Interaction Homme-Machine
Interaction multimodale

Jeudi 4 novembre 2010

Intervenante

Laurence Nigay, professeur à l'Université Joseph Fourier Grenoble 1
Laboratoire d'Informatique de Grenoble (LIG) - Équipe Ingénierie de l'Interaction Homme-Machine
B.P. 53 - 38041 Grenoble cedex 9 Laurence.nigay@imag.fr http://iihm.imag.fr/nigay/

Laurence Nigay est Professeur à l'Université Joseph Fourier et responsable de l'équipe "Ingénierie de l'Interaction Homme-Machine" (IIHM) du LIG. Le CNRS lui a décerné la médaille de bronze en 2002 pour ses travaux de recherche. Depuis septembre 2004, elle est membre de l'Institut Universitaire de France.
Elle est responsable depuis 2006 du Master 2 Professionnel Génie Informatique (M2Pro GI). Elle enseigne les modèles (conception ergonomique et conception logicielle) pour l'Interaction Homme-Machine, l'interaction multimodale, l'interaction sur supports mobiles et les collecticiels en M2Pro GI, en 3ème année de Polytech' Grenoble filière RICM (Réseaux Informatiques et Communication Multimédia), en 1ère année de Polytech' Grenoble filière TIS (Technologies de l'Information pour la Santé) et en M2 Recherche.

Ses travaux de thèse présentés en 1994 ont trait à l'interaction multimodale et à ses aspects logiciels. Ses travaux de recherche actuels ont trait à la conception et à la modélisation logicielles des systèmes interactifs. Parmi les systèmes, ses travaux portent particulièrement sur les interfaces utilisateur qui intègrent les aspects innovants de la technologie en communication homme-machine : les systèmes multimodaux, les systèmes sur supports mobiles et les systèmes de réalité augmentée et les collecticiels. Elle a effectué plusieurs séjours à l'Université de Carnegie-Mellon (USA) et a été visiteur scientifique à l'Université de Glasgow pendant un an (2002). Elle a été co-responsable du groupe de travail international WG 2.7 "Ingénierie de l'interaction" de l'IFIP jusqu'en juillet 2004 et a participé pendant 5 ans aux travaux du projet européen ESPRIT AMODEUS puis le réseau européen TMR-TACIT. Elle a ensuite participé au réseau d'excellence européen SIMILAR (FP6, 2003-2007) sur la multimodalité et a été coordinatrice du projet européen OpenInterface (FP6, 2006-2009). Dans le cadre du GDR-PRC I3, elle a été co-responsable du groupe de travail sur l'interaction multimodale et celui sur les collecticiels et est actuellement co-responsable du groupe de travail sur l'informatique mobile et ubiquitaire. Elle est aussi membre du comité de programme des colloques AVI, CHI, MobileHCI, DSVIS, EHCI, INTERACT, EIS, IHM et UBIMOB. Elle a publié plus de 150 articles dans des conférences internationales, chapitres dans des livres, articles de revues et est co-auteur du livre "Design Principles for Interactive Software" (Chapman&Hall, Groupe WG2.7 de l'IFIP).

Contenu du cours

Le cours a pour thème la conception et la réalisation des systèmes interactifs. Parmi ces systèmes, nous étudions les interfaces utilisateur qui intègrent les aspects de la technologie actuelle en interaction homme-machine : les interfaces multimodales. Nous soulignons aussi que la multimodalité est un vecteur intégrateur de nombreuses techniques d'interaction innovantes comme les interfaces tangibles, manipulables (Embodied User Interface) et les interfaces sur supports mobiles.
Le cours comprend trois parties :

- Introduction : domaine en évolution permanente, panorama des modalités d'interaction et exemples de systèmes multimodaux
- Espaces de conception et de classification des systèmes multimodaux : cette partie est dédiée à la conception ergonomique des interfaces multimodales
- Plateforme logicielle : Lien entre la conception et une approche à composants

Les concepts avancés d'interaction seront illustrés au moyen de nombreux exemples par le biais de vidéos et démonstrations.

Mots-clefs

Multimodalité, modalité d'interaction, système interactif, interface homme-machine, espace de conception, langage d'interaction, dispositif physique.

Multimodality: Introduction

Domain
Definitions
Path to evolution

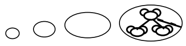
Introduction

Man-Machine Interface
Handling multimodal interaction

A modality

A multimodal system

Software architecture model for multimodal systems



Fusion of different objects from various modelling techniques:
How ?

At which level of abstraction?

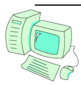
Domain and definitions

- Beyond the traditional User Interface (UI)
  - Windows: scroll, resize, move
  - Icons: representations, drag/drop
  - Menus: pop-up, pull-down
  - Pointers: mouse, digitizer, trackball, etc.
- Multimodal systems
  - Multi-modal refers to interfaces that support non-GUI interaction
  - Speech and pen input are two common examples - and are complementary

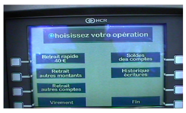
From WIMP to Post-WIMP

Integration into the env.

Mobility



Traditional interaction



[Lyytinen & Yoo 2002]

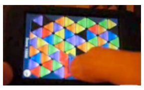
From WIMP to Post-WIMP

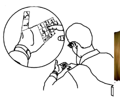
Integration into the env.

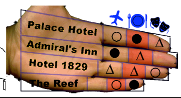
Mobility

Traditional interaction

Mobile computing







Conclusion: From WIMP to Post-WIMP

Integration into the env.

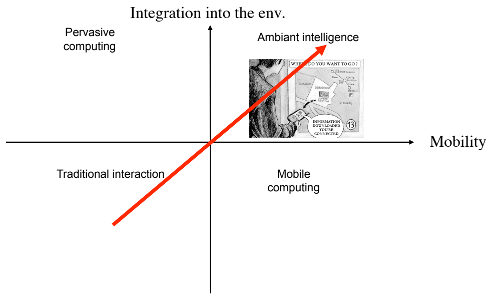
Pervasive computing

Mobility

Traditional interaction

Mobile computing

Conclusion: From WIMP to Post-WIMP



Domain and definitions

"New Interfaces" extend the sensori-motor capabilities of computer systems

Multimodal ≠ Multimedia

Multimodal ≠Speech interface

New interaction capabilities appear

8

Media - Modality

- Media
  - material (signal on a channel)
  - the support of communication
- Modality
  - a channel or path of communication between the human and the computer
    - sensorial (audition, vision, etc.)
    - of communicating (voice, gestures, facial expressions, etc.)
  - A modality is a process of receiving and producing chunks of information

9

Multimedia - Multimodality

- Multimedia system
  - transport signals of different kinds
    - For ex.: a sound clip attached to a presentation
- Multimodal system
  - interpret signs belonging to various sensory and communication modalities
    - For ex.: the combined input of speech and typing in a word processor

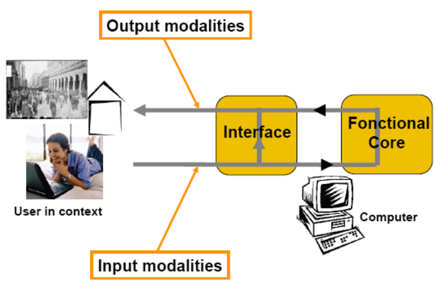
10

Multimodal and crossmodal

- Multimodal interaction makes use of several input and/or feedback modalities in interacting with a computer system.
  - Examples of modalities: manual gestures, gaze, touch, speech, head & body movements
  - Modality: human sensory channel, different representation modality, or different input method
- Crossmodal interaction makes use of a different human sensory modality to present information typically presented through another modality.

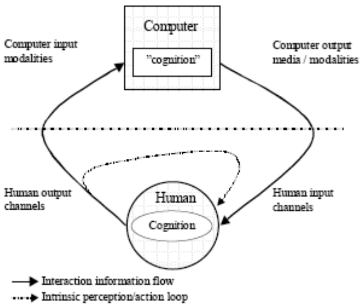
11

Input/Output modality



12

Multimodal interaction



13

Why multimodal?

- Most technologies are mature
- Seek to optimize the distribution of information over different modalities
- For adaptive, cooperative and flexible interaction among people

14

Why multimodal?

- Naturalness
  - provide more "natural" interfaces Usability
- Usability / flexibility
  - improve ease-of-use
- Robustness/Efficiency/Accuraccy
  - decrease error rates (Mutual disambiguation of recognition errors)
- Perception
- Relieve burden on the visual channel
- Support users with disabilities

15

Natural interaction and multimodality

- Natural interaction is the long-term goal of being able to communicate with machines in the same ways in which humans communicate with one another
  - Input/output audiovisual speech, facial expression, gesture, gaze, body posture, physical action, touch, etc.
- **Natural interaction is multimodal by nature**

16

Why multimodal?

- Flexibility for Robustness
  - Advantages for error recovery
    - Users intuitively pick the modality that is less error-prone
    - Language is often simplified
    - Users intuitively switch modality after an error, so that the same problem is not repeated
- Flexibility for
  - Users with disability (permanent or temporary)
  - Variable usage context (**mobile support**)
- The flexibility of a multimodal interface can accommodate **a wide range of users, tasks, and environments** for which any given single mode may not suffice

17

Input Multimodality

- Because of the user's circumstances – including her task, her background, her training, her knowledge, and the physical and interactive behaviour of the computer interface – the user may well have preferences as to how she communicates with the computer.
    - A familiar example is that if the user is engaged in a task which occupies her hands, she may prefer to use speech.
    - Another example: Suppose that the user wishes to book a flight from somewhere in Europe to Las Vegas. She may not know what is the nearest international airport, so she would prefer to indicate her destination by pointing on a map – or at the very least, by choosing from an appropriately filtered list of airports.

