Multimodal Interaction

Lesson 11
Multimodal Fission

Maria De Marsico
demarsico@di.uniroma1.it

Credits

Derived from:

- Patrizia Grifoni. **Multimodal Human Computer Interaction and Pervasive Services**. Information Science Reference. 2009

- Rousseau, C., Bellik, Y., Vernier, F., & Bazalgette, D. (2006**). A framework for the intelligent multimodal presentation of information**. *Signal Processing,* 86(12), 3696-3713. Available at http://perso.limsi.fr/Individu/bellik/publications/2006_SIGNAL_PROCESSING.pdf

Multimodal output

- An important issue for **communication** processes in general, and for **multimodal interaction** in particular, is the information **output arrangement and organization = multimodal fission**.
- Some issues to consider in designing and configuring fission:
  o information structure, intonation, and emphasis for the output by speech,
  o spatio-temporal coordination of pieces of information for visual (video, graphics, images, and texts) outputs
  o the design of appropriate output for each kind of modality
  o synchronization of the different outputs modalities

- Such activity is becoming more and more critical with the use of different interaction devices, from laptop to mobile and smart-phones, in different contexts.

Maria De Marsico - demarsico@di.uniroma1.it

Sketching the problem

- Foster, (2002) defines fission as *"the process of realising an **abstract** message through output on some **combination** of the available **channels**"*.
- This process can be conceived as consisting of three main steps:
- (1) *content selection and structuring:* selecting and organizing the content to be included in the presentation
- (2) *modality selection:* specifying modalities that can be associated with the different contents of the previous step
- (3) *output coordination:* coordinating the outputs on each channel in order to form a coherent presentation.

Maria De Marsico - demarsico@di.uniroma1.it

The problem in summary

- The fission process needs to consider *what* information has to be presented according to the interaction context and *how* this information can be presented in term of information structure, the chosen modalities for the output and their coordination/synchronization.

Maria De Marsico - demarsico@di.uniroma1.it

Information structure …

- The **fission** process, and more generally, **the information presentation** activities are closely connected with the *information structure*, independently from the different modalities.
- It was introduced by Halliday (1967) and was initially used to structure a **sentence** into parts such as *focus*, *background*, *topics*, and so on.
- **Focus** identifies "information that is **new** or at least expressed in a **new way**" (Steedman 2000).
- **Background** expresses **old** or **given** information.

Maria De Marsico - demarsico@di.uniroma1.it

… in different channels

- In each channel we can rely on elements that contribute to identify the information structure
- <u>Speech</u>: *syntactic structures*, *word order*, *intonation* and *prosody*
- <u>Visual communication</u>: *layout presentation*
- The *focus* and *background* concepts have been introduced considering **informativeness** of **phrases** composing sentences, but can be extended to **visual elements** that compose an image.

Maria De Marsico - demarsico@di.uniroma1.it

Structuring information

- **Example:** structuring **visual** information (images, graphics, video, texts) requires **spatial** and **temporal** (for dynamic visualizations) coordination.
- The use of focus and background notion can be **extended** to information structure associated with **multimodal utterance**.
- When two or more than two modalities are jointly used, **some** of them provide the **new information (focus)** and some others give the **information context (background)**.
- The modality that usually is involved in expressing the **focus** is the **prevalent** modality, i.e. the modality that can **significatively** express the information content.
- It will be convenient to choose the prevalent modality **according** to the different **users** and **contexts**.
- Examples:
  - o do not choose a prevalent output modality that uses **visual channel** for systems used by **visually** impaired people
  - o do not choose **speech** when the environment presents **sounds noises**.

Maria De Marsico - demarsico@di.uniroma1.it

Focus on context

- From Merriam-Webster online:
- **1:** the parts of a discourse that surround a word or passage and **can throw light on its meaning**
- 2**:** the **interrelated** conditions in which something exists or occurs

- We can identify
- an **intra-modality context**: defined by parts that mutually influence each other (reinforce or complement) using the same channel
- an **inter-modality** context: defined by **inter-modality spatial and temporal relations**

Maria De Marsico - demarsico@di.uniroma1.it

Focus on context

- The literature proposes a lot of definitions for context and in particular for *interaction context*.
- Schilit et al. (1994) claimed that the important aspects of context are: **where** the user is, **who** the user is with, and **what resources** are nearby.
- They define **context** to be the **constantly changing** execution environment.
- The environment is threefold:
  o **Computing environment**: available processors, devices accessible for user input and display, network capacity, connectivity, and costs of computing.
  o **User environment**: location, collection of nearby people, and social situation.
  o **Physical environment**: lighting and noise level.

