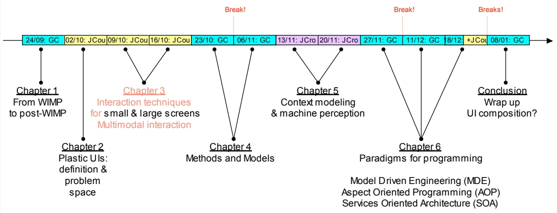
Mobile and Context-aware Interactive Systems

Gaëlle Calvary, Joëlle Coutaz and James Crowley

Université Joseph Fourier (Grenoble I)
ENSIMAG

Laboratoire d'Informatique de Grenoble (LIG)

Outline and schedule



GC: Gaëlle Calvary
JCou: Joëlle Coutaz
JCro: James Crowley

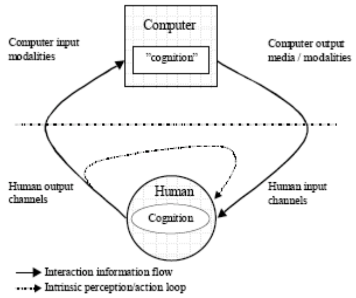
2

Outline of Chapter 3 - part 3 (multimodal interaction)

1. Introduction: motivation
2. Multimodality from the human perspective
3. Multimodality from the system perspective
4. Design guidelines

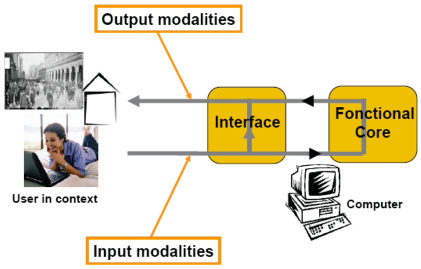
3

Multimodal interaction



4

System input modality & system output modality



5

Human input modalities (system output devices)

Sensory perception	Human sense Organ	Human input Modality	System output device
Sense of sight	Eyes	Visual	Screen
Sense of hearing	Ears	Auditive	Loud speaker
Sense of touch	Skin	Tactual	Braille device, haptic device
Sense of smell	Nose	Olfactory	Olfactory displays (whifers)
Sense of taste	Tongue	Gustatory	
Sense of balance	Organ of equilibrium	Vestibular	motored devices

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

The Phantom (haptic device)

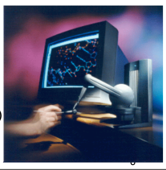


Figure 1a. Solenoid-activated perfume bottles

6

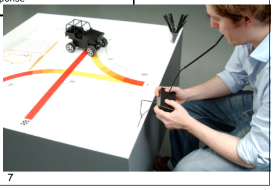
Human output modalities (system input devices)

Human motor system	System input device
muscle action controlling movement 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



http://www.irvinebrown.com/breathcar.html

7

Motivation for multimodal interaction

- Observation 1: human-to-human interaction is intrinsically multimodal
- Motivation 1: natural interaction
  - Humans should be able to communicate with machines in the same ways they communicate with one another

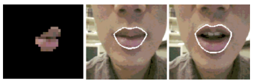
System as a tool:
"Put that there" paradigm [Bolt 80, MIT]

System as a partner:
conversational agents
"Talking heads"

8

Motivation for multimodal interaction

- Observation 2: humans optimize their information bandwidth with the environment switching between modalities or combining multiple modalities
- Motivation 2: robust and flexible 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, plastic UI)



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

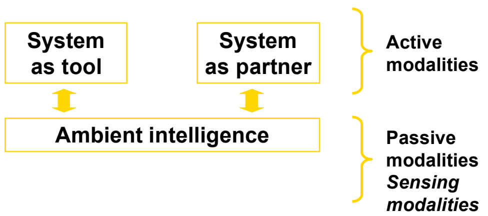
9

In short: three paradigms for multimodality

- System as a tool
  - Multiple input modalities are used to enhance direct manipulation
- System as a partner
  - Multiple modalities are used to increase the anthropomorphism of the user interface
- Ambient intelligence - machine perception (chapter 5)
  - Multiple modalities are used to sense the context of use
  - Modalities are exploited to adapt to the variation of the context of use

10

Three paradigms for multimodality



11

Outline of Chapter 3 - part 3 (multimodal interaction)

