# Multimodal Interaction

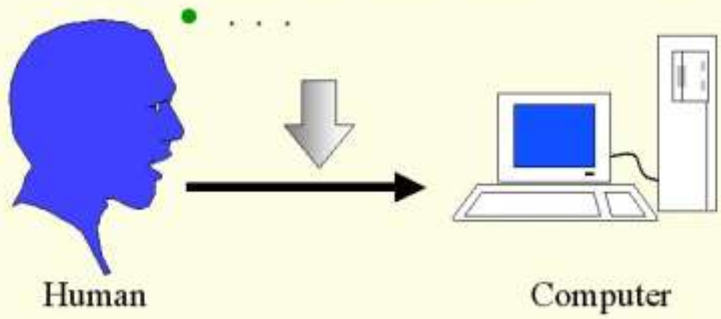
## Lesson 2
## Multimodal Interaction at a glance

Maria De Marsico
demarsico@di.uniroma1.it

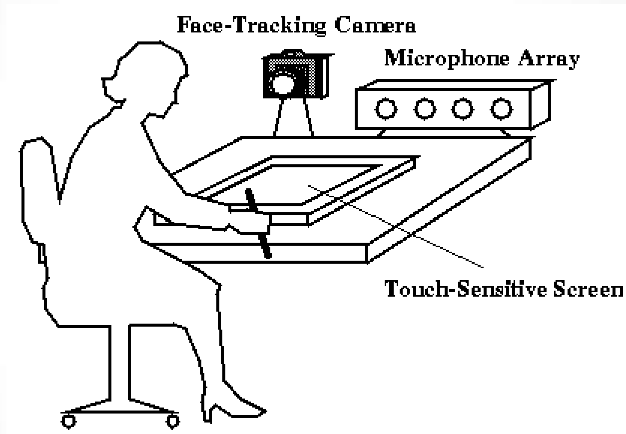


# Multimodal interfaces

- keyboard
- mouse
- voice
- gesture (pen, hand, movement)
- . . .

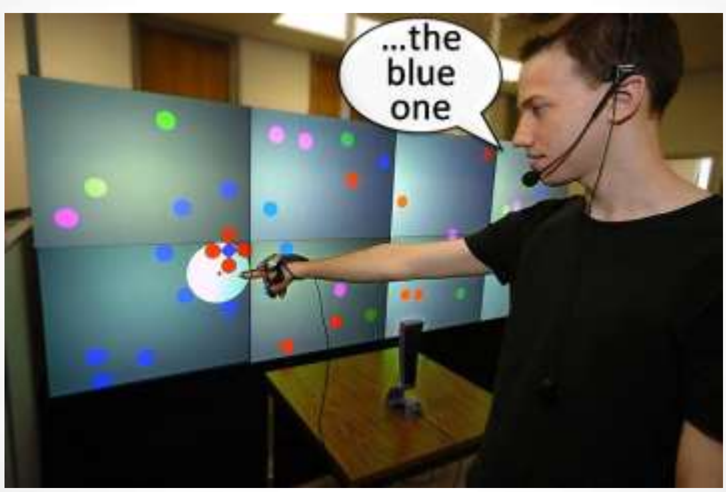


# Multimodal interfaces






# Multimodal interfaces



# Put-That-There


Richard Bolt's 1980 **voice and gesture** large display user interface, Put-That-There



# Put-That-There

http://www.cs.brown.edu/~avd/UI/Put-That-There-TwoPeopleClip.mov

# Multimedia <> Multimodal

- **Multimedia** = has to do with the structure of data
- **Multimodal** = has to do with communication channels
- **Multimodal interfaces** = are characterized by the (possibly simultaneous) use of multiple human sensory modalities and can support combined input/output modes.



