Laurence Nigay
Laurence.Nigay@univ-grenoble-alpes.fr

Modalities and Multimodalities

Domain
Definitions
Challenges

1

1

Domain and definitions

- Multimodal systems
  - Multi-Sensori-Motor Systems
  - extend the sensori-motor capabilities of computer systems

2

2

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 gesture are two common examples and are complementary

3

3

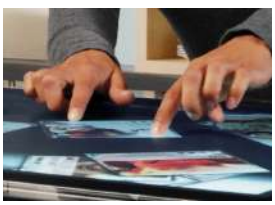
Domain and definitions

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

Multimodal ≠ Multimedia

Multimodal ≠Speech interface

New interaction capabilities appear



4

4

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/interacting (voice, gestures, facial expressions, etc.)
  - A modality is a process of receiving and producing chunks of information

5

5

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 interaction modalities**
    - For ex.: the combined input of speech and gesture

6

6

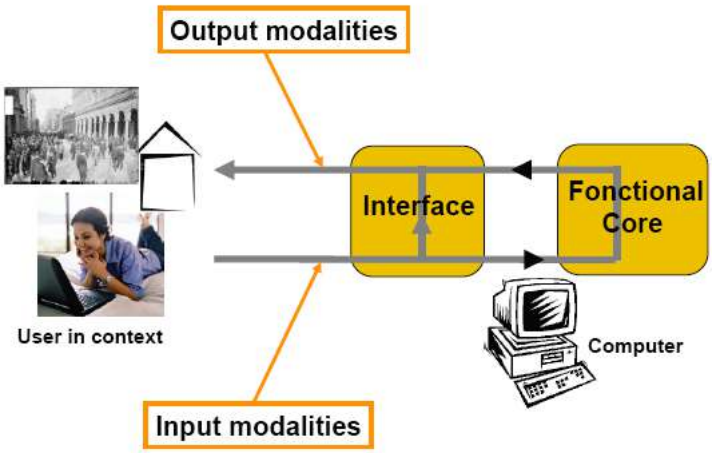
Multimodal and crossmodal

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

7

7

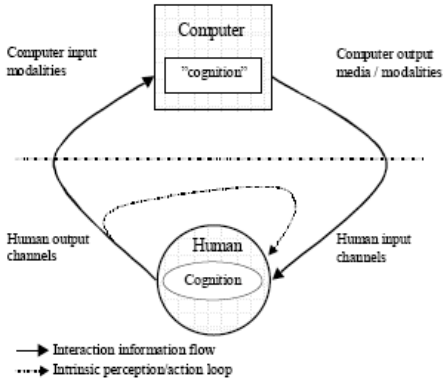
System: Input/Output modality



8

8

Multimodal interaction



9

9

Human sensory channels / System output modalities

Sensory perception	Human sense Organ	Human input Modality
Sense of sight	Eyes	Visual
Sense of hearing	Ears	Auditive
Sense of touch	Skin	Tactual
Sense of smell	Nose	Olfactory
Sense of taste	Tongue	Gustatory
Sense of balance	Organ of equilibrium	Vestibular

- Tactual
  - Tactile: Cutaneous sensitivity
  - Kinaesthetic: Awareness of movement, orientation of limbs and position
  - Haptic: combination of tactile and kinaesthetic

https://www.3dsystems.com/haptics-devices/touch

10

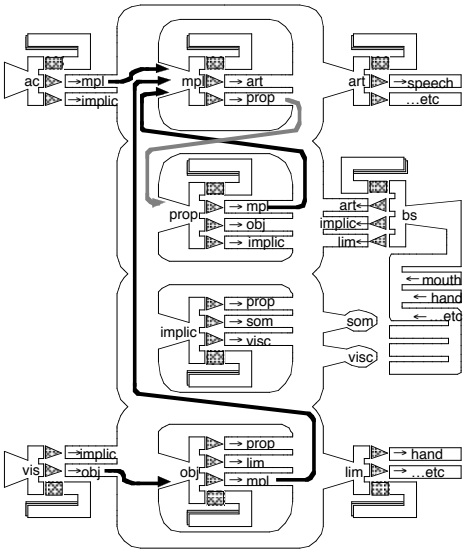
10

MacGurk effect

- Voice ba ba ba
- Lips ga ga ga
- Result: combined percept
- There are strong individual differences in terms of what is perceived: da da

Human multisensory perception

- Sight and sound
  - there can be effects of our visual perception upon the way we interpret sound
  - McGurk effect



Human output systems / System input modalities

Human motor system muscle action controlling movement	System input device
of limbs	contact or non contact sensing
hands	keyboards, pen, mouse, trackpad, etc.
eye	eye tracker
facial expression	video camera
body movement	accelerometers, magnetometers, gyrometer, etc.