1. Introduction: motivation
2. Multimodality from the human perspective
3. Multimodality from the system perspective
4. Design guidelines

12

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

13

Human perception is **multisensory**

- Humans have several different senses through which information about the environment is obtained
- Each sense is assigned to a specific form and range of energy so that we can sense the different aspects of the environment
- No information processing system is powerful enough to perceive and act accurately under all conditions
- If a single modality is not enough to come up with a robust estimate, information from several modalities are combined
- Humans combine information following two general strategies:
  - by maximizing information delivered from the different sensory modalities and thus overcoming a specific sensory deprivation (sensory combination)
  - by increasing the reliability of the sensory estimates (sensory integration)
- This applies to both **biological and technical systems**

From Jukka Raisamo and Roope Raisamo, Tampere Univ.

14

Human sensory **combination**

- The human brain reconstructs the environment from the incoming streams of – often ambiguous – sensory information and generates unambiguous interpretations of the world
- To do so many different sources of sensory information are constantly processed, analyzed and **combined**

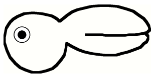
From Jukka Raisamo and Roope Raisamo, Tampere Univ.

15

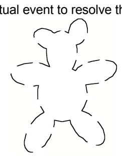
Human sensory **combination**

- Examples

- The moving train illusion
  - Is it your train or the other train that is moving?
  - The brain collects additional information about the perceptual event to resolve the ambiguity
- The incomplete figure (ICS theory)
- The ambiguous figure (ICS theory)



Propositional representations help the object representation settle on one interpretation of ambiguous figures



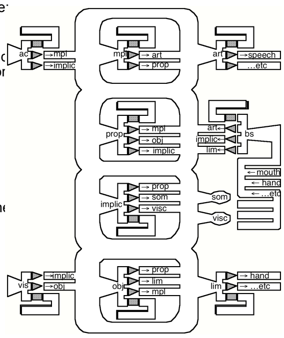
Propositional knowledge helps to 'complete' the object representation of this figure

From Jukka Raisamo and Roope Raisamo, Tampere Univ

ICS: Interacting Cognitive Subsystems [Barnard&May 93]

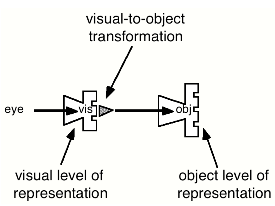
ICS is a general cognitive architecture

- It models the flow of information through different mental representations from sensation anc perception, through comprehensior to action
- It identifies cognitive aspects such as the influences of experience, memory requirements, and the potential for learning
- The architecture also constrains thi way that different sensory representations (i.e., the user's 'input' modalities) and effector representations (i.e., their 'output' modalities) can be combined.



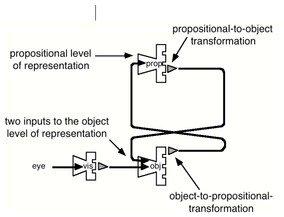
ICS: Visual perception

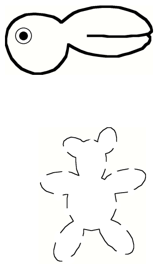
- Sensory information is transformed from a visual level of representation into an object level



ICS: Visual perception

- Exchange of representations between the object and propositional levels





Human sensory **integration**

- Often there is more than one sensory estimate available for perceiving an environmental property
- Example: judging an object size
  - Both the visual and haptic sensory modalities provide information
  - But what is the perceived size of an object that is simultaneously seen and touched?
  - The one determined by the visual estimate, the one determined by the haptic estimate, or something in between?
- Information from the different sensory modalities has to be integrated in order to form a coherent **multisensory percept**
- Sensory integration makes the resulting estimate more reliable

From Jukka Raisamo and Roope Raisamo, Tampere Univ.

Human sensory **integration**

- Visual dominance in sensory integration
  - Tactile information can be altered by visual information
    - For example, if the visual shape of an object differs considerably from its tactual shape (cf. Rock & Victor's experience, next slide)
  - The spatial location of a sound source can be drastically influenced by visual stimulation
    - For example, in television the voices are perceived to originate from the actors on the screen
  - Vision may alter speech perception
    - For example, McGurk effect: subject watch a video where:
      - A person actually utters "ba-ba- ba"
      - But the lips of that person moves as if the person were saying "ga-ga-ga"
      - In general, subjects report hearing the sound "da-da-da"

From Jukka Raisamo and Roope Raisamo, Tampere Univ.