# Mode <> Modality

- The term **multimodal** recurs across several domains.
- Its affinity and derivation from the terms ''**mode**'' and ''**modality**'' are often discussed.
- According to Merriam-Webster:
  - one of the meanings for **mode** is ''a possible, customary, or preferred way of doing something'' → gesture, gaze, or posture have generally been termed nonverbal modes of communication
  - **modality** can be ''one of the main avenues of sensation (as vision).'' → all the above modes exploit the visual channel
- In multimodal interfaces:
  - the former influences the way information is conveyed,
  - the latter refers to the exploited communication channel.
- Both express peculiar aspects of a multimodal system, which is expected to provide users with flexibility and natural interaction.

# Multimedia for Multimodal

- **Media**= hardware/software device allowing interaction between a user and a system according to a given set of modes (according to a modality)
  - For auditivemodality, we have many possible media: voice, music, sound, noise
- A **multimodal** user interface implies one or more kinds of media for each modality



# Carriers and senses

- Information is physically instantiated in some way:
  - in sound waves, light or otherwise = *physical carriers*
- To physically capture the information conveyed, humans have sensor systems, including the five classical senses of sight, hearing, touch, smell and taste
- Each carrier corresponds to a different sensor system.

| Medium | Physical carrier | Sensor system |
|---|---|---|
| Graphics (any visual) | Light (wave–particle duality) | Vision |
| Acoustics | Sound (waves) | Hearing |
| Haptics | Mechanical impact (forces) | Touch |
| Olfaction | Chemical impact (molecules) | Smell |
| Gustation | Chemical impact (molecules) | Taste |

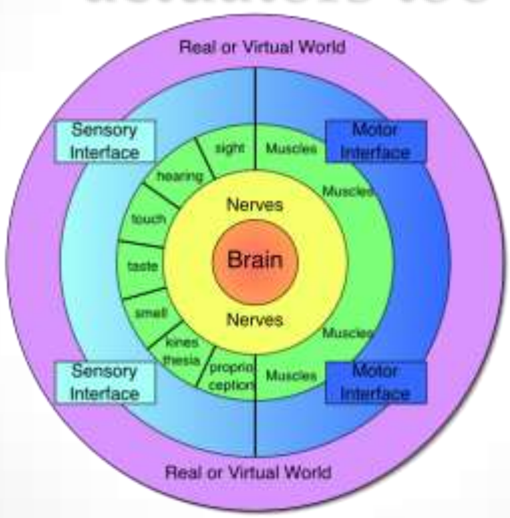
# Humans have sensors and actuators too

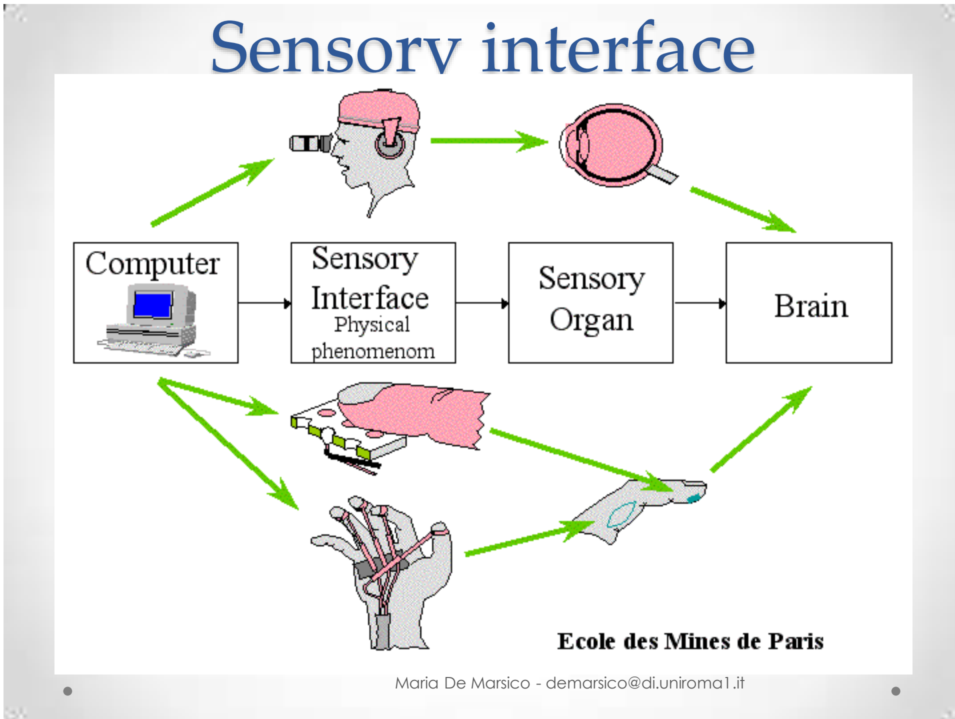
- The following images are from:

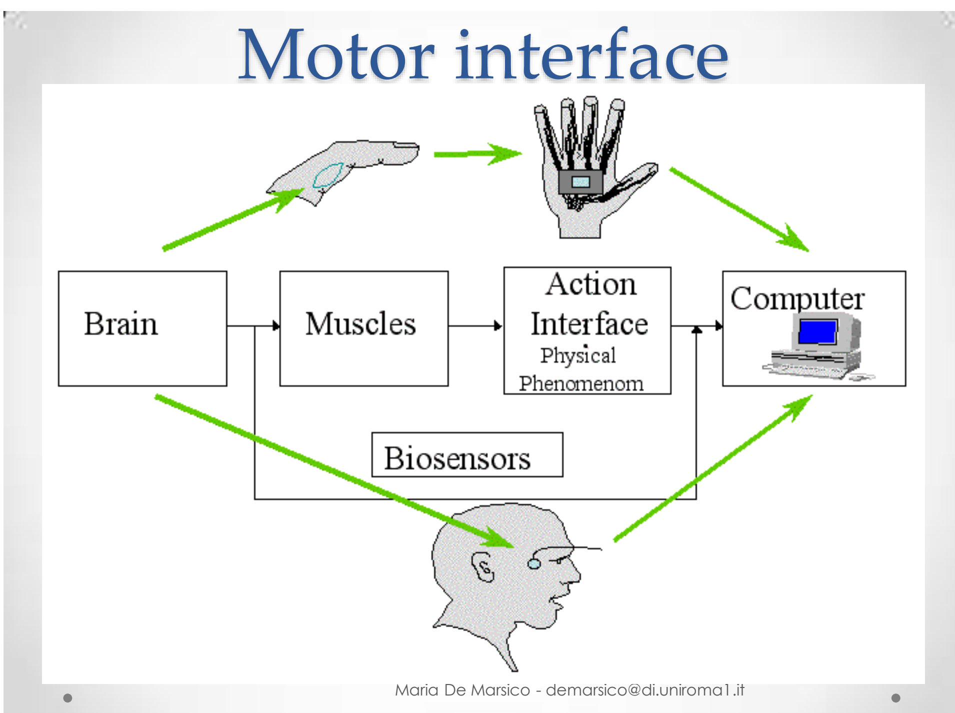
http://www.vrarchitect.net/anu/ivr/vr/printNotes.en.html a very interesting introduction to concepts of Virtual reality



# Humans have sensors and actuators too




Sensory interface
Computer
Sensory
Interface
Physical
phenomenom
Sensory
Organ
Brain
Ecole des Mines de Paris
Maria De Marsico - demarsico@di.uniroma1.it