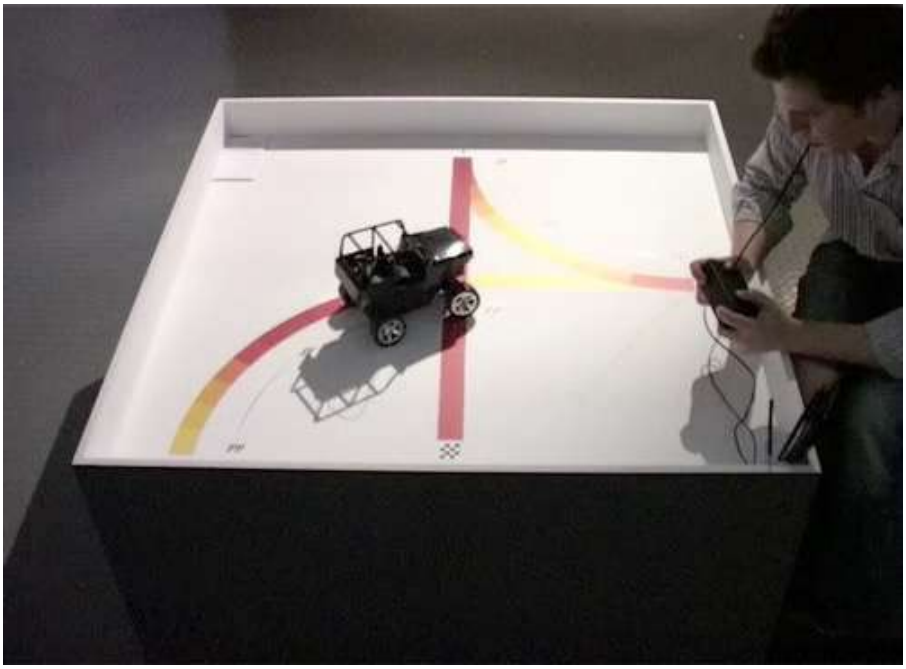
Speech Vocal utterance	microphone, speech recognition, topic recognition

Breath Pressure sensing for exhalation	Breath controllers, microphone

Bio-electric signals	EMG-signals relate to muscle activity EEG - brainwaves GSR - Galvanic skin response ECG - heart rate

13

13



14

14

Multimodal (MM) versus GUI

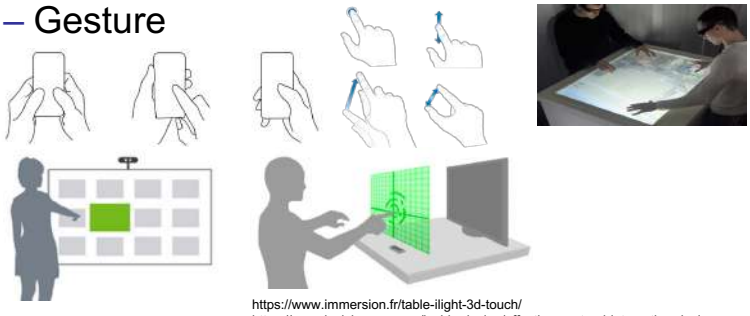
- GUI interfaces often restrict input to single non-overlapping events, while MM interfaces handle all inputs at once
- GUI events are unambiguous, MM inputs can be based on recognition

15

15

Why multimodal?

- Most technologies are mature
  - Gesture



https://www.immersion.fr/table-ilight-3d-touch/
https://www.invisionapp.com/inside-design/effective-gestural-interaction-design
https://www.clubic.com/technologies-d-avenir/article-575170-1-leap-motion-test.html

  - Speech

16

16

Why multimodal?

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

17

17

Why multimodal?

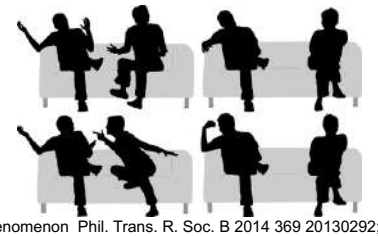
- Naturalness
  - provide more “natural” interfaces Usability
- Usability / flexibility
  - improve ease-of-use
- Robustness/Efficiency/Accuraccy
  - decrease error rates (Mutual disambiguation of recognition errors)
- Perception
- Relieve burden on the visual channel
- Support users with disabilities

18

18

Why multimodal?

- 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**



Language as a multimodal phenomenon Phil. Trans. R. Soc. B 2014 369 20130292; DOI: 10.1098/rstb.2013.0292. Published 4 August 2014

19

19

Why multimodal?

- Flexibility **for Robutness**
  - 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 **to accommodating a wide range of users, tasks, and environments**
  - Users with disability (permanent or temporary)
  - Variable usage context (**mobile support, ubiquitous computing**)

20

20

Why multimodal?