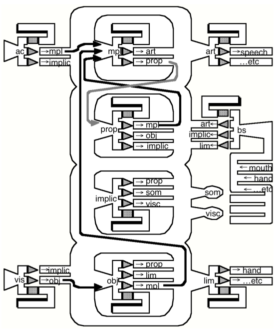
Human sensory **integration**

- Rock and Victor experience: integration of visual and haptic information
  - They asked subjects to report the perceived size of an object simultaneously seen and felt
  - Subjects looked at the object through a cylinder lense that makes a square look like a rectangle which thus created a conflict between visual and haptic information
  - Vision dominated the integrated percept
  - However, there was also always a small but consistent influence of touch on the integrated percept
  - This phenomenon of visual dominance was subsequently called "**visual capture**"

 From Jukka Raisamo and Roope Raisamo, Tampere Univ

ICS: Integrating sight and sound

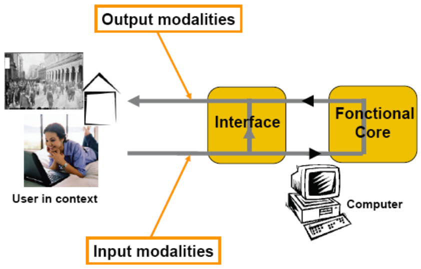
- Sight and sound
  - there can be effects of our visual perception upon the way we interpret sound
  - McGurk effect



Outline of Chapter 3 - part 3 (multimodal interaction)

1. Introduction: motivation
2. Multimodality from the human perspective
3. Multimodality from the system perspective
4. Design guidelines

System perspective: underlying concepts

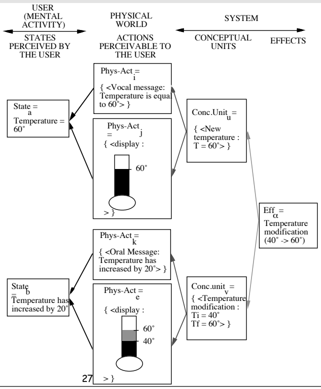


System perspective: underlying concepts

- 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

System perspective: underlying concepts

- Example: from system conceptual unit to system physical action
- The room temperature has changed



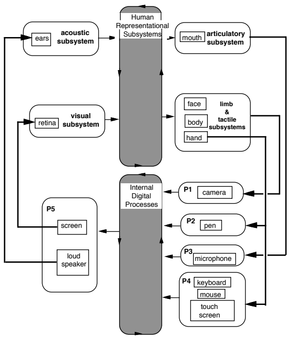
System perspective: underlying concepts

- 2 concepts as point of contact between the user and the system:
  - physical device: d
  - interaction language: l
- Physical device: a peripheral (transducer) accessible to both the user and the system that converts information into stimuli and vice versa
  - Input device: transducer (sensor) that converts (human) energy into digital units of information "processable" by the system
  - Output device: transducer (actuator) that converts digital units of information into stimuli perceivable by a human
- Interaction language: a set of well-formed expressions composed from units of information used by the system or the user to express to convey the structure (syntax) and the semantics of information

28

System perspective: underlying concepts

- Interaction language and device as 2 points of contact between the system and the user
- Physical device: perception/action = physical level of interaction
- Interaction language: cognition (representational cognitive/processing subsystems) = logical level of interaction



29

System perspective: modality definition

- Modality = <device, interaction language>

Speech = < , natural language>

30

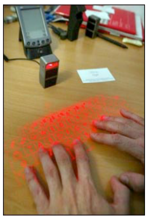
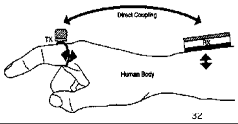
System perspective: modality definition

- Modality = <device, interaction language>
- 2 input modalities to support the same human tasks
  - <keyboard, command language>
  - <microphone, command language>
- 2 output modalities to represent the same data
  - <screen, table>
  - <screen, graph>