Why multimodal?

- What do these persons have in common?

19

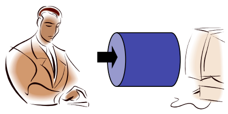
Why multimodal?

- Enabling the user
- New multimodal technologies enable the user to be better engaged in the interaction to receive more information through several modalities
- Multimodal interaction makes using of information technology possible for people with special needs, e.g., for blind and visually impaired people

20

Why multimodal?

- The combination of human output channels effectively (multimodal input interaction) increases the bandwidth of the human machine channel.
  - *This has been discovered in many empirical studies of multimodal human computer interaction*



21

Why multimodal? Nevertheless…

- Adding extra output modality requires more neurocomputational resources and will lead to deteriorated output quality resulting in reduced effective bandwidth.
- Two types of effects are usually observed:
  - a slowdown of all output processes, and
  - interference errors due to the fact that attention cannot be divided between the number of output channels.
- Two examples of this: writing when speaking, and speaking when driving a car.

22

Three paradigms for multimodality

- **Computer as tool**
- Multiple input modalities are used to enhance direct manipulation behavior of the system
  - the computer is a passive tool and tries to understand the user through all the different input modalities that the system recognizes
  - the user is responsible for initiating the actions
  - follows the principles of direct manipulation [Shneiderman, 1982]

23

Three paradigms for multimodality

- **Computer as partner**
- The multiple modalities are used to increase the anthropomorphism of the user interface
  - agent based conversational user interfaces
  - multimodal output is important: talking heads and other humanlike presentation modalities
  - speech recognition is a common input modality in these systems, and speech synthesis is used as an output modality

24

Three paradigms for multimodality

- **Proactive computing (ubicomp, PUI, …)**
- The multiple modalities are used to sense the user and the environment
  - multimodal (multisensory) input is important
  - the functionality of the system depends on the level of deduction (AI) the system is capable of
  - proactive functionality is often in the background and only indirectly visible for the user, predicting his/her actions and needs

25

Multimodality: Path to evolution

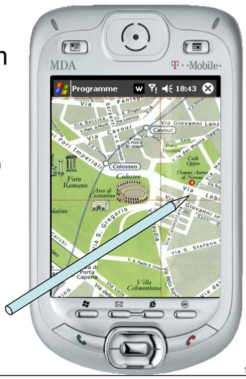
- Since 1980 “Put that there” paradigm
  R. Bolt MIT

26

Multimodality: Path to evolution

- “Put that there” paradigm
  R. Bolt MIT

„Zoom in here”



User selects a point of interest clicking with a stylus and speaking in order to focus it.

27

Multimodality: Path to evolution

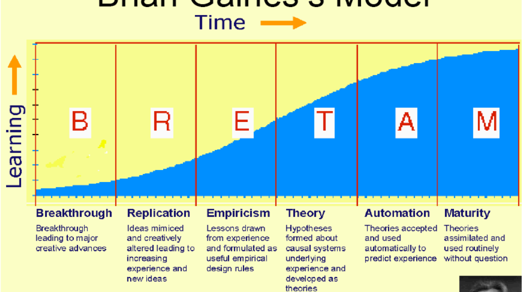
- “Put that there” paradigm
  R. Bolt MIT

„Play this sound logo”

User selects a sound logo by clicking on the title with a stylus and speaking in order to hear it

28

Multimodality: Path to evolution
Brian Gaines’s Model



In the 80’s, Brian Gaines introduced a model on how science technology develops over time

Brian Gaines

29

Readings

- Bolt, R. A. “Put-that-there”: Voice and gesture at the graphics interface. Proceedings of SIGGRAPH’80, 14, 3 (1980), 262–270
- Martin, J. C. TYCOON: Theoretical Framework and Software Tools for Multimodal Interfaces. Intelligence and Multimodality in Multimedia Interfaces, AAAI Press (1997)
- Nigay, L., Coutaz, J. The CARE Properties and Their Impact on Software Design. Intelligence and Multimodality in Multimedia Interfaces, (1997)
  http://iihm.imag.fr/publs/1997/IMMI97-ChapterNigay.pdf
- Oviatt, S. “Ten myths of multimodal Interaction”, Comm. of the ACM, 42, 11 (1999), 74-81
- Turk, M., Robertson, G. Eds, Perceptual user Interfaces. Comm. of the ACM, 43, 3 (2000), 32-70

30

Readings

- ACM SIGCHI: ACM's Special Interest Group on Computer-Human Interaction
  - http://www.sigchi.org/

- ICMI conference
  - International Conference on Multimodal Interfaces
- CHI conference
  - Computer Human Interface
- UIST conference
  - User Interface Software and technology
- MobileHCI conference
  - Human Computer Interaction with Mobile Devices and Services

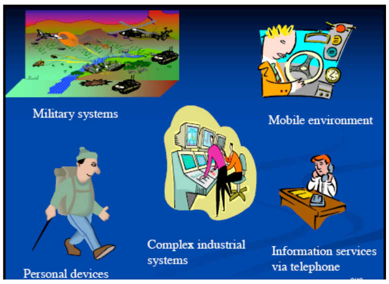
31

Multimodal systems

Application domains
Examples

1

Application domains



2

Three paradigms for multimodality

- **Computer as tool**
  - Multiple input modalities are used to enhance direct manipulation behavior of the system
- **Computer as partner**
  - The multiple modalities are used to increase the anthropomorphism of the user interface
- **Proactive computing (ubicomp, PUI, …)**
  - The multiple modalities are used to sense the user and the environment

3

Three paradigms for multimodality

Computer as tool | **Computer as partner** — **Active modalities**

Proactive computing (ubicomp, PUI, …) — **Passive modalities** ***Sensing modalities***

4

Active/Passive modalities

- Active modalities are used by the user to issue a command to the computer (e.g., a voice command)
- Passive modalities are used to capture relevant information for enhancing the realization of the task, information that is not explicitly expressed by the user to the computer such as eye tracking location/orientation tracking etc.
- Combination of active and passive modalities

5

Three paradigms for multimodality

Computer as tool | **Computer as partner** — **Active modalities**

Proactive computing (ubicomp, PUI, …) — **Passive modalities** ***Sensing modalities***

6

Computer as tool
Multimodality

- Computer as tool
- The user is responsible for initiating the actions
- Multiple input/output modalities are used to enhance direct manipulation behavior of the system
  - Interaction modalities

7

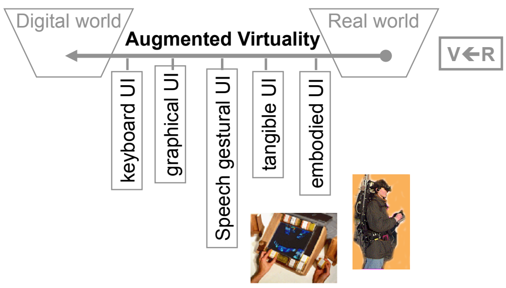
Computer as tool
Augmented Reality / Augmented Virtuality

Purpose of the task = real world

Purpose of the task = computer

8

Augmented virtuality



9

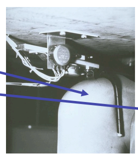
Once upon a time HCI ...

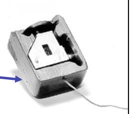
- Ivan Sutherland 1963 (PhD thesis MIT)
- Sketchpad
  - Drawing tool
  - Optical pen and buttons
  - Direct manipulation
  - Icons
  - Zoom
  - Copy/Paste

10

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

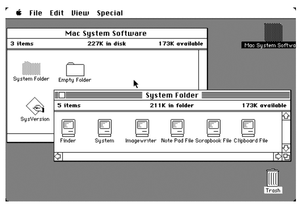




11

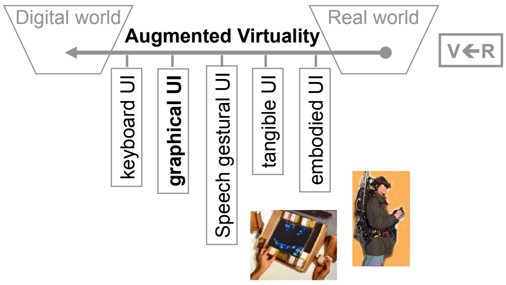
Once upon a time HCI ......