- Because of the user's circumstances – including her task, her background, her training, her knowledge, and the context– the user may well have preferences as to how she interacts 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.

21

Why multimodal?

- What do these persons have in common?

22

22

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

23

Summary: Why multimodal?

- **Natural interaction**
  - Observation: Human-to-Human interaction is intrinsically multimodal
  - Motivation: Humans should be able to communicate with machines in the same ways they communicate with one another

24

Summary: Why multimodal?

- **Flexible interaction**
  - Observation: Humans optimize their information bandwidth with the environment switching between modalities or combining multiple modalities
  - Motivation: Robust efficient multimodal interaction
    - To accommodate users with different needs and preferences (e.g. disabilities, hands-busy)
    - To improve system robustness in different contexts of use
    - To adapt to the context of use (pro-active computing)

      Example
      - Speech Recognition degrades in noisy environments
      - Use of Image based modeling of the lips can improve accuracy

25

25

Summary: Why multimodal?

- **Natural interaction**
  - Motivation: Humans should be able to communicate with machines in the same ways they communicate with one another

- **Flexible interaction**
  - Motivation: Robust efficient multimodal interaction
    - To adapt to the context of use (**pro-active computing**)

26

26

Summary: Why multimodal?

- **Natural interaction** 1
  - Motivation: Humans should be able to communicate with machines in the same ways they communicate with one another
- **Flexible interaction** 2
  - Motivation: Robust efficient multimodal interaction
    - To adapt to the context of use (**pro-active computing**)