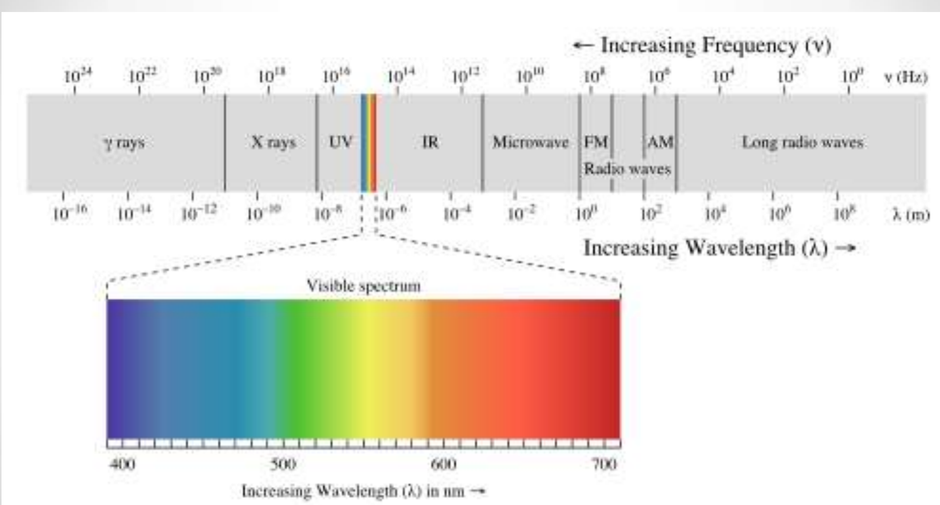

Motor interface
Brain
Muscles
Action
Interface
Physical
Phenomenom
Computer
Biosensors
Maria De Marsico - demarsico@di.uniroma1.it

# Limits

- Transmitted information to be perceived by a human recipient, must respect the limitations of the human sensors
- For instance
  - the human eye only perceives light in, approximately, the 400–700 nm wavelength band. Light intensity and other factors matter as well
  - the human ear can perceive sound in the 18–20,000 Hz frequency band and only if of sufficient intensity
  - touch information (often collected by the hands, but can be collected by all body parts) must be above a certain mechanical force threshold to be perceived, and its perception also depends on the density of touch sensors in the part of human skin exposed to touch.
- Genetically determined limits
- Due to a variety of factors, human thresholds differ from one individual to another and over time.



# Light waves

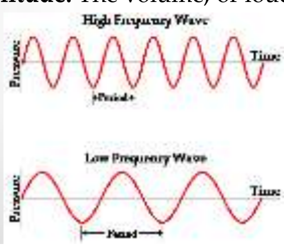


# Sound

**Wavelength**. This is the period between waves of sound; the number of cycles completed per second is the frequency in Hertz. Expressed in Hertz (Hz), cycles per second, the human ear perceives frequencies ranging from 20 Hz to 20,000 Hz, although as humans age they tend to lose their ability to hear high frequency sounds. Hearing ranges for some animals include:

| | |
|---|---|
| domestic cats | 100-32,000 Hz |
| domestic dogs | 40-46,000 Hz |
| African elephants | 16-12,000 Hz |
| bats | 1000-150,000 Hz |
| rodents | 70-150,000 Hz |

**Amplitude.** The volume, or loudness of a sound = height of the wave

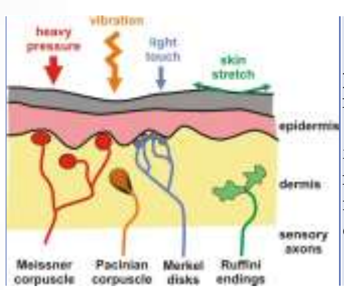





# Tactile Perception

(from http://www.pc.rhul.ac.uk/staff/J.Zanker/PS1061/L6/PS1061_6.htm)

- signals and sensors




**pressure, deformation, strain** need to be measured

for this, we have a large number of **mechanoreceptors** in our skin, muscles, connective tissue