Maria De Marsico - demarsico@di.uniroma1.it

Focus on context

- Anind Dey et al. (2001): interaction context as "*any information that can be used to **characterize** the **situation** of an **entity**. An entity is a **person** or **object** that is considered **relevant** to the **interaction** between a user and an application, including the user and application themselves.* Context is typically the **location**, **identity**, and **state** of people, groups, and computational and physical objects.".

Maria De Marsico - demarsico@di.uniroma1.it

Visual information and structure

- **Gestalt** theory was introduced for the visual perception, but it influenced the philosophy and the culture during all the 20th century.
- It was based on the **holistic** view according to **a whole is more than the sum of units that compose it**.
- Information structure concept and perceptual theories converge in some principles.
- An important principle is the *Figure/ground principle*, which shows the human perceptual *tendency to separate figures from their backgrounds.*
- **Figures** correspond to the **focus**, while the **ground** is the environment or **background** surrounding the figure.

Maria De Marsico - demarsico@di.uniroma1.it

Gestalt laws

- Gestalt laws drive figure/ground separation are separated into seven categories:
- Proximity, Similarity, Closure, Good Continuation, Common Fate, Good Form, and Experience

Maria De Marsico - demarsico@di.uniroma1.it

Georges Seurat

A Sunday Afternoon on the Island of La Grande Jatte
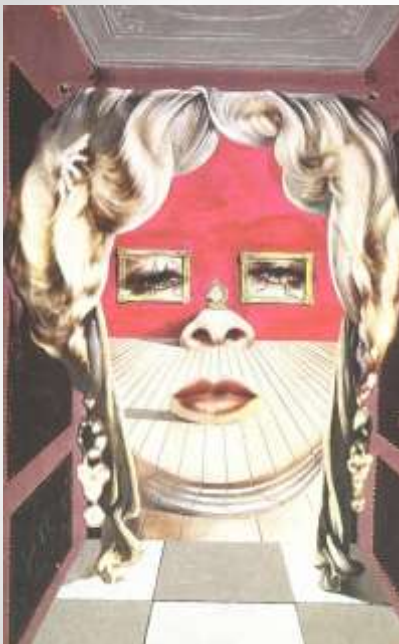1884-1886

Maria De Marsico - demarsico@di.uniroma1.it

Georges Seurat
The Circus   1890

Salvador Dali'
Mae West 1934 - 36

Proximity 1

All else being equal, perception tends to **group** stimuli that are **close** together as **part of the same object**, and stimuli that are far apart as two separate objects. This allows for the grouping together of elements **into larger sets**, and reduces the need to process a larger number of smaller stimuli. For this reason, **people tend to see clusters of dots** on a page instead of a large number of individual dots (Seurat).
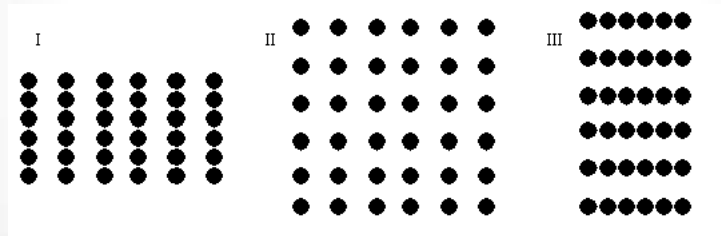
Maria De Marsico - demarsico@di.uniroma1.it

Proximity 2

Maria De Marsico - demarsico@di.uniroma1.it

Proximity 3

In the relatively brief history of computing, the methods by which users interact with digital information have been largely artificial and unintuitive. Although people perceive information in the world through a combination of sight, sound, touch, feel, and smell, people using computers have perceived information primarily through static, monochromatic displays of text and numbers. The development of the NCSA Mosaic Web browser in 1993 made the digital information of the Internet more accessible by incorporating graphics and sound with text and by making navigation easy with a point-and-click interface. The subsequent exponential growth of the World Wide Web is a dramatic demonstration of how information can be made more accessible by incorporating visualization techniques.

Visualization is only one aspect of a broader range of methods of interacting with digital information that we will see in the future. Haptic (touch) feedback is beginning to see use in specialized applications, and the use of audio is becoming more common. People perceive information, however, primarily through vision, and the display of digital information is likely to continue to be designed primarily for visual perception. Although the computer world as a whole has not yet moved too far from monochromatic displays of text and numbers, the emerging field of information visualization has already produced many interesting examples of how information can be made more accessible through visual representations.