3

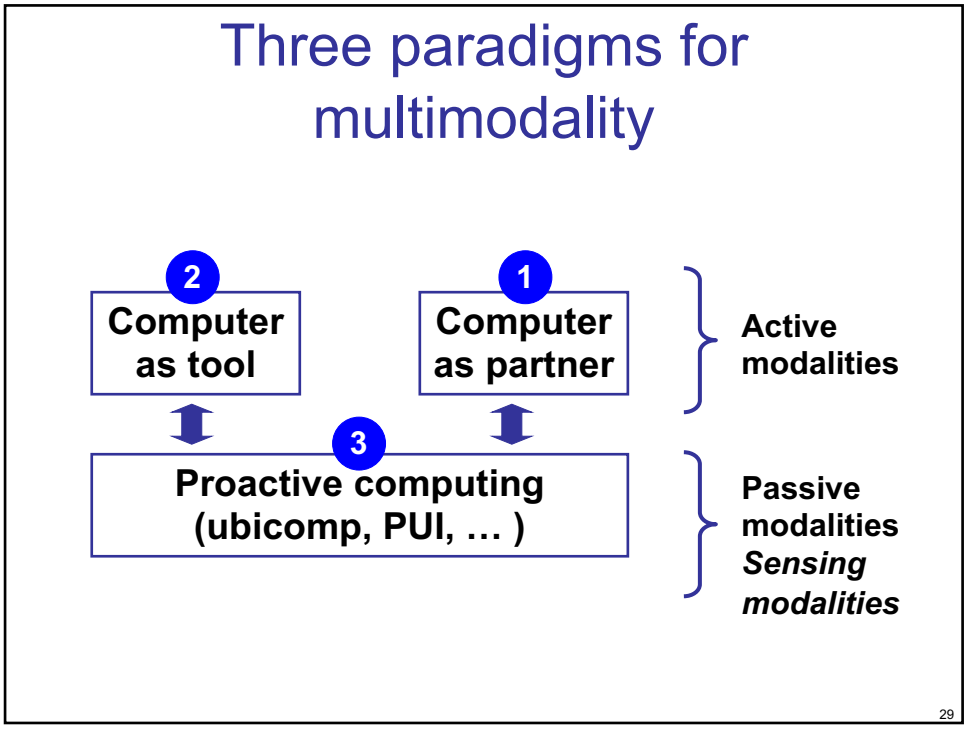
→ **Three paradigms for multimodality**

27

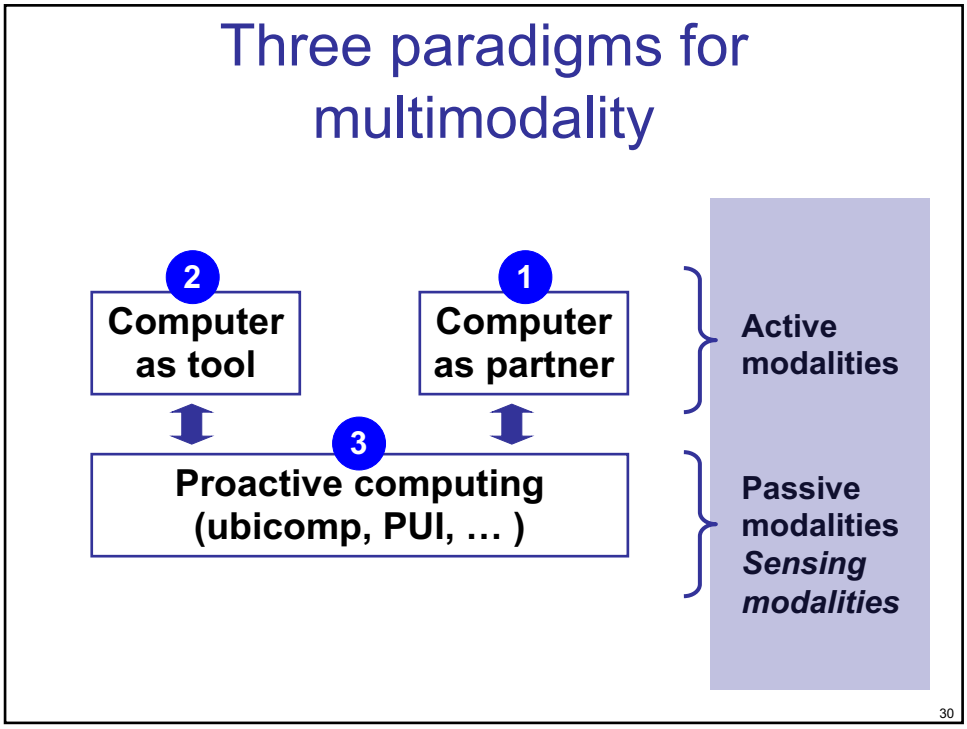
Three paradigms for multimodality

- **Computer as partner** 1
  - The multiple modalities are used to increase the anthropomorphism of the user interface
- **Computer as tool** 2
  - Multiple input modalities are used to enhance direct manipulation
- **Proactive computing (ubicomp, PUI, …)** 3
  - Multiple modalities are used to sense the context of use

28


Three paradigms for multimodality
2
Computer as tool
1
Computer as partner
Active modalities
3
Proactive computing (ubicomp, PUI, …)
Passive modalities
Sensing modalities
29

29


Three paradigms for multimodality
2
Computer as tool
1
Computer as partner
Active modalities
3
Proactive computing (ubicomp, PUI, …)
Passive modalities
Sensing modalities
30

30

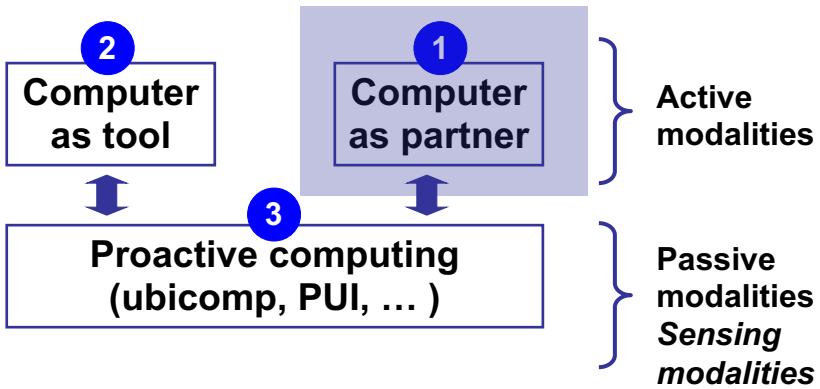
Passive/Active Modality

- **ACTIVE** MODALITIES
  - For inputs, active modalities are used by the user to issue a command to the computer
  - For example: a pedal to validate a selection 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: eye tracking, tracking position.

31

Three paradigms for multimodality



32

32

Computer as partner Multimodality

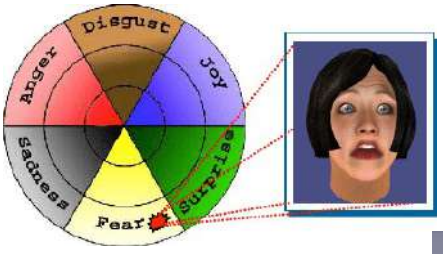
- The multiple modalities are used to increase the anthropomorphism of the user interface
  - agent based conversational user interfaces
  - multimodal output is important: talking heads and other humanlike presentation modalities
  - speech recognition is a common input modality in these systems, and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

33

33

Computer as partner Multimodality

- **Multimodal output** is important:
- talking heads and other humanlike presentation modalities



LUCIA: An Open Source 3D Expressive Avatar

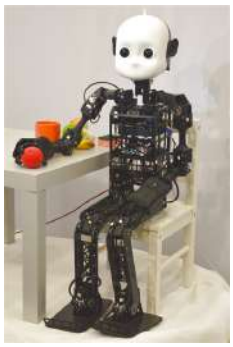
NICO – Neuro-Inspired COmpanion:
A Developmental Humanoid Robot Platform for Multimodal Interaction
http://nico.knowledge-technology.info