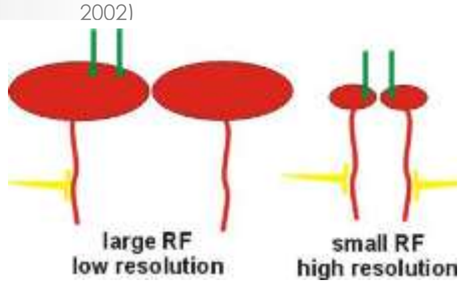
different types of mechanoreceptors respond to different physical conditions
Merkel receptor – slow (~10 Hz, continuous) > light touch
Meissner corpuscle – medium fast (~50 Hz) > pressure
Ruffini cylinder – fast (~ 100 Hz, continuous) > stretching
Pacinian corpuscle – very fast (~400 Hz, transient) > vibration

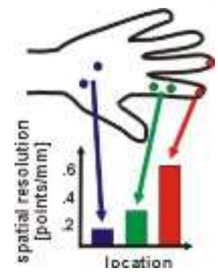
NOTE: each of these receptors is **specific in their tuning** for frequency, they show adaptation effects, and have **receptive fields**...

# Tactile Perception

(from http://www.pc.rhul.ac.uk/staff/J.Zanker/PS1061/L6/PS1061_6.htm)

- **tactile receptive fields**
  - two **pressure points** are perceived as separate if they are stimulating the receptive fields of different mechanosensors (adapted from Goldstein 2002)




the **receptive field size** determines **spatial resolution** (the number of points that can be detected in a given skin area)

tactile resolution (receptive field size) **varies for different areas of the body surface**: finger tips are better than palm of the hand, etc...

# Electronic Nose and Virtual Olfactory Display

The following material is from:
Fabrizio Davide, Martin Holmberg, Ingemar Lundström. **Virtual olfactory interfaces: electronic noses and olfactory Displays. In: Identity Community and Technology in the Internet Age** Edited by *G. Riva* and *F. Davide*, IOS Press: Amsterdam, 2001

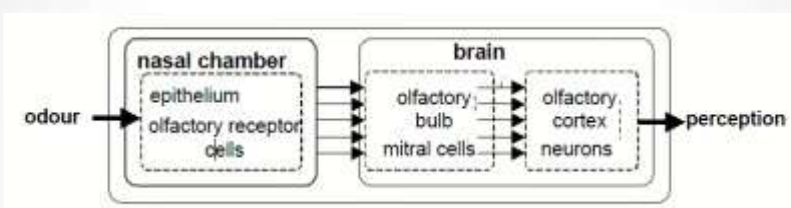
- "**Smell** has the tremendous power of having long range, and it has been far more important for survival during the evolution than sight and hearing, as witnessed by the incredible amount of genes codifying olfactory receptors in humankind (nearly 1000 over 100.000 involved, an enormous percentage among the others gene families)."

# Some terms

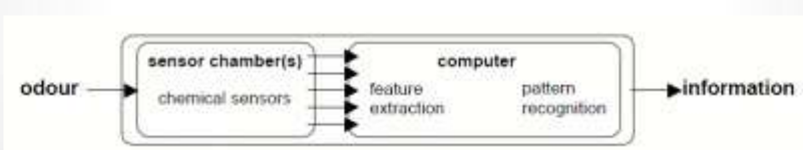
- **Odour**: the property or quality of a thing that stimulates or is perceived by the sense of smell
- **Odorants**: the chemical substances (volatiles) which come from an object and stimulate the smell are called
- **Olfaction**: the act of smelling odorants
- **Electronic nose:** an electronic system that, just like the human nose, tries to characterise different gas mixtures. It uses a number of individual sensors (typically 5- 100) whose selectivities towards different molecules overlap
- **Virtual olfactory display** (VOD): a system made of hardware, software and chemicals, able to present olfactory information to the (virtual environment) user
- In the science of **transducers** a VOD is simply seen as a transducer from the information domain (usually electric domain) to the chemical domain (in gas phase)