This paper surveys the variety of ways visualization is being used to make information more accessible. Most of the applications and techniques discussed in this paper have been in development for less than a decade. Other visualization applications have been in existence for quite a bit longer; scientific visualization, for example, is a well-established field. But scientific visualization is also a relatively specialized field, focused on data that describes physical objects and scientific measurements. It is a way to visualize real objects that are otherwise difficult to see and manipulate, such as molecular structures, or to view simulations of scientific phenomena, such as the flow of air over wings. Because of this specific focus, scientific visualization has been the domain of a relatively small number of trained scientists.

While developments in scientific visualization are quite interesting, this paper concentrates on information visualization designed for a broader audience. Specifically, this paper surveys visualization techniques and applications designed to enable a wide variety of computer users to more easily navigate information spaces, to better display retrieved information, and to improve their understanding of information.

Maria De Marsico - demarsico@di.uniroma1.it

Similarity 1

All else being equal, perception lends itself to seeing stimuli that physically **resemble** each other **as part of the same object**, and stimuli that are different as part of a different object. This allows for people to distinguish between adjacent and overlapping objects based on their visual texture and resemblance.
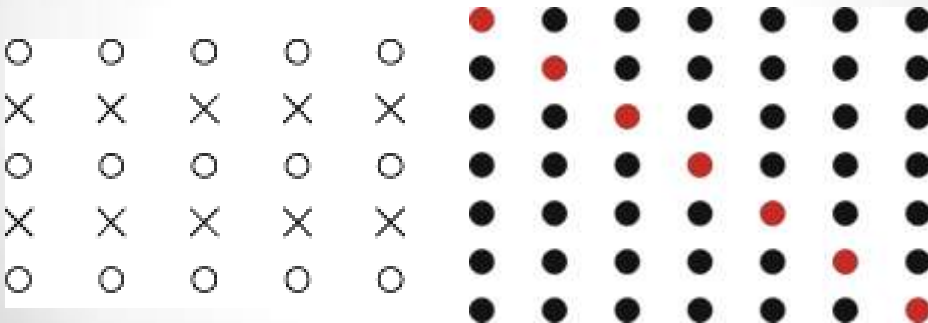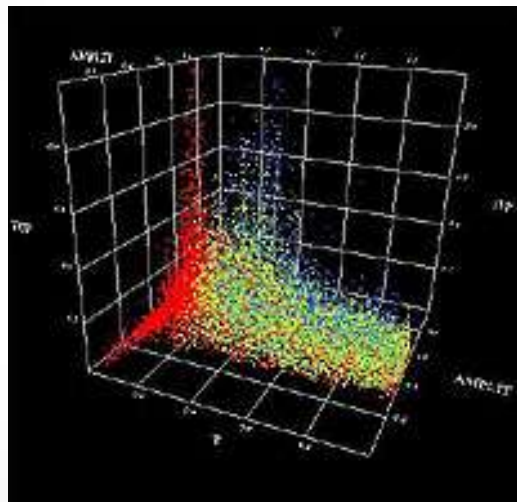

Maria De Marsico - demarsico@di.uniroma1.it

Similarity 2

Similarity 3

Closure 1

All else being equal, lines delimiting a closed surface are more easily perceived as a unit than not closing ones.
Mind tends also to see **complete** figures or forms even if a picture is incomplete, partially hidden by other objects, or if part of the information needed to make a complete picture in our minds is missing.