34

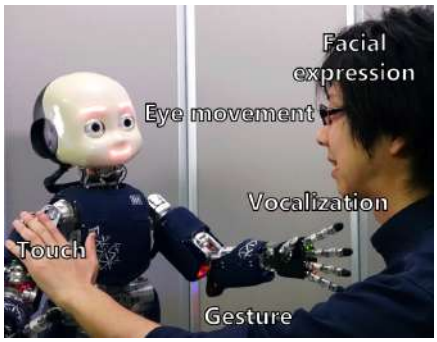
34

Computer as partner Multimodality

- **Multimodal human-robot interaction**



NICO – Neuro-Inspired COmpanion:
A Developmental Humanoid Robot Platform
for Multimodal Interaction
http://nico.knowledge-technology.info

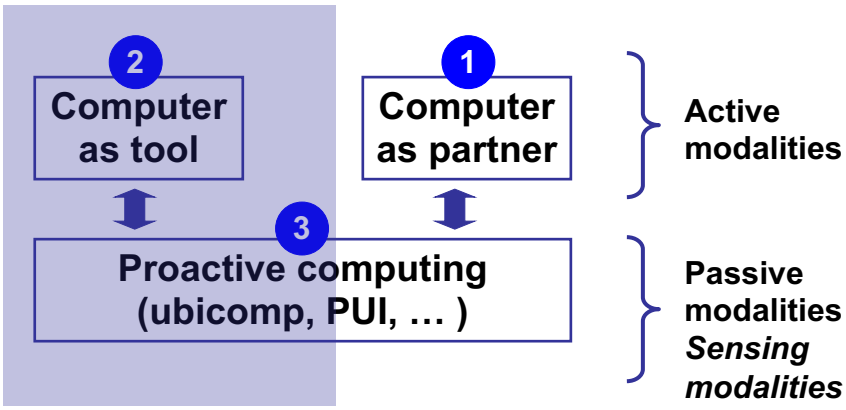


https://doi.org/10.1145/3125739.3125757

35

35

Three paradigms for multimodality



36

36

Computer as tool Multimodality

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

37

37

Multimodality: challenges

- Using multimodal input generally requires advanced recognition methods:
  - For each modality
  - For combining redundant information
  - For combining non-redundant information: “open this file (pointing)”
- Information is combined at two levels:
  - Feature level (early fusion)
  - Semantic level (late fusion)

38

38

Multimodality: challenges

- Early fusion - “feature level” fusion
  - applies to combinations like speech+lip movement
    - Speech Recognition degrades in noisy environments
    - Use of Image based modeling of the lips can improve accuracy

  - Difficult because:
    - Of the need for MM training data
    - data need to be closely synchronized
    - Computational and training costs

39

39

Multimodality: challenges

- Late fusion - “semantic level” fusion”
- for combinations of complementary information, like gesture+speech.
  - Recognizers are trained and used separately
  - Unimodal recognizers can be available off-the-shelf
  - It is still important to accurately time-stamp all inputs: typical delays are known between e.g. gesture and speech

40

40

Multimodality: challenges

- Challenge 1: Theory of modality and multimodality
  - a vast world of possibilities
    => Characterization of the modalities
- Challenge 2: Fusion mechanism
  - Criteria for triggering the fusion: time and ? … space
  - Ambiguity and the fusion mechanism
  - Uncertainty of the data processed by the fusion mechanism
- Challenge 3: Pervasive computing
  - Dynamicity
    => Plugging at runtime new modalities to the fusion mechanism
- Challenge 4: Development tools
  - Tools for quickly developing multimodal interaction
    - ICARE , OPENINTERFACE, context-toolkit for passive-implicit modalities, QUICKSET…

41

41

Multimodality: Path to evolution

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

42

42

43

43

Multimodality: Path to evolution

- "Put that there" paradigm

„Zoom in here"

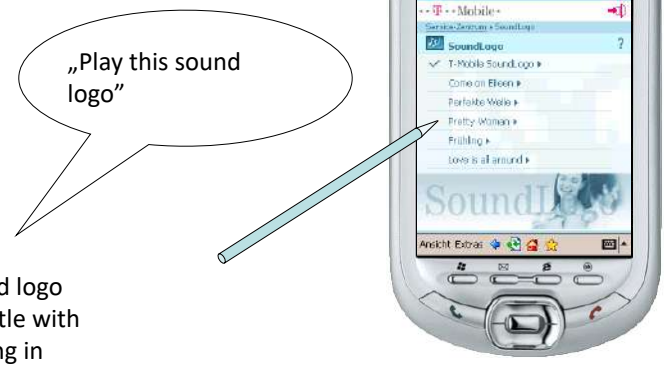
User selects a point of interest clicking with a stylus and speaking in order to focus it.