# Electronic Nose




Human olfactive system summarized



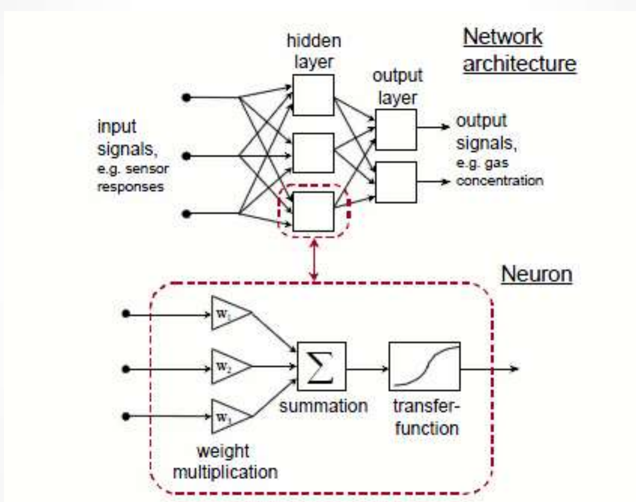

Electronic nose schema

# Electronic Nose

| HUMAN | ELECTRONIC |
|---|---|
| ~ 10 million receptors, self generated | 5-100 chemical sensors manually replaced |
| 10-100 selectivity classes | 5~100 selectivity patterns |
| Initial reduction of number of signals (~1000 to 1) | "smart" sensor arrays can mimic this? |
| Adaptive | Perhaps possible |
| Saturates | Persistent |
| Signal treatment in real time | Pattern recognition hardware may do this |
| Identifies a large number of odours | Has to be trained for each application |
| Cannot detect some simple molecules | Can detect also simple molecules ($H_2$, $H_20$, $C0_2$…) |
| Detects some specific molecules | Not possible in general at very low concentrations |
| Associative with sound, vision, experience, etc | Multisensor systems possible |
| Can get "infected" | Can get poisoned |



# Electronic Nose: example



# Virtual Olfactory Display

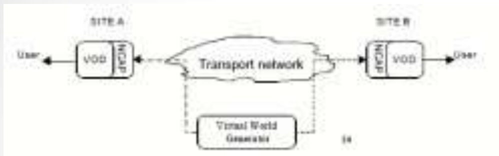
- Also called "**odour generators**" (AromaJet, DigiScents, TriSenx)
- They use a number of chemicals stored in a type of cartridge
- Upon receiving a signal describing an odour, they release a mixture of these chemicals, for example by using pumps similar to the ones used in ink printers. The resulting gas mixture is then blown towards the user with a small fan
- **Problem**: no standardised way of describing the odours has been created, so, one smell will be represented in different ways by different manufacturers
- **Design Issues**
  - **Saturation**: the perceived concentration of a certain odorant becomes stable as its concentration overcomes a specific level
  - **Interference**: odorants simultaneously presented may shield each other at a certain degree, so that the perception thresholds change
  - **Persistence**: how long does an odour last before fading a way and when should the display represent it?
  - **Smell field**: *smell field* for similarity with sound, a human can locate source with an error of 7-10°. Therefore a virtual olfactory display may be asked to position the odour in a sufficient smell field (order of 90-150°) with a sufficient angular resolution (in the order of 10-45°).



# Virtual olfaction <> Teleolfaction

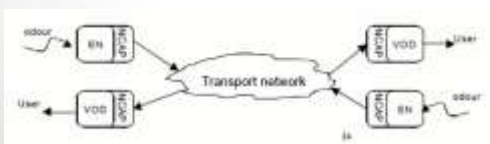
- **Virtual olfaction:** the act of smelling a mixture of odorants produced by a **virtual olfactory display**
  - centered on human smell but distinguishes the source of the odorant
- **Teleolfaction**: a form of virtual olfaction, the act of smelling a mixture of odorants, whose composition is related to a mixture present in a remote place
  - Further distinguishes the source of the olfactory information
- **Teleolfaction** deals with making **copies of reality**, and involves the problem of **fidelity**.