- Macintosh 1984
  - Direct manipulation
  - Mouse and Keyboard



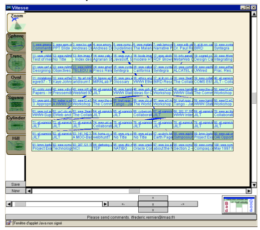
12

Augmented virtuality



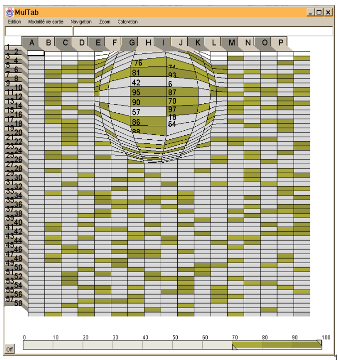
Graphical UI

- Vitesse: LIG-IIHM
- Multiple output modalities
  - *Video VITESSE/MultipleView.avi*

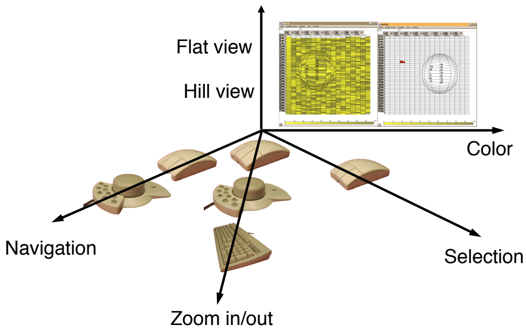


Graphical UI

- MULTAB
- LIG-IIHM



Experimental evaluation: MULTAB

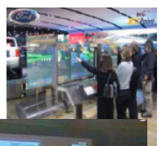


Graphical UI

- MERL table
  - *Video: TableRondeVernier.mpg*

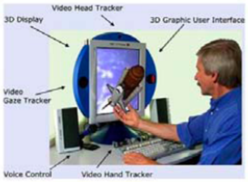
Graphical UI

- Multi-surface interaction

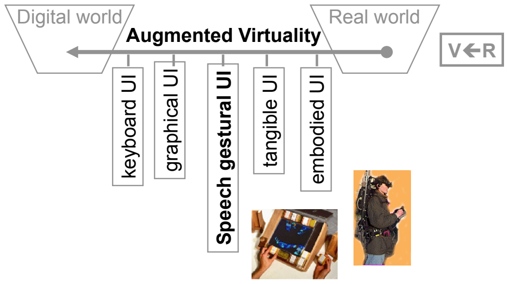


Manipulation and stereo

- Input modalities
  - 3D gesture
  - Speech
- Sensing modalities
  - Head tracker
  - Eye gaze tracker



Augmented virtuality



Meditor: Multimode Text Editor

- MEDITOR: Y. Bellick LIMSI-Paris
- Combines keyboard, Braille terminal, a French text-to-speech synthesiser, and a speech recognition system
- Allows Blind people to perform simple Document editing tasks

Meditor

Meditor Commands

- <u>To put a word in italic</u>
- The user says "*italic*" while clicking on any character of the word on the Braille terminal.
- <u>To place a character into an exponent position</u>
- The user says "*character exponent*" while clicking on the corresponding character.
- <u>To delete a part of the text</u>
- 1) The user says "*begin selection"* while clicking on the first character of the string to be deleted, 2) then says "*end selection*" while clicking on the last character, and 3) says "*delete*" to complete the command. The message feed-back "*selection deleted*" is then generated by the speech-synthesizer.

Speech + gesture

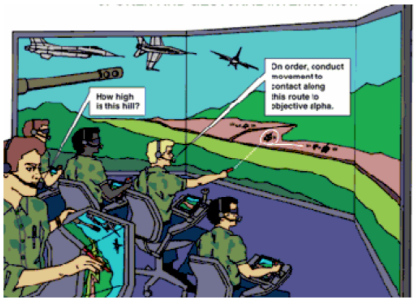
- VoicePaint LIG-IIHM
  - Graphical editor
  - Mouse + speech
    - Change colors using speech while drawing using the mouse

Quickset
Spoken and gestural interaction

- Speech
- Pen input
  - Pointing (selection)
  - Gesture recognition

25

Quickset
Spoken and gestural interaction



26

Quickset
Spoken and gestural interaction

27

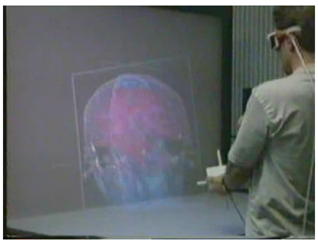
Quickset - RASA: Multimodality and Augmented Reality

- *Video DEMOMULTI/Rasa.mpg*

28

Gestural UI

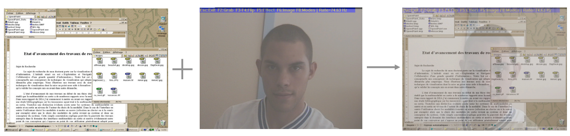
- 3D cubic mouse
  - *Video CubicMouse.mov*



29

Gestural UI

- Mirror Pixels
- Fusion of two images, one is the user interface, one is a video pointing to the user or to his/her hand

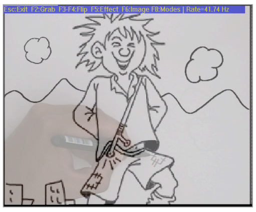


Interface Video Result

30

Gestural UI

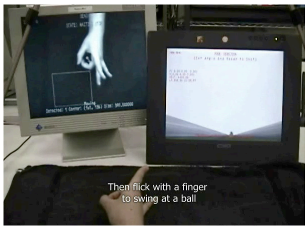
- Mirror Pixels: Drawing application
  - Input modality based on a pen
    - *Video PIXMIRROR/App_Dessin.mpg*



31

Gestural UI

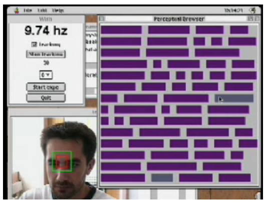
- Flicking gesture
  - Vision-based tracking
    - *Video: DEMOMULTI/BilleGestureVision-UIST06.mpg*



32

Gestural UI

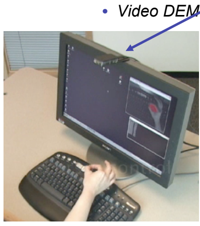
- Perceptual Browser LIG-IIHM
  - Two modalities
    - Head movement (vision based tracking) + mouse
      - *Video PBrowserMac.mpg*

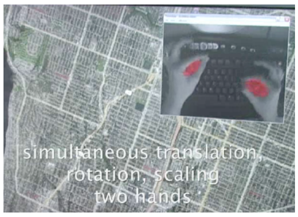


33

Gestural UI

- Thumb and Fore-Finger Interface Microsoft
  - Modality based on vision-tracking
    - *Video DEMOMULTI/HandsOverKeyboard.wmv*

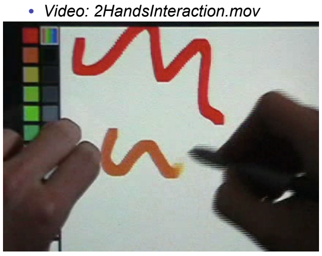




34

Two-handed interaction

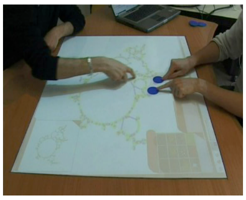
- Drawing editor: Berkeley
  - Two modalities (one per hand)
    - *Video: 2HandsInteraction.mov*



35

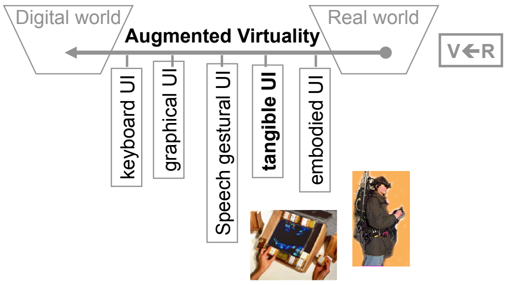
Two-handed interaction

- NavGRAPH: LIG-IIHM
- NAVRNA: a system to visualize, explore and edit RNA Ribonucleic Acid
  - *Video:* NAVGRAPHE/AVI_2.wmv



36

Augmented virtuality



37

Tangible UI

- AmbientRoom : MIT Medialab
  - *Video: AmbientRoom.mpg*

38

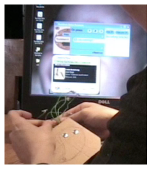
Tangible UI

- Lumino:
  tangible blocks for tabletop computers
  - *Video: Lumino.mp4*

39

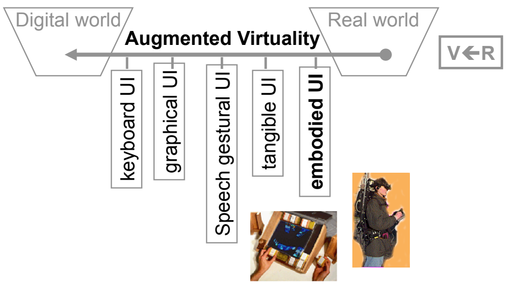
Tangible UI

- Rapid Construction of physical interfaces CMU
  - *Video: PhysicalPrototype-UIST06.mov*



40

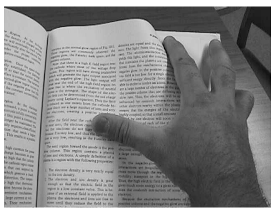
Augmented virtuality



41

Embodied User Interface

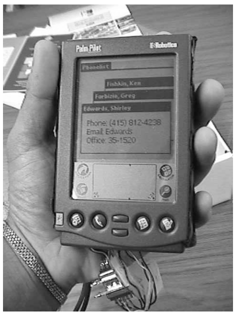
- Embodied UI: Rank Xerox



42

Embodied User Interface

- Embodied UI: Rank Xerox



43

Embodied User Interface

- Tilt and gesture based user input
  - Compaq project

Rock'n'Scroll

Video: RocknScroll.mov

44

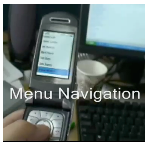
Embodied User Interface

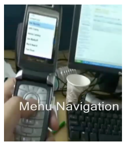
Peephole Displays
Ka-Ping Yee
Uni. of California, Berkeley
Video: CHI03VIDEO/Yee.avi

45

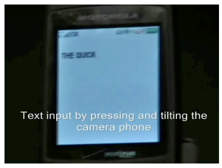
Embodied User Interface

- Camera phone based motion sensing





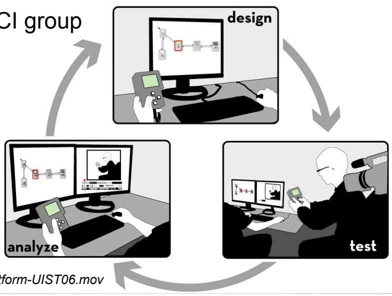
Uni. of California, Berkeley
& IBM
Video: Phone3DMvt-UIST06.mov



