammer, Norway, June 1995), 115–120.
- Dumas, B., Lalanne, D., Oviatt, S. 2009. Multimodal Interfaces: A Survey of Principles, Models and Frameworks. In Denis Lalanne, Jürg Kohlas eds. Human Machine Interaction, LNCS 5440, Springer- Verlag, pp. 3-26 (2009).
- Oviatt, S.L. 1999. Ten myths of multimodal interaction. In Communications of the ACM, 42(11), New York: ACM Press, pp. 74-81 (1999).

Multimodal Interfaces

[2.6] Perspectives

Key concepts



- Physicality – Natural Interfaces
  - Augmented reality
  - The disapearing and ubiquitous computer (distributed in the environment)
- Mobility
  - Personal data available everywhere
  - Mobility of user and of interaction device
- Plasticity
  - A form of adaptation to
    - ✓ The context of interaction
    - ✓ The various interaction devices
    - ✓ The user and its interaction patterns
- Multimodality is a key element to achieve these concepts

Which Modality?

- Best match between
  - Requirements
  - Available technologies
  - User, Task, Context
- Some technologies are better for collaborative work in co-presence (e.g. tangibles), some are better for precise selection individual tasks (e.g. mouse), some are better for entertainment (e.g. gestures)…
- It is all about choosing the best technology for its purpose, for its user

Multimodal Interfaces

[2.7] What you should know by now ?
&
what’s next?

What you should be able to answer…

- What is multimodal interaction?
- What is a multimodal versus a multimedia system?
- What is a modality?
- When are speech, eyes gaze, pen-based interaction useful?
- What is augmented reality? augmented virtuality?
- What are the advantages of multimodal interaction?
- What are the 10 myths related to multimodal interaction?

What’s next? (MMI_03)

- A model of multimodal communication
- Fundamental problems to solve with multimodal systems
- The CARE and CASE models
- Multimodal Fusion and the known mechanisms
- Multimodal Fission
- Key properties of multimodality
- Present your multimodal interface design