From Wikipedia: Closure is also thought to have evolved from ancestral survival instincts in that if one was to partially see a predator their mind would automatically complete the picture and know that it was a time to react to potential danger even if not all the necessary information was readily available.
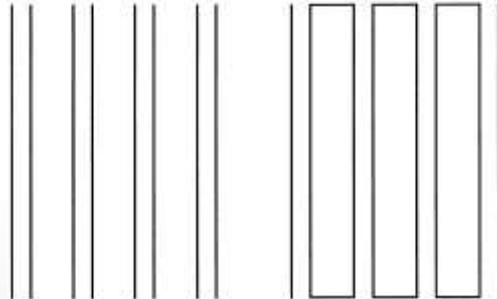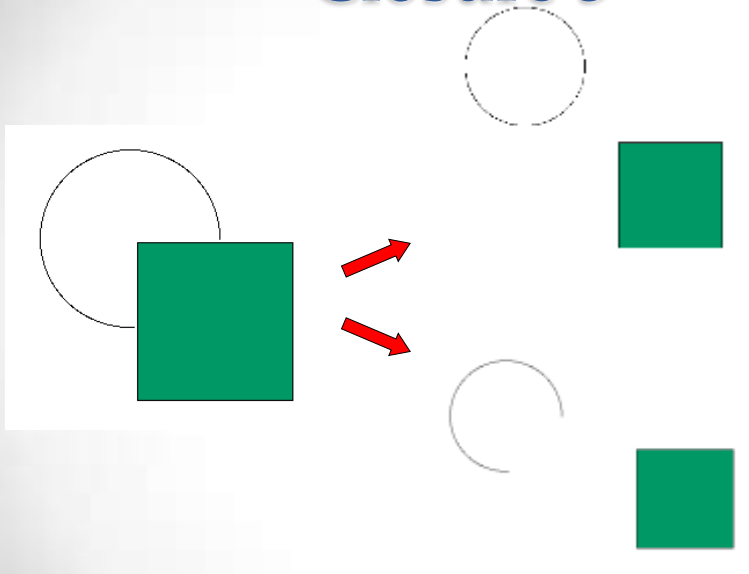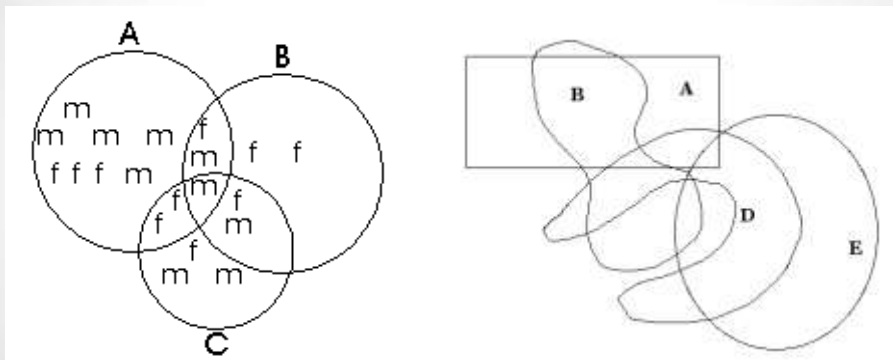
Maria De Marsico - demarsico@di.uniroma1.it

Closure 2
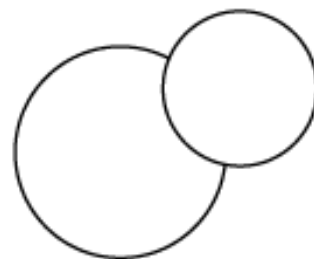
Maria De Marsico - demarsico@di.uniroma1.it

Closure 3

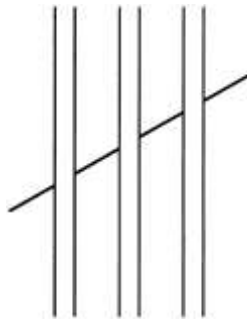

Maria De Marsico - demarsico@di.uniroma1.it

Closure 4
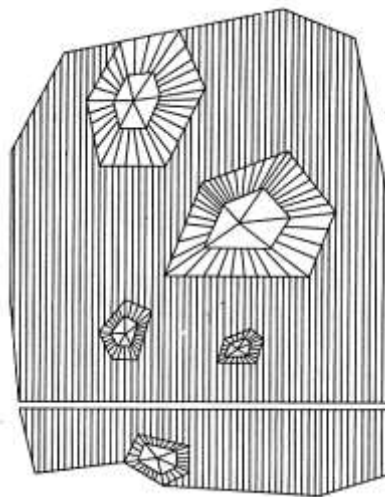

Maria De Marsico - demarsico@di.uniroma1.it

13

Common fate 1

When visual elements are seen "moving" in the same direction (at the same speed), perception associates the movement as part of the same stimulus.
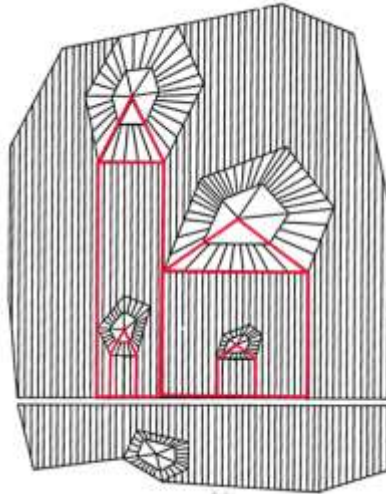
Maria De Marsico - demarsico@di.uniroma1.it