46

Embodied User Interface

- Prototyping tool for embodied UI
  - Stanford HCI group

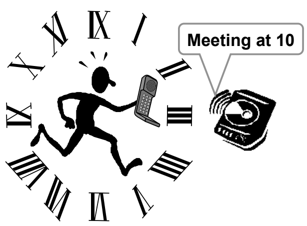


–*Video: MapPDADesignplatform-UIST06.mov*

47

3D sound

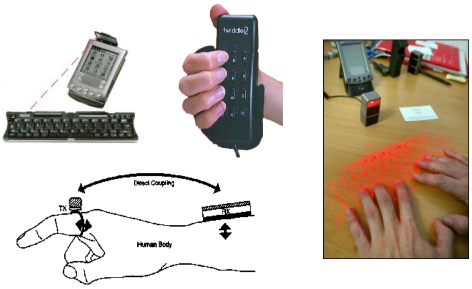
- Spatialized sound
- Mobile setting



Soundbeam Necklet

48

Keyboard

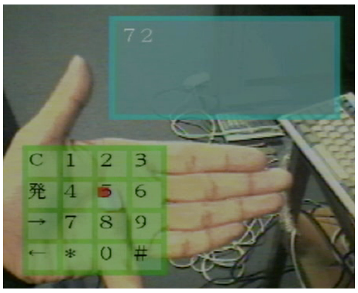


Keyboard

- http://mozillalabs.com/conceptseries/2010/09/23/seabird/

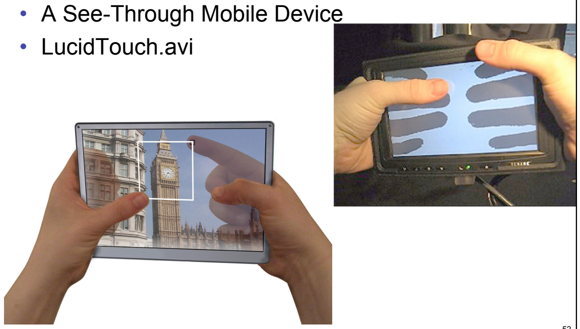
Keyboard

- Oulu University (Finland)
- *Video: Main.mpeg*



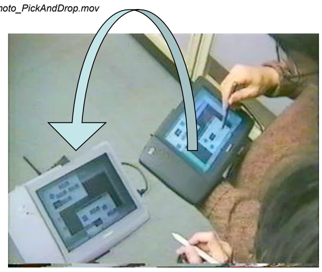
Interaction: LucidTouch:

- A See-Through Mobile Device
- LucidTouch.avi



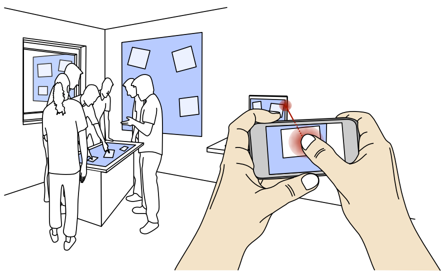
Multi-surface interaction

- Pick&Drop
  - *Video: Rekimoto_PickAndDrop.mov*

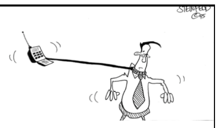


Multi-surface interaction

- TouchProject
  - *Video: TouchProjector.wmv*



Direct manipulation and mobility

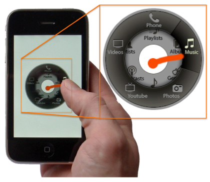


- Key issues:
  Transparency/Usability of mobile devices
- Challenges for HCI
  - Limited interactional resources

  - Interaction in mobility

Direct manipulation and mobility

- Wavelet menu
  - Video: MenuWavelet.video



- Problem space
  - Space on screen
  - No keyboard for shortcuts
  - One-hand interaction
  - Eye-free interaction

56

MATCH

- Multimodal Access to City Help
- A Multimode Portable Device that accepts speech and pen gestures created by ATT&T

57

MATCH

- Part of a multi-million, multi-year contract from DARPA
- Enables users to interact using speech, pen, or synchronized combinations of speech and pen
- Essentially a testbed for designing portable multimodal applications
  - *Video: DEMOMULTI/CityPlannerATT.mpeg*

58

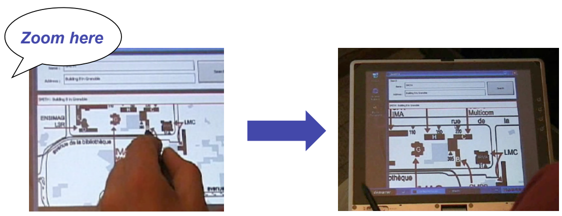
MATCH Testing Statistics

- Exchanges 338
  - Speech only 171 (51 %)
  - Multimodal 93 (28%)
  - Pen only 66 (19%)
  - GUI actions 8 (2%)

59

YellowPages

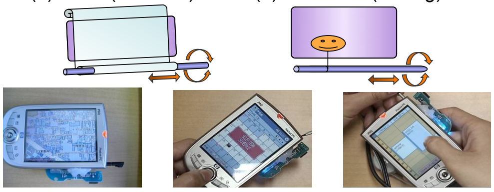
- YellowPages: LIG-IIHM
  - Voice command + Pointing using stylus
    - *Video:YellowPages-CLIPS.wmv*



60

RodDirect

- Two types of stylus movement in the stylus holder are associated with parameters
  - (1) Twist (Rotation) (2) Push/Pull (Sliding)



Map Browser | Application Switching | Scheduler

JAIST: Japan Advanced Institute of Science and Technology, JAPAN
Video: RodDirectDemo.mpg

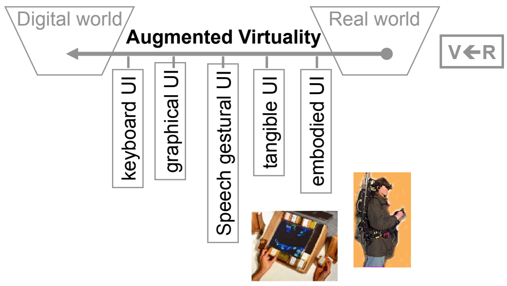
Multimodality on mobile phone

- Contact manager application on a mobile phone: LIG-IIHM
  - Devices : keyboard + microphone
  - Enabling forms filling using the stylus and speech commands

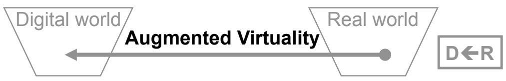
Commercial product

- www.kirusa.com/multimodality.html

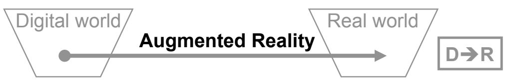
Augmented virtuality



Computer as tool
Augmented Reality / Augmented Virtuality



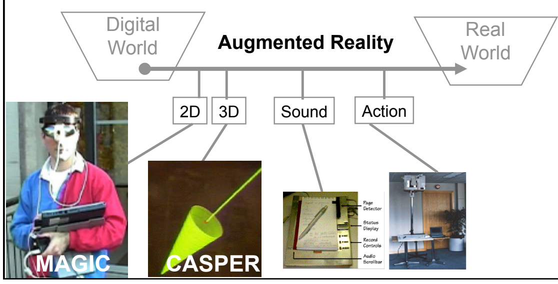
Purpose of the task = computer



Purpose of the task = real world

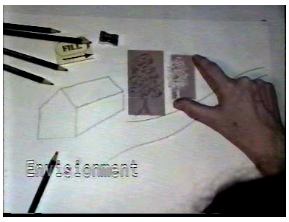
Augmented Reality

- New interaction modalities



Augmented Reality

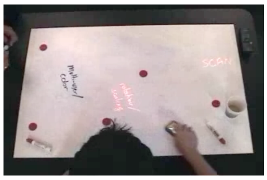
- First AR system
  - P. Wellner DigitalDesk
    - Video: DigitalSketching.mpg



67

Augmented Reality

- MagicTable: LIG-IIHM
  - Brainstorming
    - Video: MAGICBOARD/magicboard.mov



68

Context-aware interactive system

- Sensing modalities

69

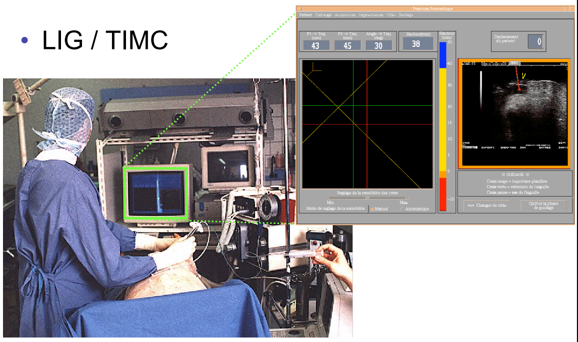
Context-aware interactive system

- Input modalities
  - Sensing modalities
    - 3D Location
    - 3D Orientation
- Output modality
  - Textual information displayed on HMD
  - Integration of virtual information and actions in the real world of the user through modalities

70

Augmented surgery: CASPER

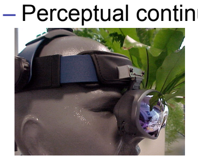
- LIG / TIMC

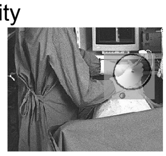


71

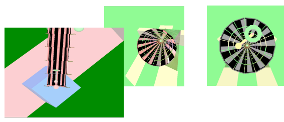
Augmented surgery: CASPER

- Output modality
  - Perceptual continuity



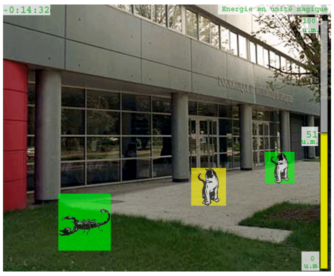


  - Cognitive continuity



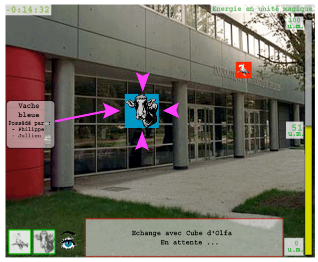
72

TROC



73

TROC



74

Conclusion

- Three paradigms for multimodality

Computer as tool	**Computer as partner**	**Active modalities**
Proactive computing (ubicomp, PUI, …)		**Passive modalities** ***Sensing modalities***