System perspective: modality definition

- Input M = <keyboard-device, text>



System perspective: modality definition

Input M =
<tactile screen, Gesture>

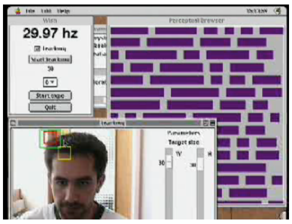
Input M =
<microphone, NL>

Input M = <PDA, Gesture>
Embodied modality

Input M =
<stylus, direct manipulation>

System perspective: modality definition

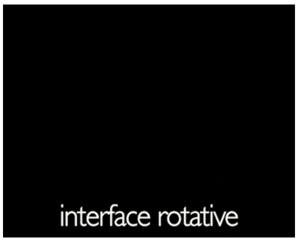
- Input M = <camera, head movement>
- Ouput M = <screen, 2D graphics>



34

System perspective: modality definition

- Input M = <camera-token, direct manipulation>
- Ouput M = <video-projector+table, 2D graphics>



35

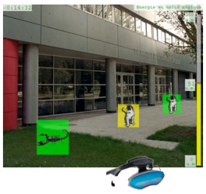
System perspective: modality definition

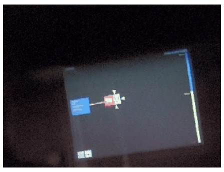
- Input M = <bottle-sensor, grab gesture>
- Output M = <loud speaker, analogical sound>

36

System perspective: modality definition

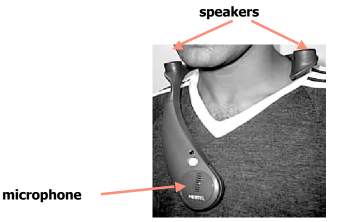
- Input Modalities (*sensing modalities*)
  - M1 = <GPS, localization>
  - M2 = <magnetometer, orientation>
- Output modality
  - M = <Head-Mounted Display, 3D graphics>





System perspective: modality definition

- Input M = <microphone on the chest, NL>
- Output M = <2 directional speakers on the shoulders, spatialized audio>

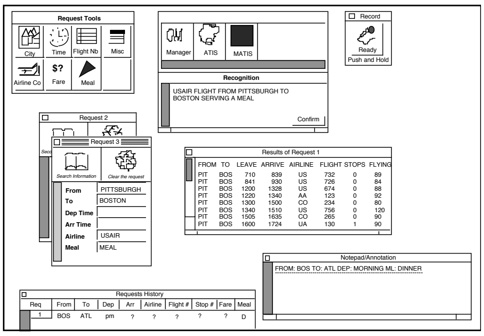


The MIT Soundbeam Neckset

System perspective: multi-modality

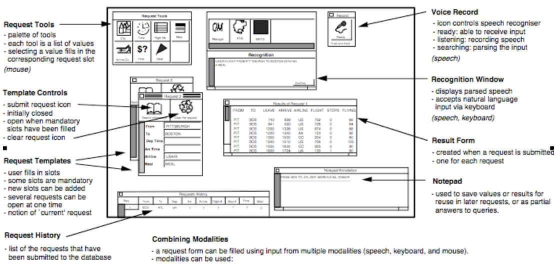
- Mono device - mono language
- Mono device - multi language
- Multi device - mono language
- Multi device - multi language
- Plus different ways of using them: the CARE properties

Running example: MATIS [L. Nigay]



MATIS: Multi-modal Airline Travel Information System *(LGI Grenoble & CMU)*

- form-based interface to a database of information about airline flights between cities.

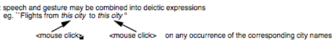


Combining Modalities

- a request form can be filled using input from multiple modalities (speech, keyboard, and mouse).
- modalities can be used:

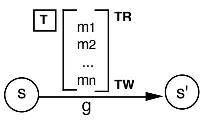
- sequentially: eg. use speech to specify values for some slots, then others are filled using request tools and mouse
  eg. use speech to identify slot, then mouse or keyboard to provide the value

- concurrently: speech and gesture may be combined into deictic expressions
  eg. ``Flights from *this city* to *this city*''



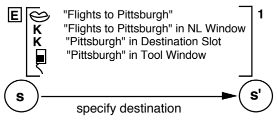
CARE properties

- The formal expression of the CARE properties relies on the notions of state, goal, modality, and temporal relationships
  - Modality mi= <d, l> = an interaction method that an agent (user, system) can use to reach a goal.
  - Goal g = a state s' that an agent intends to reach from s using modalities m1, mn
  - TR = temporal relationships between the use of modalities m1, .. Mn (parallelism, sequentiality, cardinality) = // | ; | 1
  - TW = temporal window within which the modalities are used
  - T = C | A | R | E = Complementarity, Assignment, Redundancy, Equivalence