Common fate 2a

Maria De Marsico - demarsico@di.uniroma1.it

Common fate 2b

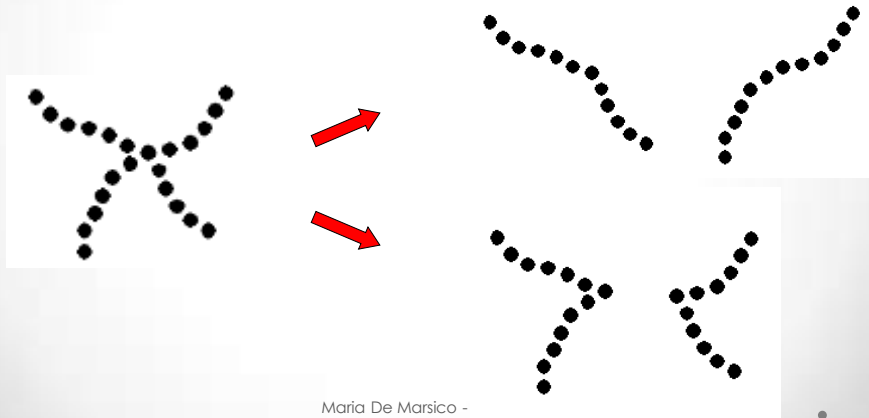Maria De Marsico - demarsico@di.uniroma1.it

Common fate 3

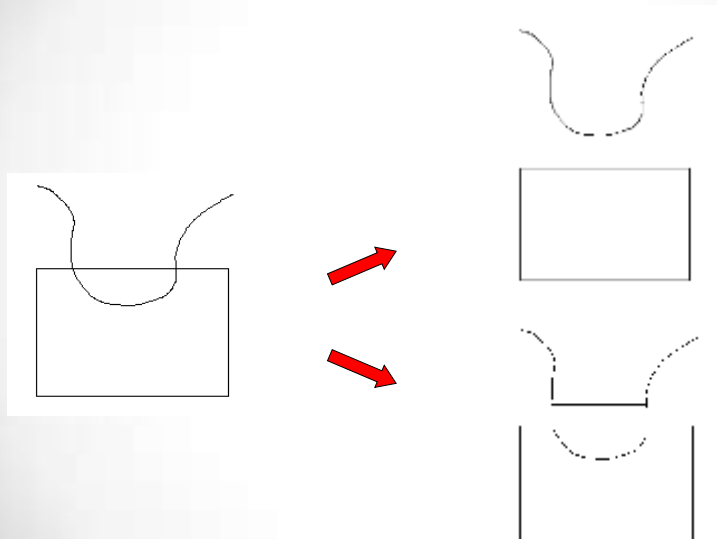Maria De Marsico - demarsico@di.uniroma1.it

15

Good continuation 1

When there is an **intersection** between two or more objects, people tend to perceive **each object as a single uninterrupted object**. This allows differentiation of stimuli even when they come in **visual overlap**. We have a tendency to group and organize lines or curves that follow an **established direction** over those defined by sharp and abrupt changes in direction...
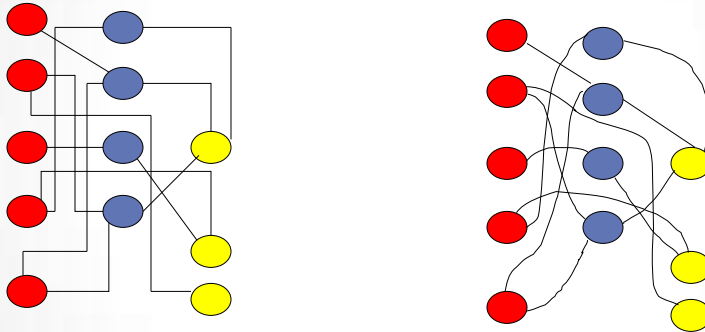
Maria De Marsico -

Good continuation 2

Maria De Marsico - demarsico@di.uniroma1.it
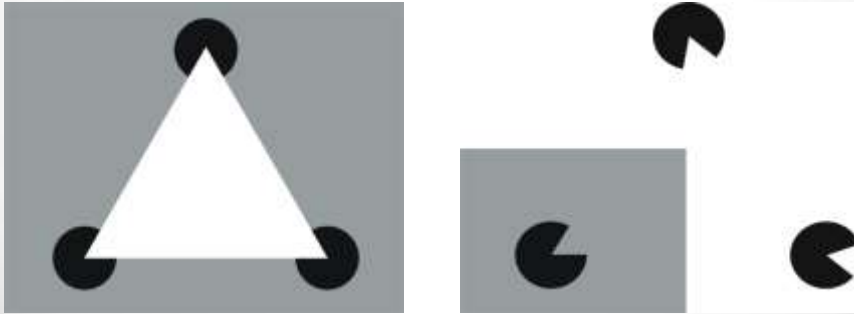
16