75

Conclusion: Computer as tool

Digital world — **Augmented Virtuality** — Real world — D←R

Purpose of the task = computer

Digital world — **Augmented Reality** — Real world — D→R

Purpose of the task = real world

76

Conclusion

- Modality and multimodality:
- **A VAST space of possibilities to be explored**
- Multimodality is an integrating vector for several recent interaction paradigms that include:
  - perceptual user interfaces
  - tangible interfaces
  - Visualisation:
    - Modality = Output graphical interaction technique
  - Augmented Virtuality / Augmented Reality:
    - Modality based on physical objects
    - Multimodality:
      - Real world (Action/Perception)
      - Digital world (Action/Perception)

77

Readings

- Multimodal Communication for the Blind
  http://www.limsi.fr/Individu/bellik/francais/meditor.htm
- Non-Visual Interfaces for Wearable Computers
  http://www.dcs.gla.ac.uk/~stephen/papers/IEE_wearables _00.pdf
- MATCH: An Architecture for Multimodal Dialogue Systems
  http://www2.research.att.com/~johnston/matchacl02.pdf
- MIPAD: A Multimodal Interaction Prototype
  http://research.microsoft.com/pubs/75373/2001-huang-icassp.pdf
- RodDirect: http://css.jaist.ac.jp/~miuramo/roddirect/index.php
- DigitalDesk: Communications of the ACM Volume 36 , Issue 7 (July 1993) Special issue on computer augmented environments: back to the real world

78

Readings

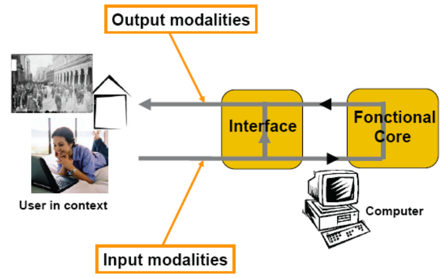
- Berkeley Institute of Design - Uni. of California at Berkeley http://bid.berkeley.edu/
- Stanford HCI group http://hci.stanford.edu/research/
- Georgia Tech http://www.gvu.gatech.edu
- MIT Media Lab http://www.media.mit.edu/research/
- Carnegie Mellon Uni. - HCI Institute http://www.hcii.cmu.edu/
- Uni. of Glasgow - Multimodal Interaction Group http://www.dcs.gla.ac.uk/~stephen/
- Microsoft Research http://research.microsoft.com/research/

79

Multimodality: Design

Underlying concepts
Design space
Software Platforms

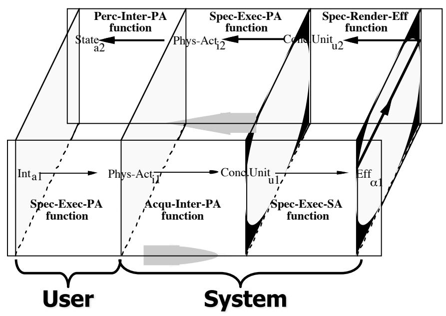
Underlying concepts



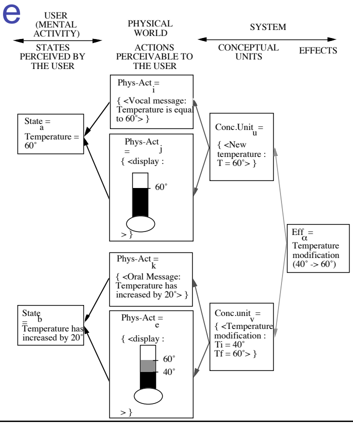
Underlying concepts
The Pipeline Model

- A data flow model:
  - user's intention -> user's physical actions
  - system's acquisition function:
    - user's physical actions -> input conceptual units
  - system's action:
    - input conceptual units -> an effect (a system state change)
  - system's rendering function:
    - effect -> output conceptual units
    - output conceptual units -> system's physical actions
  - user's perception, interpretation, evaluation
    - systems' physical actions -> new mental model

The Pipeline Model



The Pipeline Model



The Pipeline Model

- 2 concepts as point of contact between the user and the system:
  - interaction language
  - physical device
- Interaction language: set of well formed expressions used by the system or the user to exchange information
- Inter. language & phys. device = 2 facettes of an expression
  - interaction language = the structure (Hemjslev's form)
  - physical device = the observable (Hemjslev's substance)

The Pipeline Model: the utility

- A bridge between user and system perspectives using simple concepts: L&D
  - interaction language
  - physical device
- Derivation of properties that may be of interest for user modelers: the CARE props.
  - complementarity, assignment, redundancy, equivalence
- Classification of interactionally rich systems in terms of L&D: the UOM method
  - multiplicity of L&D
  - in a given state, options for the system/the user between multiple L&D
  - in a given state, usage by the system/the user of L&D
- Implication on software architectures
  - which components are L-dependent, D-dependent, etc.

7

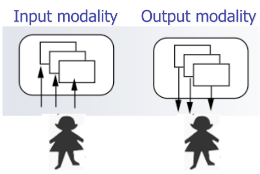
Underlying concepts Definition of a modality

- Built-in cognitive capability of the system for interpretation and rendering
- Input modality
  Interpretation function: sequence of transformations from input "raw information"
- Output modality
  Rendering function: sequence of transformations to output "raw information"

8

Definition of a modality

- Modality = (device, interaction language)
  - A set of sensors (input devices) or effectors (output devices)
  - A processing facility based on a language



9

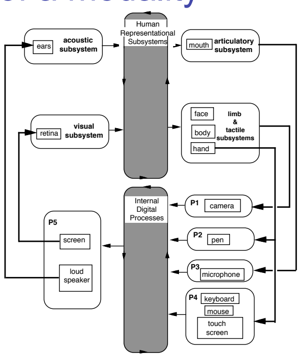
Definition of a modality

- Modality = (device, interaction language)
  - A set of sensors (input devices) or effectors (output devices)
    Perception/Action
  - A processing facility based on a language
    Cognition

10

Definition of a modality

- Theory ICS
  - APU Cambridge



11

Definition of a modality

- Modality = (device, interaction language)
- Multimodality
  - Multi device Mono Language
  - Multi device Multi Language
  - **Mono device Multi Language**
  - e.g. table and graph displayed on screen as two different modalities
    - M1 = (screen, table) and M2 = (screen, graph)

12

Definition of a modality

- Modality = (device, interaction language)
- Recent interaction paradigms such as perceptual User UI tangible UI and embodied UI open **a vast world of possibilities**
  - M1 = (microphone, natural language)
  - M2 = (keyboard, command language)
  - M3 = (mouse, direct manipulation)
  - M4 = (PDA, 3D gesture) ***embodied UI***
  - M5 = (HMD, 3D graphics) ***AR***
  - M6 = (bottle-sensor, 3D gesture) ***tangible UI***
  - M7 = (GPS, localization) ***perceptual UI***
  - *M8 = (Tongue display, 2D shape)*

13

Definition of a modality

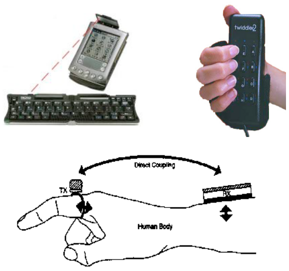
- Input Modality = <d, l>

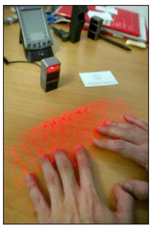
Speech = < , natural language>

14

Definition of a modality

- Input M = <device, text>

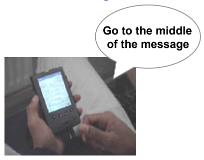




15

Definition of a modality

Input M =
<camera, Gesture>



Input M =
<microphone, NL>

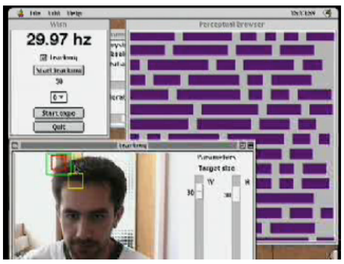
Input M = <PDA, Gesture>
Embodied modality

Input M =
<stylus, direct manipulation>

16

Definition of a modality

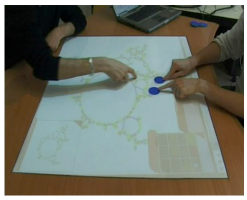
- Input M = <camera-head, gesture>



17

Definition of a modality

- Input M =
  <camera-token, direct manipulation>



18

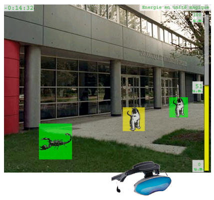
Definition of a modality

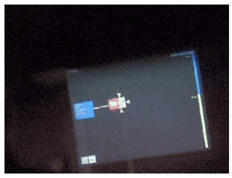
- Input M = <bottle-sensor, gesture>

19

Definition of a modality

- Input Modalities (*sensing modalities*)
- M1 = <GPS, localization>
- M2= <magnetometer, orientation>