44

44

Multimodality: Path to evolution

- "Put that there" paradigm



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

45

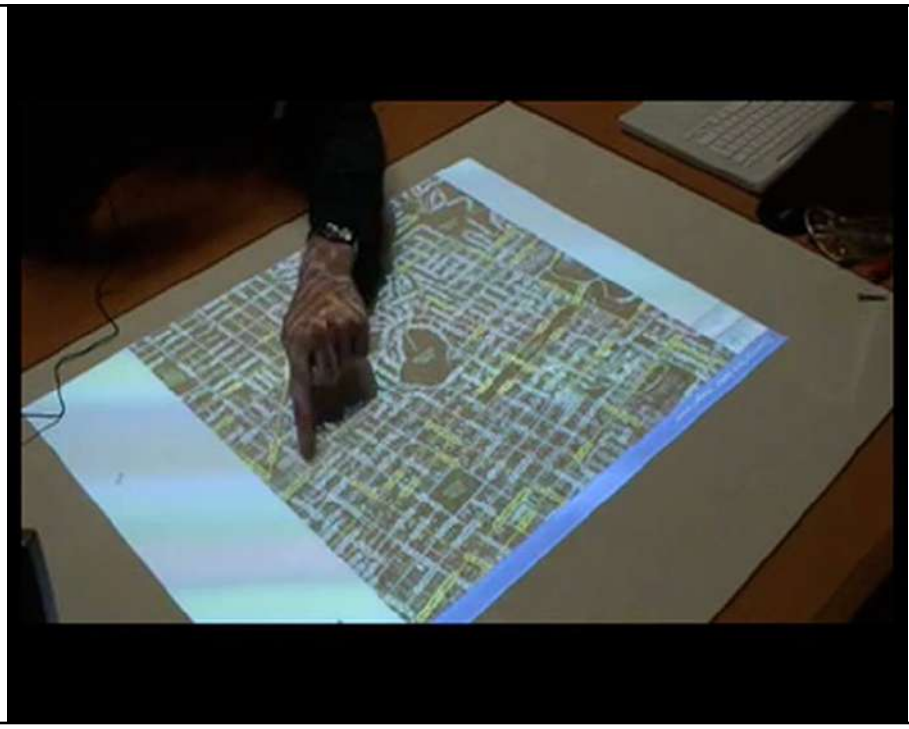
45

Multimodality: Path to evolution

- "Put that there" paradigm

46

46

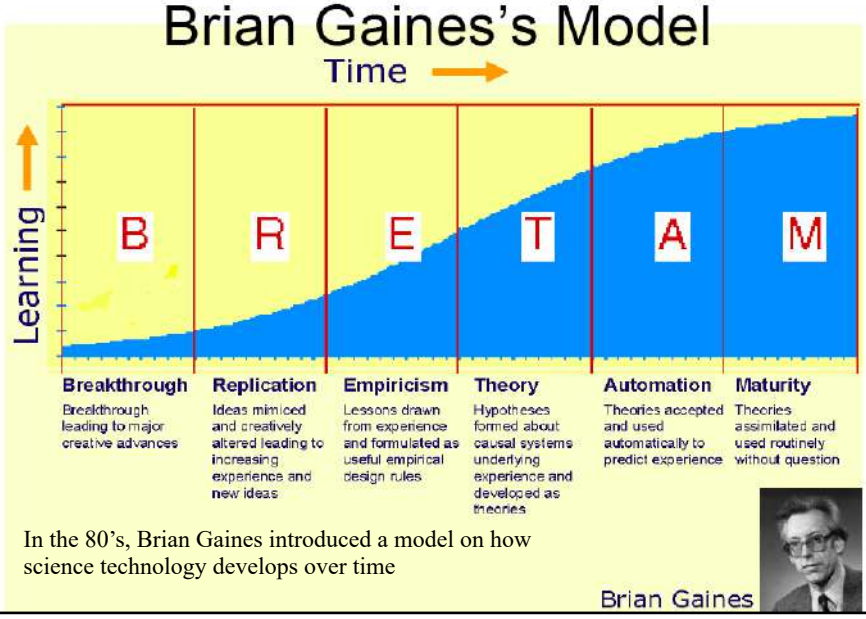


47

47

Multimodality: Path to evolution

Brian Gaines's Model



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

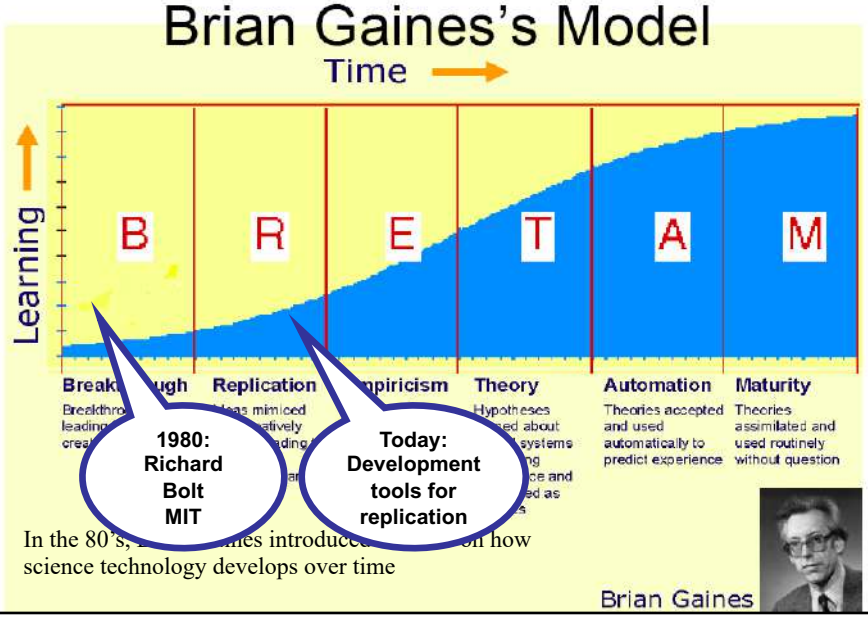
Brian Gaines

48

48

Multimodality: Path to evolution

Brian Gaines's Model



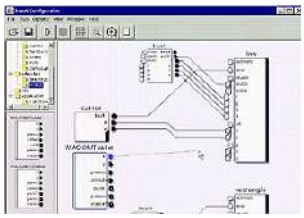
In the 80's, [illegible] introduced [illegible] on how science technology develops over time