CARE properties: Functional Equivalence

- Modalities of set M are functionaly *equivalent* for reaching s' from s, if it is necessary and sufficient to use any *one* of the modalities. M is assumed to contain at least two modalities.
- $Equivalence\ (s, M, s') \Leftrightarrow (Card(M) > 1) \wedge (\forall m \in M\ Reach\ (s, m, s'))$
- *Reach (s, m, s') : state s' can be reached from s using Modality m*

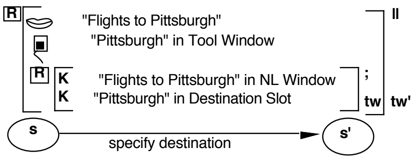


CARE properties: Assignment

- Modality m is *assigned* in state s to reach s', if no other modality is used to reach s' from s
- In contrast to equivalence, assignment expresses the absence of choice:
  - either there is no choice at all to get from one state to another,
  - or there is a choice but the agent always opts for the same modality to get between these two states.
- Thus we can define two types of assignment:
  - $StrictAssignment\ (s, m, s') \Leftrightarrow Reach\ (s, m, s') \wedge (\forall m' \in M.\ Reach(s, m', s') \Rightarrow m' = m)$
  - $AgentAssignment\ (s, m, M, s') \Leftrightarrow (Card(M) > 1) \wedge (\forall m' \in M.\ (Reach\ (s, m', s') \wedge (Pick\ (s, m', s')) \Rightarrow m' = m)\ )$
  - $Pick(s, m, s')$ predicate that expresses the use of *m* among a set of modalities to reach *s'* from *s*

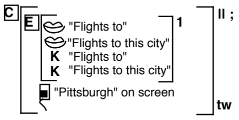
CARE properties: Redundancy

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

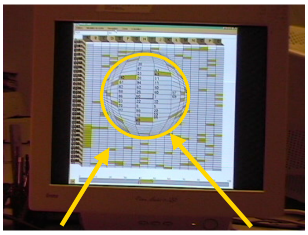


CARE properties: Complementarity

- Modalities of a set M must be used in a *complementary* way to reach state s' from state s within a temporal window, if all of them must be used to reach s' from s: none of them taken individually can cover the target state.
- *Complementarity (s, M, s', tw)*
  $\Leftrightarrow (Card(M) > 1) \wedge (Duration(tw) \neq \infty) \wedge (\forall M' \in \mathbf{P}M\ (M' \neq M \Rightarrow \neg REACH\ (s, M', s'))) \wedge REACH\ (s, M, s') \wedge (Sequential\ (M, tw) \vee Parallel\ (M, tw))$
- REACH(s,M,s') means that state s' can be reached from state s using the modalities in set M.



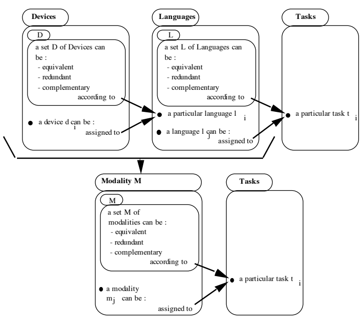
Complementarity of two output modalities