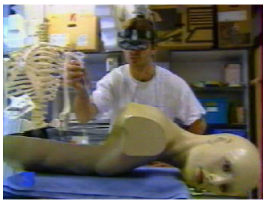


20

Definition of a modality

- OUTPUT Modality = <d, l>

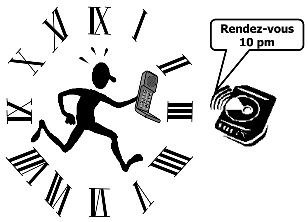
  M = <HMD, 3D graphics>



21

Definition of a modality

- Output M = <loudspeakers, NL>
- 3D sound:

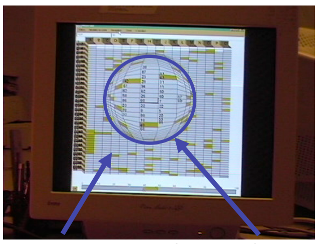


Soundbeam Neckset

22

Definition of a modality

- Complementarity of output modalities



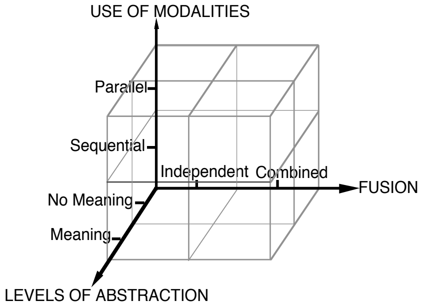
23

Underlying concepts

- Modality = (device, interaction language)
  - Input modality
    - Interpretation function: sequence of transformations from input "raw information"
  - Output modality
    - Rendering function: sequence of transformations to output "raw information"
- Four intertwined ingredients (for both):
  - 1. Levels of abstraction
  - 2. Context
  - 3. Fusion and fission
  - 4. Granularity of concurrency

24

Underlying concepts



Dimension 1: Levels of Abstraction

Expresses the variety of representations supported by the system:

Raw (digital channels)

Interpretation ↓ ↑ Rendering

High level (task domain)

Interpretation function: Ability to abstract

Rendering function: Ability to materialize

Dimension 1: Levels of Abstraction

Example: Speech input and output

Interpretation function **Ability to abstract to**	Rendering function **Ability to materialize from**
Digital signal	Symbolic representation of meaning
Word or a pattern of words	Pre-stored text message (text to speech)
Meaningful sentence	Pre-recorded vocal message

We consider two values only:
MEANING / NO MEANING

Dimension 1: Levels of Abstraction

- The capacity of abstraction may vary with the context
- Example : VI text editor
  - command mode: text is processed -> high level
  - input mode: text is recorded only -> raw
- Context of commands
  high level interpretation
- Context of task-domain data
  low level interpretation

Dimension 2: Use of Modalities

- Supported use of modalities
- Sequential:
  Use of the modalities one after another
- Parallel:
  Use of multiple modalities simultaneously
  - Multiple devices used simultaneously

Dimension 3: Fusion

- Fusion: Combination of chunks
- It occurs at multiple levels of abstraction
- Lowest level: chunks from distinct modalities
- Higher levels: chunks from dictinct contexts

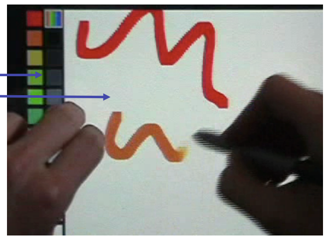
Dimension 3: Fusion

- **Lowest level: chunks from distinct modalities**
- Independent: (Absence of fusion)
  Independent interpretation/rendering process for each modality
- Combined: (Presence of fusion)
  Fusion of data expressed using different modalities
  - "Put that there" paradigm
  - => Combination of different types of data

31

Dimension 3: Fusion

- **Higher levels: chunks from dictinct contexts**
- Single input channel, multiple context
- For example:
  - Fusion of events
    - Palette
    - Drawing area



32

Multimodal versus multimedia

A multimodal system:

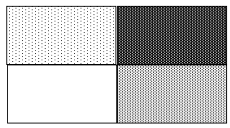
Value "Meaning" along the axis "Levels of Abstraction"

=> Four types of multimodal systems

33

Multimodal system: four types

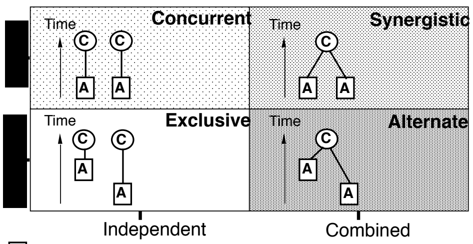
Exclusive: (Sequential, Independent)
Alternate: (Sequential, Combined)
Concurrent: (Parallel, Independent)
Synergic: (Parallel, Combined)



34

Multimodal system: four types

A multimodal system:
Value "Meaning" along the axis "Levels of Abstraction"



35

Multimodal system: four types

- Examples: www.kirusa.com

36

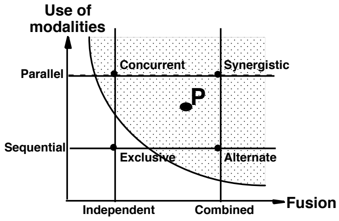
How to classify a system

To classify a given system:

1. A set of its features fi:
   fi = (pi,wi) wi: the weight
   pi: the position
2. The position P of the system is defined by:

$$P = \frac{1}{\Sigma_w} \times \sum_i pi \times wi$$

$$\Sigma_w = \sum_i wi$$



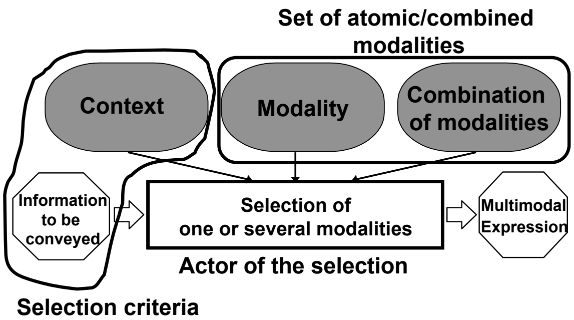
Multimodality: Design

Underlying concepts

Design space

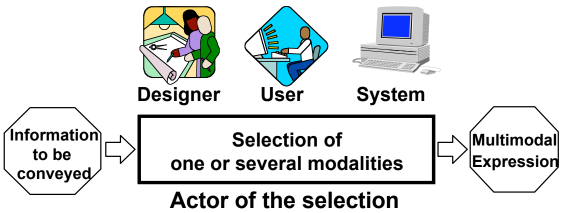
Software Platforms

Multimodality: Design space

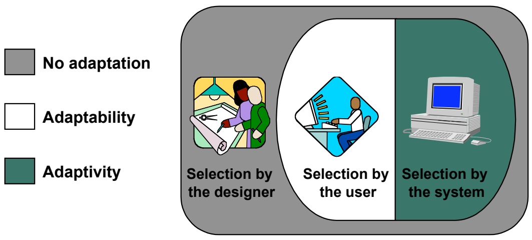


Multimodality Actor of the selection

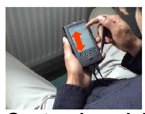
- Who is performing the selection



Multimodality Actor of the selection



Multimodality Adaptability



Gestural modality

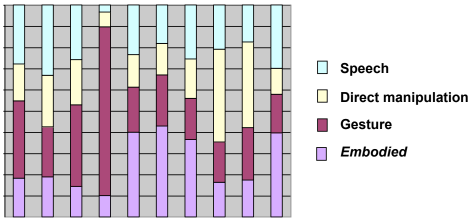
Speech

Embodied modality

Direct manipulation

Multimodality
Adaptability

- Usage of the modalities
- All sessions / All subjects



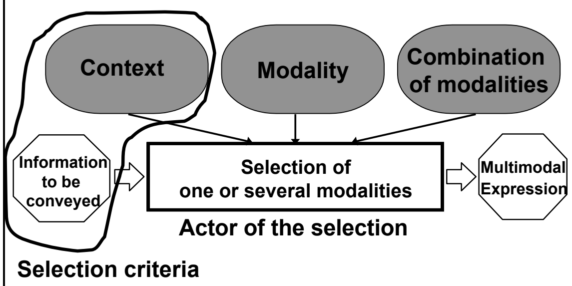
43

Multimodality
Adaptativity

- Selection of the modalities by the system
- Context-aware systems

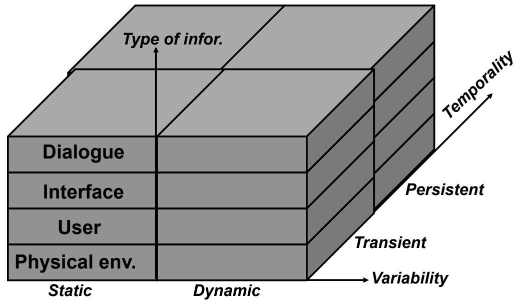
44

Multimodality: Design space



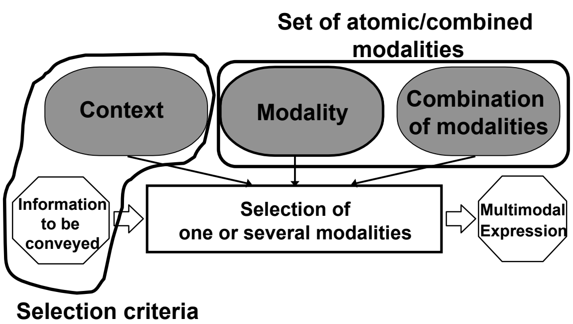
45

Multimodality
Selection criteria: Context



46

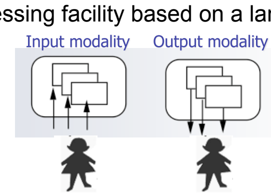
Multimodality: Design space



47

Multimodality
Characterisation of a modality

- Definition of a modality
- Modality = (device, interaction language)
  - A set of sensors (input devices) or effectors (output devices)
  - A processing facility based on a language



48

Multimodality
Characterisation of a modality

- **ACTIVE** MODALITIES
  – For inputs, active modalities are used by the user to issue a command to the computer such as a pedal to move a laparoscope in a CAS system.
- **PASSIVE - IMPLICIT** MODALITIES
  – Passive modalities are used to capture relevant information for enhancing the realization of the task, information that is not explicitly expressed by the user to the computer (PUI). For example tracking position.