# Virtual, Remote, Communicative <> Fidelity




Virtual Olfaction
No fidelity required


Remote Olfaction
High fidelity required




Communicative Olfaction
Fidelity is not critical

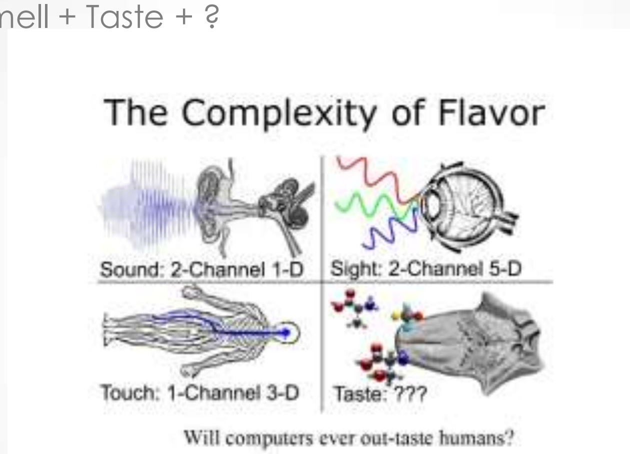
Teleolfaction

NCAP = Network Capable Application Processor



# Virtual flavors

- Smell + Taste + ?




From: Dan Maynes-Aminzade. Edible Bits: Seamless Interfaces between People, Data and Food. Presented at CHI 2005

http://www.vcasmo.com/video/aminzade/11072

# Virtual flavors

- «Until we can wire directly into human brains and transmit flavors as electrical impulses, we need a way to dispense chemical flavors using a limited repertoire of constituent ingredients. In light of the high dimensionality of flavor just described, we cannot realistically hope to reproduce the full breadth of possible flavors, nor the variations in smell and texture found in everyday food items (if we could, we would put gourmet chefs out of business). Instead, we strove for as large a breadth and expressiveness of flavors possible given a small "palate palette" of twenty flavoring agents. In selecting these flavoring agents, we turned to the food industry for suggestions.»

Maria De Marsico - demarsico@di.uniroma1.it

# Virtual flavors

Augmented Reality Flavors_ Gustatory Display
Based on Edible Marker and Cross-Modal Interaction
http://www.youtube.com/watch?v=qMyhtrejct8

Meta Cookie
http://www.youtube.com/watch?v=qMyhtrejct8

Maria De Marsico - demarsico@di.uniroma1.it

# Why multimodal?

## Easiest answers

- Ease


- Naturalness

- Engagement
- Pleasantness

Maria De Marsico - demarsico@di.uniroma1.it

# Why multimodal?

## Further answers

- The ultimate advantage of multimodal interfaces is increased **usability**
- **Redundant** or **complementary** information is conveyed by modes
- Higher **flexibility** : multimodal interfaces can accommodate a wide range of users, tasks and environments for which each single mode may not be sufficient
- Different types of information may be conveyed using the most **appropriate** or even less error prone modality, while alternation of different channels may prevent from fatigue in computer use intensive tasks
- **Redundancy** of information through different communication channels : supporting accessibility, since users with different impairments may benefit from information and services otherwise difficult to obtain
- Increased **robustness** of the interaction: the weaknesses of one modality may be offset by the strengths of another
- More **semantically rich** input streams can support mutual **disambiguation** for the execution phase
- As in human-human communication, the correct decoding of transmitted messages requires interpreting the mix of audio-video signals