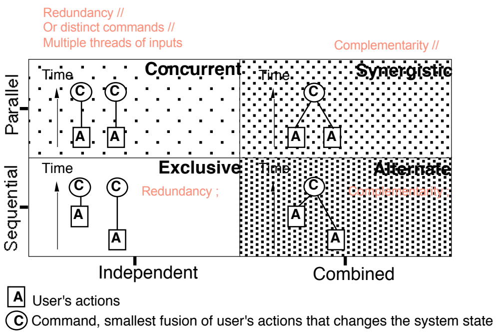
Output M1 = <screen, table>

Output M2 = <screen, deformed table>

CARE properties: in short

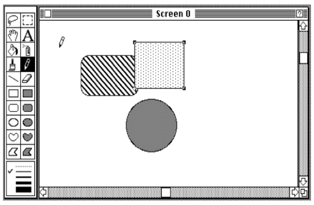


Another way of characterizing multimodal interactive systems



Concurrent multimodal UI

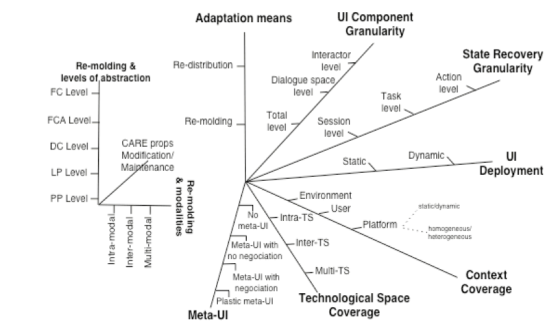
- As you draw, you can talk to change the thickness or the color of the pen



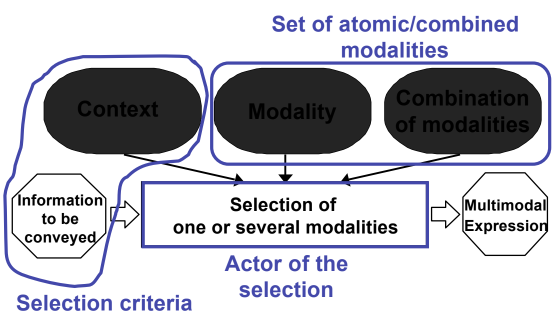
Exercise

- Go to http://www.kirusa.com/multimodality.html
- Characterize the system in terms of the CARE properties

Back to the problem space of plastic UI

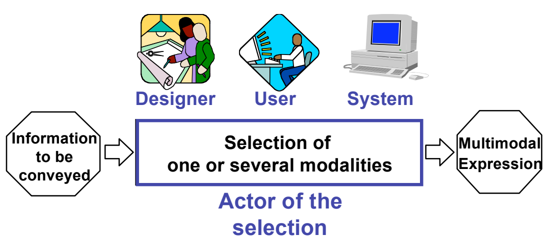


52



53

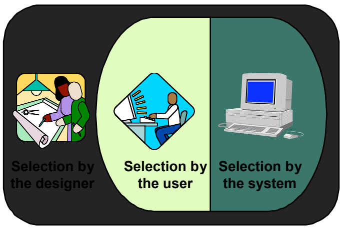
Which actor performs modalities selection ?



54

Multi-modality selection

- No adaptation
- Adaptability
- Adaptivity



55

Outline of Chapter 3 - part 3 (multimodal interaction)

1. Introduction: motivation
2. Multimodality from the human perspective
3. Multimodality from the system perspective
4. Design guidelines

56

Speech input/output

- Spoken communication with machines (both input and output) may be advantageous:
  - when the user's hands or eyes are busy
  - when only limited keyboard and/or screen is available
  - when the user is disabled
  - when pronunciation is the subject matter of computer use
  - when natural language interaction is preferred

57

Speech output

- Speech **output** is preferable when the
  - message is short.
  - message will not be referred to later.
  - messages deal with events in time.
  - message requires an immediate response.
  - visual channels of communication are overloaded.
  - environment is too brightly lit, too poorly lit, subject to severe vibration, or otherwise unsuitable for transmission of visual information.
  - user must be free to move around.

58

Pen input

- Multifunctionality (text, digits, pointing, gestural marks, symbols,graphics, sketching & art, signatures, direct manipulation, etc.)
- Visual feedback, permanent record
- Preferred for spatial & graphic tasks, selection of objects, numeric & symbolic data, & signatures

59

Pen input

- Precise spatial input (compared with speech, or even manual gesturing & touch)
- Easier for some populations (young children)
- Easy portability
- Direct input

60

Eye-gaze

- 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*
- *Eye-gaze applications: self-care applications for severely-impaired users (e.g., quadriplegics)*

61

Eye-gaze

- Eye-gaze patterns: wrong assumptions
  - Users eyes stop to look at things
  - Users look at things intentionally
  - What users are looking at is an indication of what they're thinking
  - The eyes and hands manipulate things simultaneously
  - Eye trackers track eye movements reliably

- Gaze isn't a good mouse replacement!