Good continuation 3

Good form 1

Mind tends to group structures in the perceptual field such as to identify **balanced**, **simple** entities with all parts respecting a similar principle

Good form 2

Good form 3

Good form 4

interpretazione = 2D

interpretazione = 3D

Maria De Marsico - demarsico@di.uniroma1.it

Good form 5
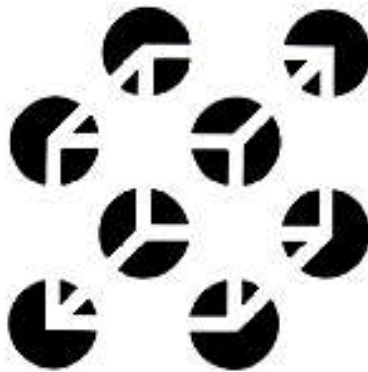
Maria De Marsico - demarsico@di.uniroma1.it

Experience 1

Mind tends to recognize patterns that are significant/familiar to us and thus fill in any information that may be missing.

Giuseppe Arcimboldo
(Milano, 1526 –
Milano, 11 luglio 1593)
Vertumnus

Giuseppe Arcimboldo
L' aria

Maria De Marsico - demarsico@di.uniroma1.it

Giuseppe Arcimboldo
Ortaggi in una ciotola

O …

Maria De Marsico - demarsico@di.uniroma1.it

Giuseppe Arcimboldo
L' ortolano

Reversible still life

Giuseppe Arcimboldo Fruit basket (Reversible still life)

Salvador Dali'
Galatea of the spheres

Maria De Marsico - demarsico@di.uniroma1.it

Maria De Marsico - demarsico@di.uniroma1.it

Important!

- What happens when laws conflict with each other, or when there are more possible arrangements?

- Ambiguous (multi-stable) images
- Impossible images

Maria De Marsico - demarsico@di.uniroma1.it

Ambiguous perception 1

Maria De Marsico - demarsico@di.uniroma1.it

Ambiguous perception 2
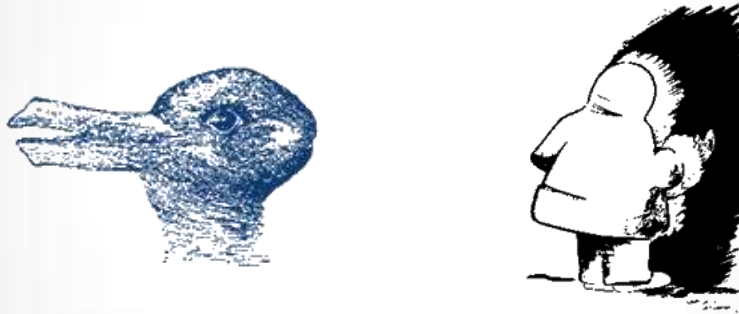
Ambiguous perception 3
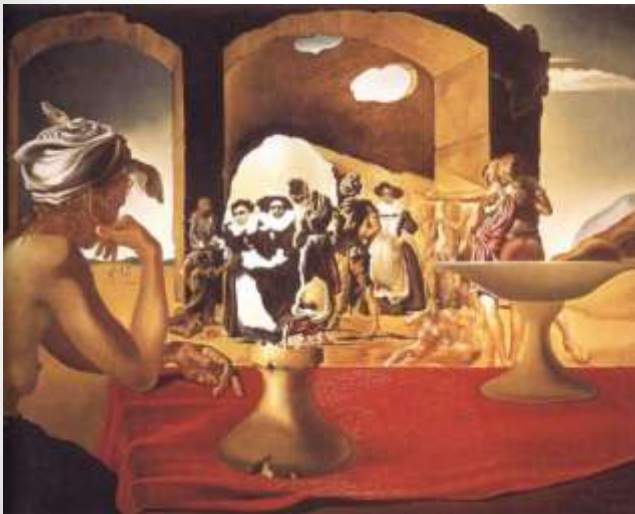
Salvador Dalí