49

Multimodality
Characterisation of a modality

- **Human sense**
- **Passive/Active**
- **Private / Public**
- **Spatial**
  - **Location**
- **Temporal**
  - **Transient/Persistent**

Modality

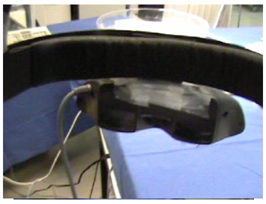
- **Dimension*: 1D 2D …***
- ***O. Bernsen 93***
  - **Linguistic**
  - **Analogue**
  - **Arbitrary**

Physical level | **Logical level**

Modality = <device, | **interaction language >**

50

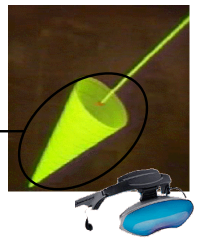
Multimodality
Characterisation of a modality



51

Multimodality
Characterisation of a modality

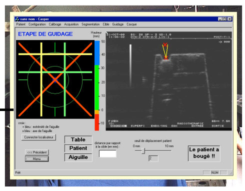
- **Physical level**
  - **Human sense: Sight**
  - **Spatial:**
    Location = operating field
  - **Temporal: Persistent**
- **Logical level**
  - **3D**
  - **Analogue**
  - **Non arbitrary**



52

Multimodality
Characterisation of a modality

- **Physical level**
  - **Human sense: Sight**
  - **Spatial: Location = screen**
  - **Temporal: Persistent**
- **Logical level**
  - **2D**
  - **Non Analogue**
  - **Arbitrary**

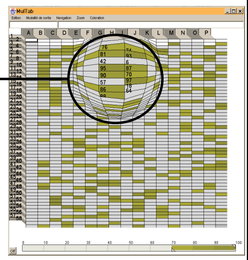


53

Multimodality
Characterisation of a modality

- Characterisation of a modality

- **Physical level**
  - **Human sense: Sight**
  - **Spatial: Location = screen**
  - **Temporal: Persistent**
- **Logical level**
  - **3D**
  - **Analogue**
  - **Non arbitrary**



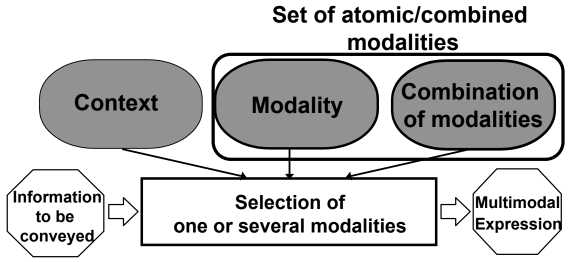
54

Multimodality
Characterisation of a modality

- Phycons as input modalities

- **Physical level**
  - **Human manipulation**
  - **Spatial: Location = desk**
  - **Temporal: Persistent**
- **Logical level**
  - **3D gesture**
  - **Analogue**
  - **Non Arbitrary**

Multimodality: Design space

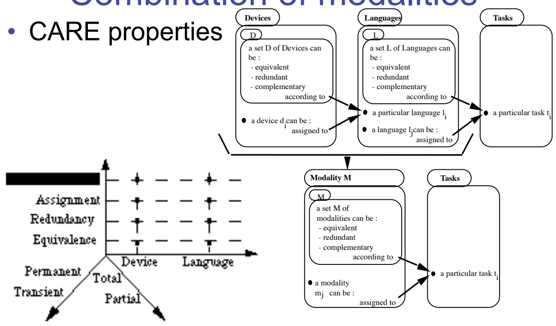


Multimodality
Combination of modalities

- Several studies
  - UOM 94 / TYCOON 95 / CARE 95
- CARE properties
  - Relationships between Devices, Interaction languages and Tasks
    - C : Complementarity
    - A : Assignment
    - R : Redundancy
    - E : Equivalence

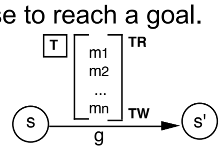
Multimodality
Combination of modalities

- CARE properties



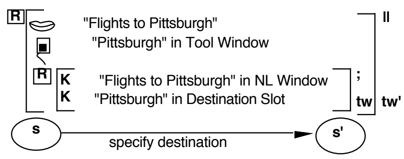
Multimodality
Combination of modalities

- CARE properties
- The formal expression of the CARE properties relies on the notions of state, goal, modality, and temporal relationships.
- A modality is an interaction method that an agent can use to reach a goal.



Multimodality
Combination of modalities

- Redundancy : Modalities of a set M are used redundantly to reach state s' from state s, if they have the same expressive power (they are equivalent) and if all of them are used within the same temporal window, tw.



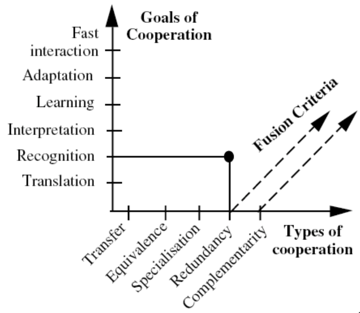
Multimodality
Combination of modalities

- Redundancy : Modalities of a set M are used redundantly to reach state s' from state s, if they have the same expressive power (they are equivalent) and if all of them are used within the same temporal window, tw.
  - Redundancy (s, M, s', tw) ⇔ Equivalence (s, M, s') $\wedge$ (Sequential (M, tw) $\vee$ Parallel (M, tw))
  - Parallel (M, tw) ⇔ (Card (M) > 1) $\wedge$ (Duration(tw) $\neq \infty$) $\wedge$ ($\exists t \in tw \cdot \forall m \in M \cdot$ Active (m, t))
  - Sequential (M, tw) ⇔ (Card (M) >1) $\wedge$ (Duration (tw)$\neq\infty$) $\wedge$ ($\forall t \in tw \cdot$ ($\forall m, m' \in M \cdot$ Active(m, t) $\Rightarrow \neg$Active(m', t)) $\wedge$ ($\forall m \in M \cdot \exists t \in tw \cdot$ Active(m, t))

Multimodality
Combination of modalities

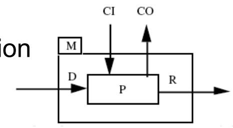
- TYCOON

*Each type of cooperation may be involved in several goals. For instance, redundancy between messages uttered and typed on the keyboard by the user may improve recognition. Only redundancy and complementarity need **fusion** which may use combination of several criteria (dotted arrows).*



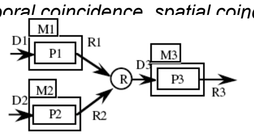
Multimodality
Combination of modalities

- TYCOON
- Logical formalism to describe the combination
- M = { P, D, R, C }
  - A process P
    - controlled by a set of parameters C (CI Input parameters CO Ouptut parameters)
    - analyzing a set of data D
    - to give a set of results R



Multimodality
Combination of modalities

- TYCOON M = { P, D, R, C }
- Redundancy
  for each possible result r3 of modality M3, the results r1 obtained by modality M1 and r2 obtained by modality M2 have been merged by an intermediate process R and have the same value for an attribute att. The criterion used by R is a parameter of the redundancy definition and may be a combination of temporal coincidence, spatial coincidence...



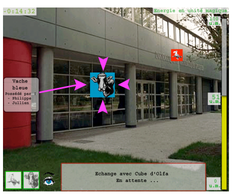
Multimodality
Combination of modalities

TROC: a game based on the technique of barter

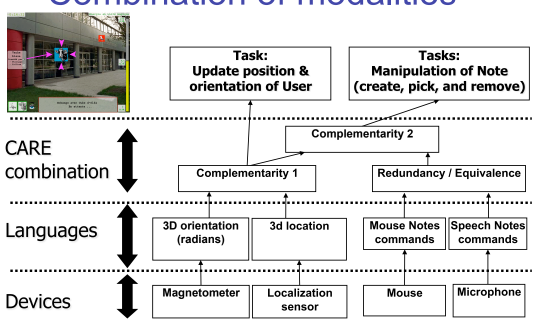
M1 = (Magnetometer, orientation)

M2 = (GPS, location)

Complementarity of M1 and M2 for selecting an object



Multimodality
Combination of modalities



Multimodality
Combination of modalities

- Several studies
  - UOM 94 / TYCOON 95 / CARE 95
- New combination space
  - Different schemas and aspects of combinations
  - 5 aspects: temporal, spatial, articulatory, syntactic and semantic
  - 5 schemas: [Allen 83]

67

Multimodality: Combination of modalities

Combination schemas

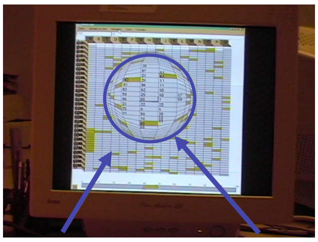
Combination aspects

Temporal	Anachronism	Sequence	Concomitance	Coincidence	Parallelism
Spatial	Separation	Adjacency	Intersection	Overlaid	Collocation
Articulatory	Independence	Fission	Fission Duplication	Partial Duplication	Total Duplication
Syntactic	Difference	Completion	Divergence	Extension	Twin
Semantic	Concurrency	Complementarity	Complementarity & Redundancy	Partial Redundancy	Total Redundancy