62

Multimodal input/output

- Designing multimodal input and output
  - Match output to acceptable user input style
    - if the user is constrained by a set grammar, do not design a virtual agent to use unconstrained natural language

- Adaptivity
  - Multimodal interfaces should adapt to the needs and abilities of different users, as well as different contexts of use. Dynamic adaptivity enables the interface to degrade gracefully by leveraging complementary and supplementary modalities according to changes in task and context.
    - Allowing gestures to augment or replace speech input in noisy environments, or for users with speech impairments

63

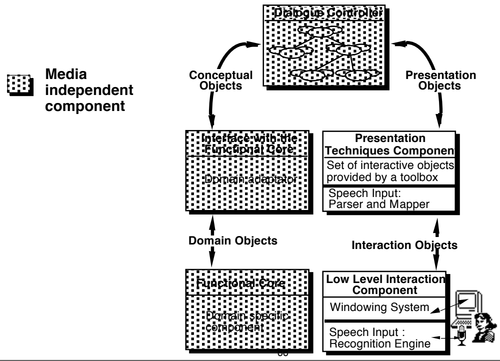
Multimodal input/output

- Consistency
  - System output independent of varying input modalities
    - the same keyword provides identical results whether user searches by typing or speaking
- Feedback
  - Users should know which modalities are available to them
- Error Prevention/Handling
  - If an error occurs, allow users to switch to a different modality

Outline of Chapter 3 - part 3 (multimodal interaction)

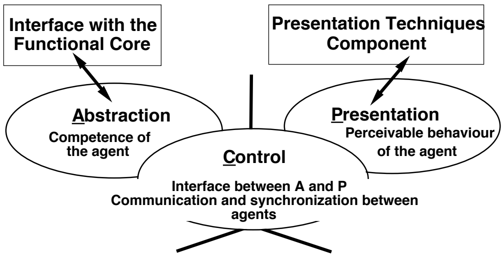
1. Introduction: motivation
2. Multimodality from the human perspective
3. Multimodality from the system perspective
4. Design guidelines
5. Software architecture modeling and the fusion pb (if we have time)

The PAC-Amodeus model



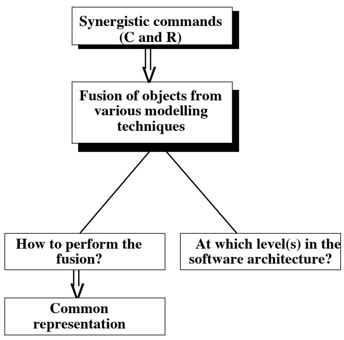
The PAC-Amodeus model and the Dialogue Controller

A multi-agent decomposition of the Dialogue Controller



Media independent PAC agent

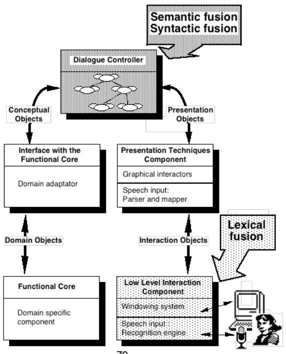
The problem of fusion



Levels of fusion

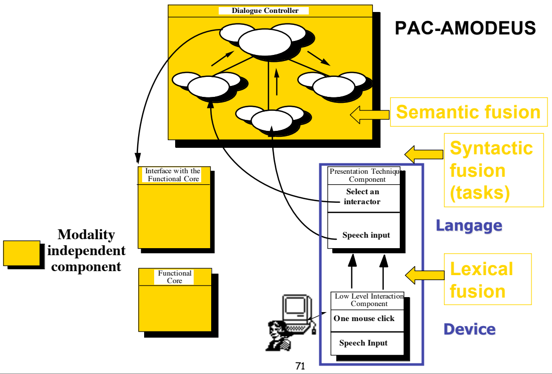
- Lexical fusion
  Shift key + mouse click -> long select
- Syntactic fusion
  Multimodal events -> a complete command
- Semantic fusion
  Results of commands -> new results

Levels of fusion



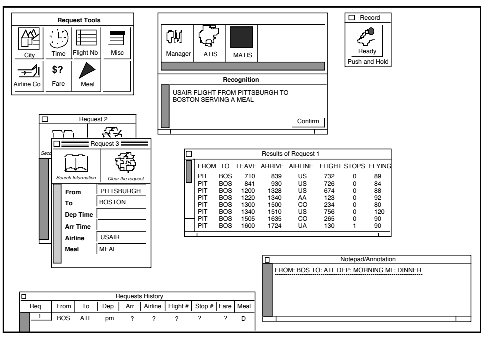
70

Levels of fusion



71

Running example: MATIS [L. Nigay]