The Slave Market
with Disappearing
Bust of Voltaire

Ambiguous perception 4

Maria De Marsico - demarsico@di.uniroma1.it

Percezione ambigua 5
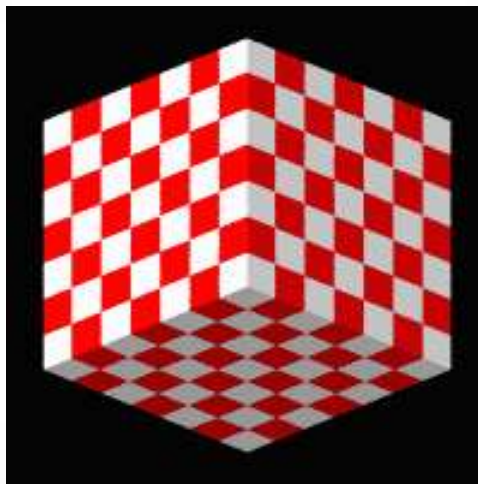
Maria De Marsico - demarsico@di.uniroma1.it

Tristable Images
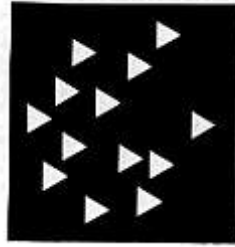
Maria De Marsico - demarsico@di.uniroma1.it

Impossible images 1

Maria De Marsico - demarsico@di.uniroma1.it

Impossible images 2
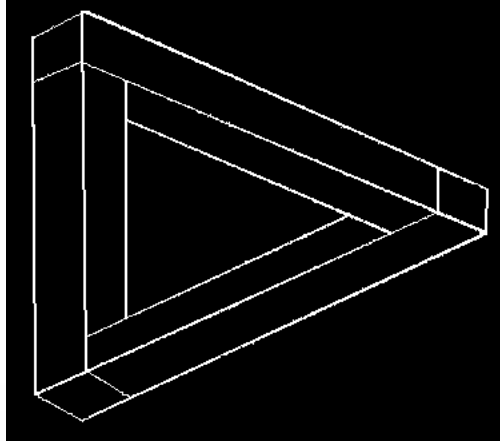
Maria De Marsico - demarsico@di.uniroma1.it

Impossible images 3
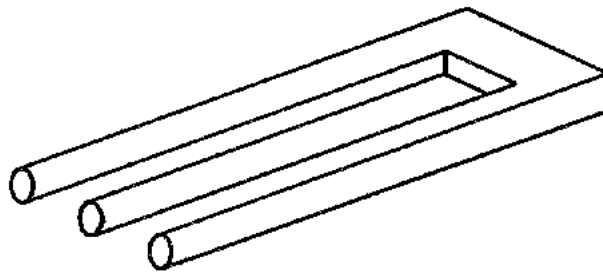
Maria De Marsico - demarsico@di.uniroma1.it

Impossible images 4

Maria De Marsico - demarsico@di.uniroma1.it

Impossible images 5
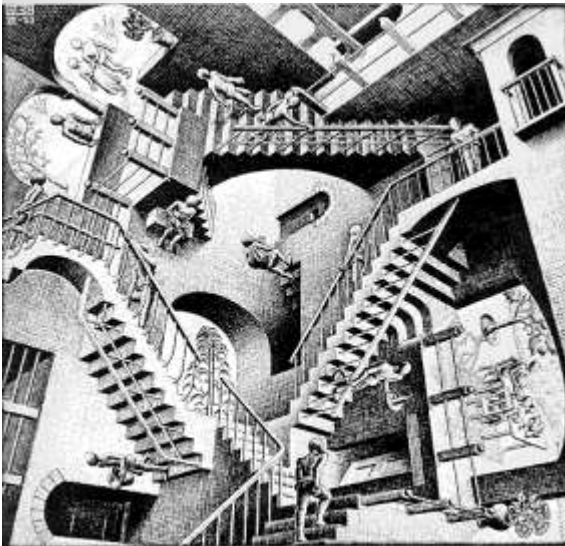
Maurits C Escher
I'm Going to Shoot that Builder
1953

Maria De Marsico - demarsico@di.uniroma1.it

Maurits C Escher
Waterfall

Reminder

- Once we have decided *what* information has to be presented according to the interaction context, we must decide *how* this information can be presented

- Visual **modality** in itself may support different **modes**
  - **static: text, tables, images**
  - **dynamic.: gesture**

- We have to identify modality and possibly mode

WWHT Model

- The following slides are mostly inspired by Rousseau, C., Bellik, Y., Vernier, F., & Bazalgette, D. (2006**). A framework for the intelligent multimodal presentation of information**. *Signal Processing, 86*(12), 3696-3713.