49

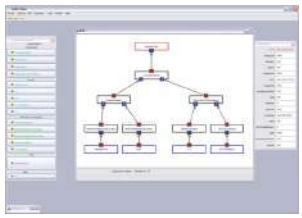
49

Multimodal platforms

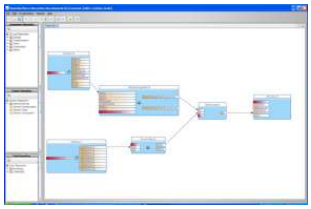
- Mediation chains from devices to commands



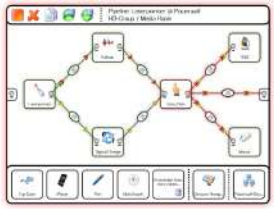
ICON (Dragicevic, 2001)



ICARE (Bouchet, 2004)



OI (Serrano, 2008)

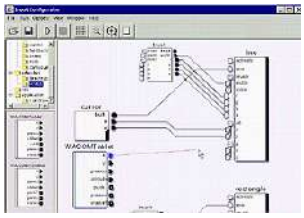


SQUIDY (König, 2010)

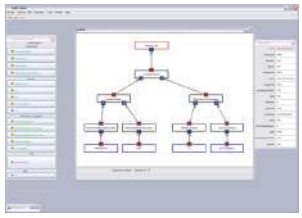
50

Multimodal platforms

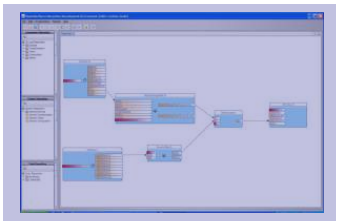
- Mediation chains from devices to commands



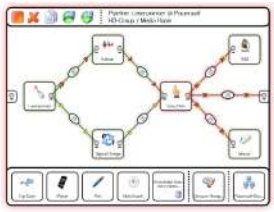
ICON (Dragicevic, 2001)



ICARE (Bouchet, 2004)



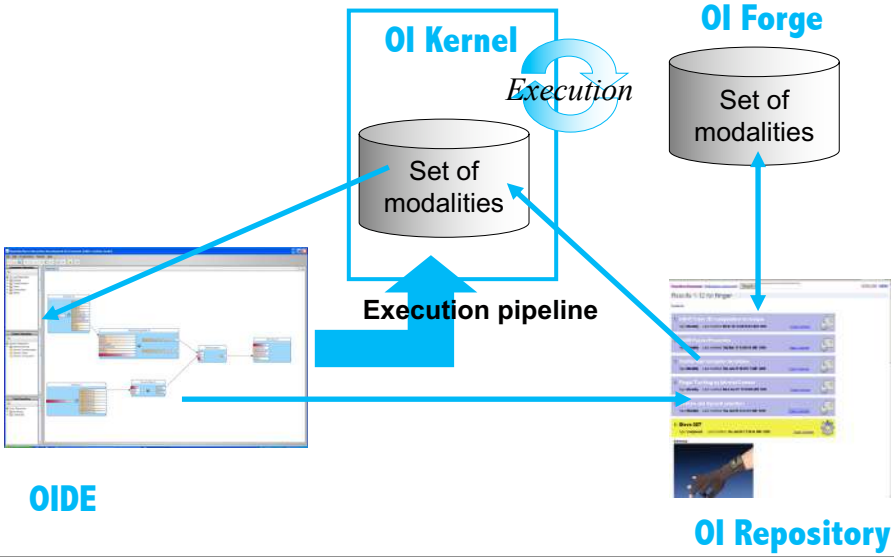
OI (Serrano, 2008)