68

Multimodality: Combination of modalities

- Complementarity of output modalities



M1 = <screen, table> M2 = <screen, deformed table>

69

Multimodality: Combination of modalities

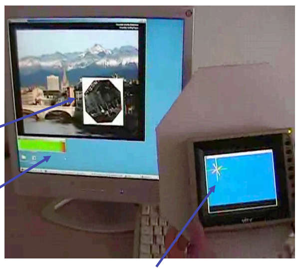
- Combination of
  M1 = <screen, table> and
  M2 = <screen, deformed table>

Temporal	Anachronism	Sequence	Concomitance	Coincidence	Parallelism
Spatial	Separation	Adjacency	Intersection	Overlaid	Collocation
Articulatory	Independence	Fission	Fission Duplication	Partial Duplication	Total Duplication
Syntactic	Difference	Completion	Divergence	Extension	Twin
Semantic	Concurrency	Complementarity	Complementarity & Redundancy	Partial Redundancy	Total Redundancy

70

Multimodality: Combination of modalities

- Puzzle



M1 = <screen, 2D image>

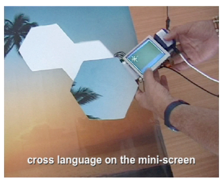
M2 = <screen, color>

M3 = <mini-screen, crosses>

71

Multimodality: Combination of modalities

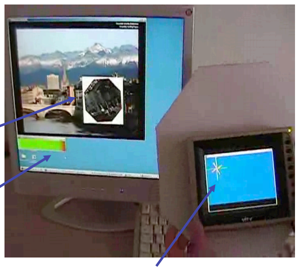
- Puzzle
  - *Video: PUZZLE-OUTPUT/Puzzle-CHI.avi*



72

Multimodality: Combination of modalities

- Puzzle



M1 = <screen, 2D image>

M2 = <screen, color>

M3 = <mini-screen, crosses>

Multimodality: Combination of modalities

- Combination of
  M2 = <screen, color> and
  M3 = <mini-screen, crosses>

Temporal	Anachronism	Sequence	Concomitance	Coincidence	Parallelism
Spatial	Separation	Adjacency	Intersection	Overlaid	Collocation
Articulatory	Independence	Fission	Fission Duplication	Partial Duplication	Total Duplication
Syntactic	Difference	Completion	Divergence	Extension	Twin
Semantic	Concurrency	Complementarity	Complementarity & Redundancy	Partial Redundancy	Total Redundancy

Multimodality: Design

Underlying concepts

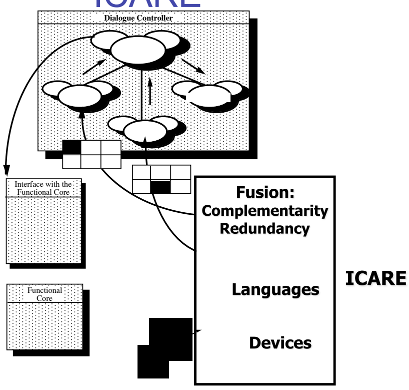
Design space

Software Platforms

ICARE

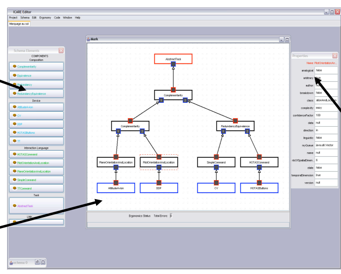
- ICARE:
- A component-based approach for the design and development of multimodal interfaces (CHI'04)
  - elementary components that describe pure modalities
  - composition components (Complementarity, Redundancy and Equivalence)
- Editor to graphically assemble components
- Automatic generation of the code (fusion mechanism)

ICARE



ICARE

Components



Properties of the selected component

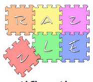
Assembly of components

ICARE

MEMO RA / PDA

Aircraft cockpit simulator

Puzzle in RA



Multimodal IDentification

79

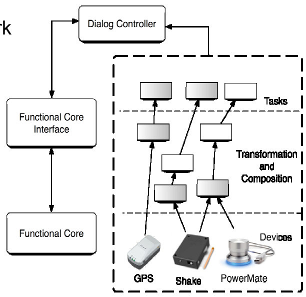
OpenInterface OI

- Open source framework for multimoal interaction
- http://www.oi-project.org/
  for download, demos and publications

80

OI Framework

- Scope of the framework



81

OI Framework

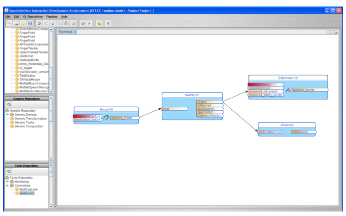
- The framework is made
  - **1) Runtime kernel: underlying platform**
  - **2) OIDE OpenInterface Interaction Development Environment**
    - Construction tool
    - Debugging / Logging tool
  - **3) Repository of interaction modalities**

82

OI Framework

- The framework is made
  - **1) Runtime kernel: underlying platform**
  - **2) OIDE OpenInterface Interaction Development Environment**

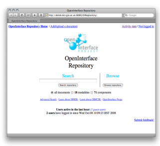
Construction tool: Assembling components



83

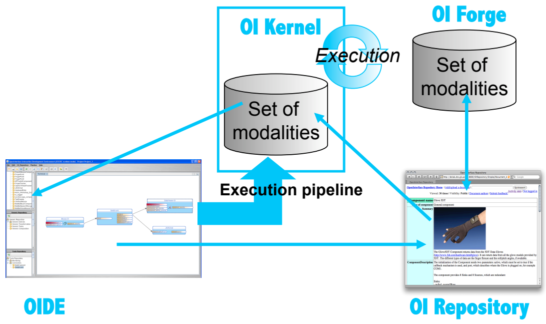
OI Framework

- The framework is made
  - **1) Runtime kernel: underlying platform**
  - **2) OIDE OpenInterface Interaction Development Environment**
  - **3) Repository of interaction modalities**

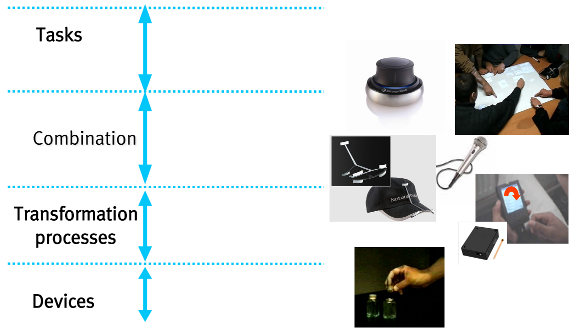


84

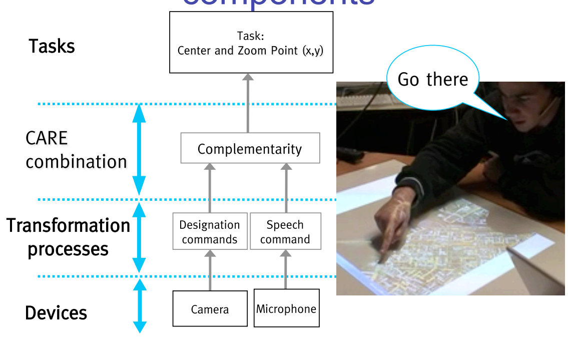
OI Framework



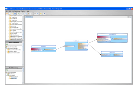
Example: Assembling components



Example: Assembling components



OI Framework



- 5 classes of components:
  - Device
  - Transformation
  - Connector
  - Debug, Data usage
  - Task

Design: Main points

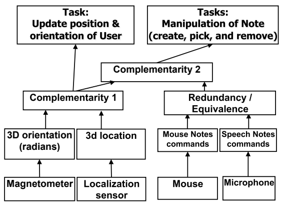
- Design space for multimodal interaction
  - Characteristics of a modality
  - Composition space
- Mapping of functionalities onto modalities not always straightforward
  - Support from guidelines and tools
  - Experimental study

Design: Main points

- Design space for multimodal interaction
  - Characteristics of a modality
  - Composition space
- For the design
  - For each task:
    - a CARE diagram with devices, languages and CARE composition
    - Characteristics of the devices and languages
    - Description of the CARE composition based on the composition space

Design: Main points

- For the design
  - For each task: a CARE diagram with devices, languages and CARE composition
- For example:



91

Readings

- Bernsen, N. Modality Theory in support of multimodal interface design. Proceedings of Intelligent Multi-Media Multi-Modal Systems, (1994), pp. 37-44
- Bouchet, J., Nigay, L., Ganille, T. ICARE Software Components for Rapidly Developing Multimodal Interfaces. Proceedings of ICMI'04, ACM Press, pp. 251-258
  http://iihm.imag.fr/publication/
- Coutaz, J., et al. Four easy pieces for assessing the usability of multimodal interaction: The CARE properties, Proceedings of Interact'95, Chapman&Hall, pp. 115-120
  http://iihm.imag.fr/publication/
- Martin, J. C. TYCOON: Theoretical Framework and Software Tools for Multimodal Interfaces. Intelligence and Multimodality in Multimedia Interfaces, AAAI Press (1997)
- Nigay, L., Coutaz, J. The CARE Properties and Their Impact on Software Design. Intelligence and Multimodality in Multimedia Interfaces, (1997) http://iihm.imag.fr/publication/
- Vernier, F,. Nigay, L. A Framework for the Combination and Characterization of Output Modalities, Proceedings of DSV-IS2000, Springer-Verlag, pp. 32-48
  http://iihm.imag.fr/publication/
- OpenInterface http://www.oi-project.org/

92