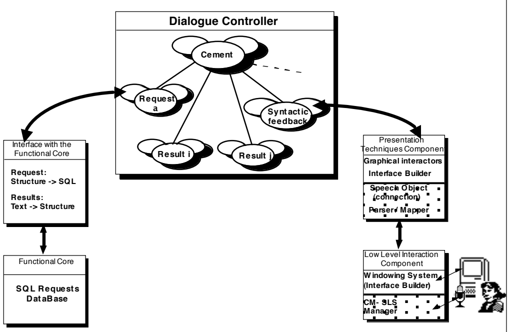
72

MATIS Overview

- **No prevailing modality: Same power of expressiveness**
- **Exclusive use of the modalities**
  <Show me the American flights from Pittsburgh to Boston>
- **Synergistic use of two input modalities mixing speech and mousing**
  <Show me USAir flights from Pittsburgh to this city>
  + <Selection of a city using the mouse>
- **Multithreading: work on multiple requests in an interleaved way**

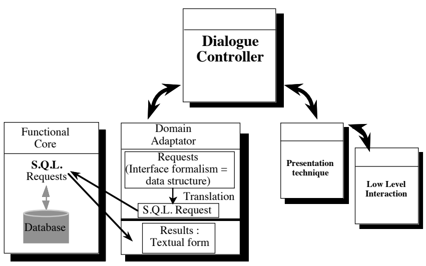
73

Software design of MATIS



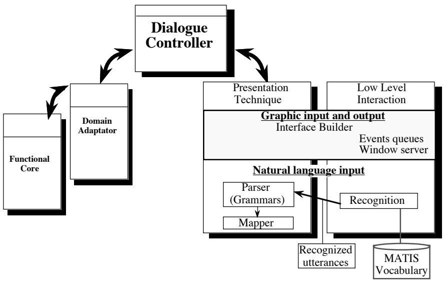
74

Software design of MATIS



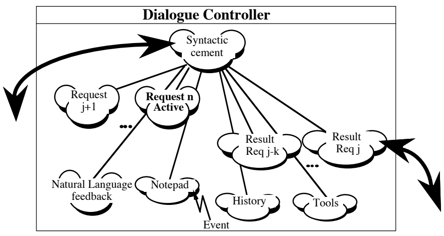
75

Software design of MATIS



76

Software design of MATIS

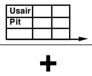


77

Fusion and Concurrency

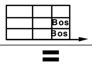
An example

Speech:<*Usair flights from Pittsburgh to this city*>



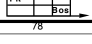
+

Mouse:<Selection of Boston inside a results window>



=

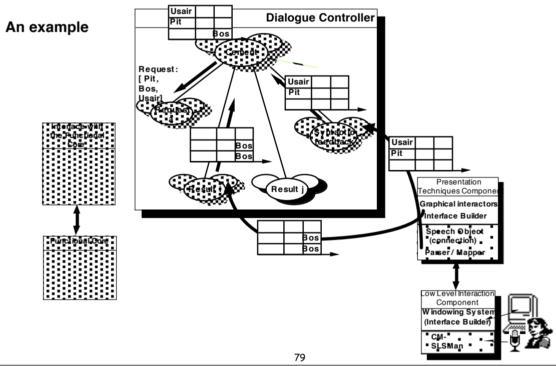
A new request: Usair flights from Pittsburgh to Boston



78

Fusion and Concurrency

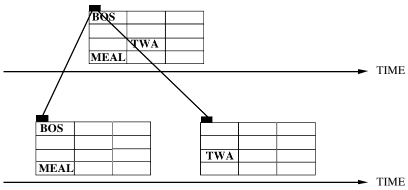
An example



79

Fusion

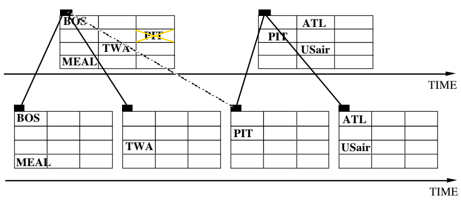
Speech : "Flights from Boston serving a meal" while selecting "TWA" using the mouse



80

Fusion

Speech :
"USAir flights from Atlanta to this city" while selecting "PIT" using the mouse



81