Maria De Marsico - demarsico@di.uniroma1.it

# Why multimodal?

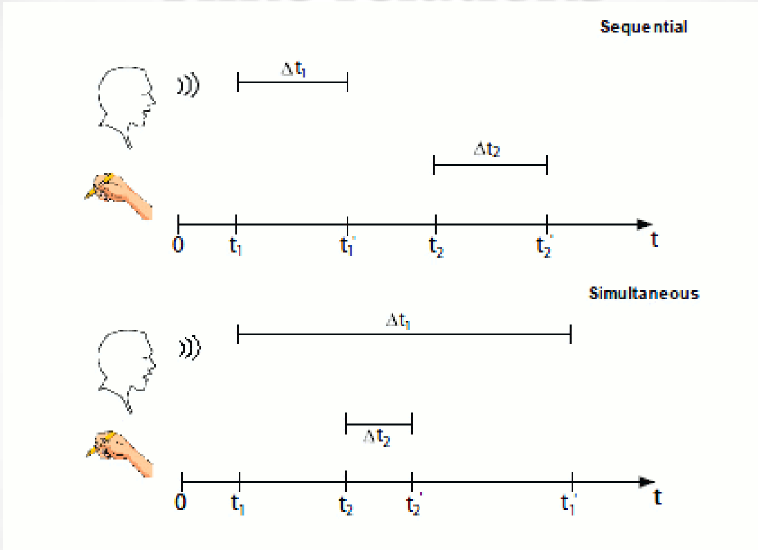
- It makes great differences to usability whether abstract information items are being represented in one or another modality, or in more than one
- Four examples
  - The blind cannot use any graphics modality as represented on a standard display
  - The seeing can, but try to represent the contents of any sentence using images only, and no text
  - This won't work, of course, but it is straightforward to read the sentence aloud for the blind
  - Despite blind gained most accessibility attention, deaf users have great difficulties with text, so it is no suitable to use too much text, and it is not sufficient at all to transduce auditory content by text



# Problem

- How to **integrate** and **synchronize** different modes
- **Synchrony** of different ''tracks'' of interaction in different modes <> **simultaneity**
- At present, each unimodal technique is developed separately
- Integration of more modal technologies → deep **understanding** of the ''**natural**'' integration patterns that characterize people's combined use of different communication modes
- The issue of integration may become even more **complex** when a multimodal interface is designed to support **collaborative** work, namely the work by multiple users who may interact through the interface using several input/output modes, either synchronously or asynchronously, and either locally or remotely.

# Time relations




From: P. Grifoni. Multimodal Human Computer Interaction and Pervasive Services. Information Science Reference. 2009



# Modality relations

| 1. Type of relation | 2. What it does | 3. Co-ordination | 4. Aimed at user groups |
|---|---|---|---|
| Complementarity | Several modalities necessary to express a single communicative act | Tight | Same |
| Addition | Add up different expressiveness of different modalities to express more information | Loose | Same or different |
| Redundancy | Express partly the same information in different modalities | Tight | Same or different |
| Elaboration | Express partly the same information in different modalities | Tight Loose | Same or different |
| Alternative | Express roughly the same information in different modalities | Loose None | Same or different |
| Stand-in | Fail to express the same information in a less apt modality | None | Same or different |
| Substitution | Replace more apt modality/modalities by less apt one(s) to express the same information | None | Special |
| Conflict | The human system cannot handle modality addition | Tight | None |

From: N. Ole Bernsen, L. Dybkjær. Multimodal Usability. Springer 2009

# Example applications

- Interaction in Mobile Environments
  - http://www.youtube.com/watch?v=W8FHg0g5iPs
- Geographic Information Systems
  - http://www.youtube.com/watch?v=qhET4a-Q0ow
- Interaction in Adverse Settings
  - http://www.youtube.com/watch?v=Lik61mHCcAk
- Multimodal Biometric Databases
  - http://www.youtube.com/watch?v=I2LCofq-Bts
- Interaction in Impairment Conditions



# Readings


- F. Davide, M. Holmberg, I. Lundström. Virtual olfactory interfaces: electronic noses and olfactory Displays. In: Identity Community and Technology in the Internet Age, Edited by *G. Riva* and *F. Davide*, IOS Press: Amsterdam, 2001
- N. Ole Bernsen, L. Dybkjær. Multimodal Usability. Springer 2009
- P. Grifoni. Multimodal Human Computer Interaction and Pervasive Services. Information Science Reference. 2009