SQUIDY (König, 2010)

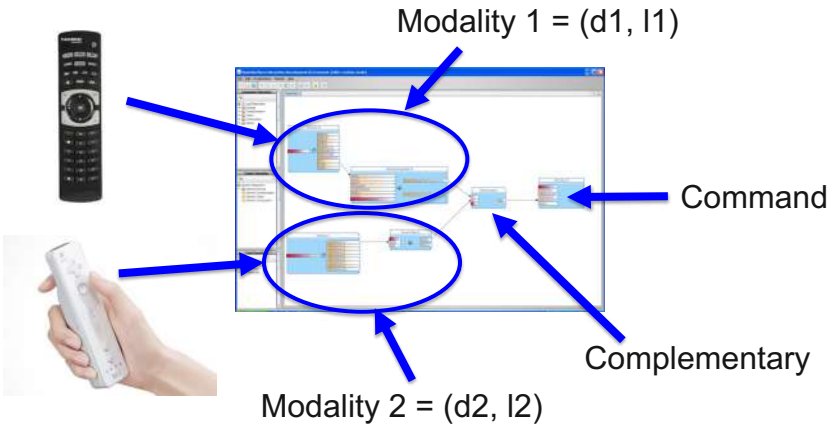
51

OpenInterface platform



52

OpenInterface platform



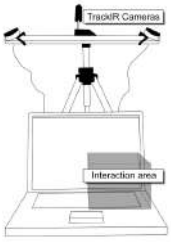
53

Components - Modalities

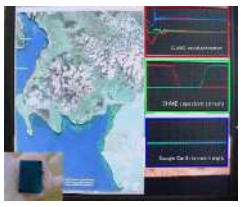
- 82 components
  including 30 devices

CubeTile
(immersion)

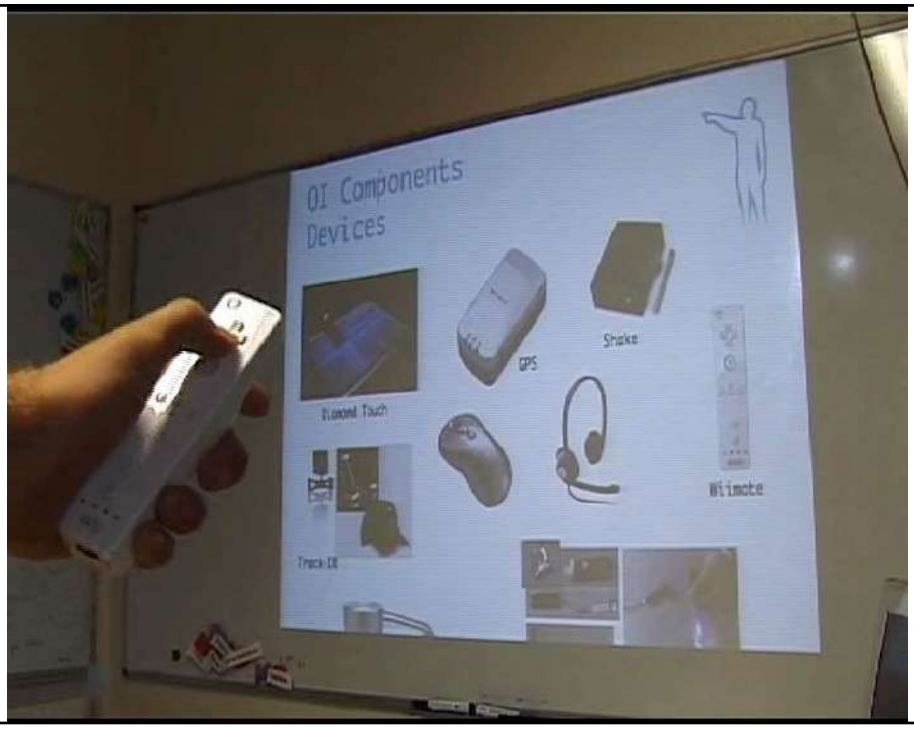
Jack and Stane
(University of Glasgow)



AirMouse
(Laboratory of Grenoble)



54



55

55

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

56

56

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

57