- **WWHT (What-Which- How-Then**) is a conceptual model for multimodal presentation of information and for the design of the multimodal systems output (Rousseau et al., 2006).
  - *What* is the information to present?
  - *Which* modality or modalities combination should we use to present this information?
  - *How* to present the information using the chosen modalities?
  - *Then*, how to handle the evolution of the resulting presentation?

Maria De Marsico - demarsico@di.uniroma1.it

Requirements

- The process of intelligent information presentation is based on four elements:
- information to present,
- interaction components,
- interaction context,
- behaviour.

Maria De Marsico - demarsico@di.uniroma1.it

Example

- The example that will be extensively used is that of the interaction with a mobile phone

Maria De Marsico - demarsico@di.uniroma1.it

Information to present

- Information is generally created by the functional core, forwarded by the dialog controller and presented by the output module.

- Example: the output module of a mobile phone may present the following information: "call of X", "message of X", "low battery level", etc.

Maria De Marsico - demarsico@di.uniroma1.it

Interaction Components

- Rousseau et al. use the **reverse** definition of mode/modality
- Output **modes** correspond to human sensory systems (visual, auditory, tactile, etc.).
- An output **modality** is defined by the **information structure** as it is **perceived** by the user (text, image, vibration, etc.) and not as it is represented internally by the machine.
  - Example: if a text is scanned then it may be represented internally by an image, but the perceived modality for the user is still text and not image.
- An output **medium** is an output device allowing the expression of an output modality (screen, speaker, vibrator, etc.).
- Output **media** are **independent** elements of the interactive system to achieve a better modularity.

Maria De Marsico - demarsico@di.uniroma1.it

Relations among components

- A **mode** can be associated with a set of **modalities** and each modality can be associated to a set of **media**.
  - Example: the "vibrator" medium allows the expression of the "vibration" modality which is perceived through the "tactile" mode.
- Two types of relations between the interaction components can be distinguished: "**primary**" and "**secondary**".
- A **primary** relation refers to a **wanted** effect whereas a **secondary** relation is a **side** effect.
  - Example: the vibration of a mobile phone is used to be perceived by the user in a tactile way. This implies a **primary** relation between "**tactile**" mode and "**vibration**" modality.
  - The **sound** generated by the vibrations is an example of **side** effect. So, a **secondary** relation between "**auditory**" mode and "**vibration**" modality can be added.

Maria De Marsico - demarsico@di.uniroma1.it

Example Relations

Modes	Modalities	Media
Visual	Photography / Text	Screen
Auditory	Ringing / Synthetical voice	Speaker
Tactile	Vibration	Vibrator

Primary relation
Secondary relation
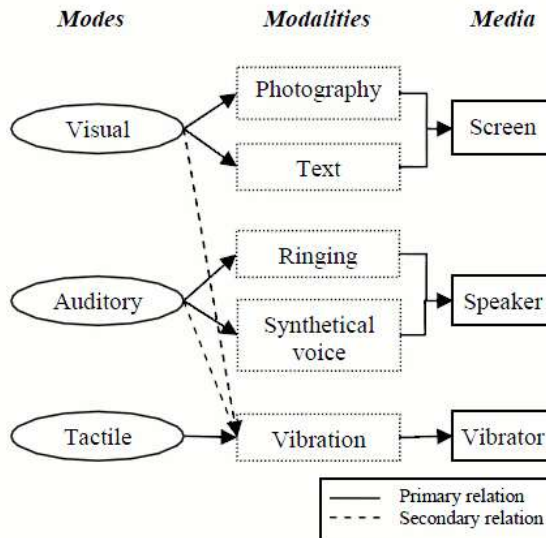
Example: Interaction components for a mobile phone

Maria De Marsico - demarsico@di.uniroma1.it

Interaction context

- Reminder: «Interaction context as "*any information that can be used to **characterize** the **situation** of an **entity**. An entity is a **person** or **object** that is considered **relevant** to the **interaction** between a user and an application, including the user and application themselves.*"(Anind Dey et al. (2001))

Maria De Marsico - demarsico@di.uniroma1.it

Example of context modeling

Criteria	Values	Model
Deaf person	Yes, No	User
Visually impaired person	Yes, No	User
Phone mode	Increased, Normal, Silent	System
Screen availability	Available, Unavailable	System
Speaker availability	Available, Unavailable	System
Vibrator availability	Available, Unavailable	System
Audio channel availability	Free, Occupied	System
Battery level	0-100	System
Noise level	0-130	Environment

Example: Interaction context for a mobile phone, where relevant information about user, system and environment is summarized

Behaviour

- The expression of information requires a multimodal **presentation suited** to the current interaction **context**.
- The presentation is composed by a set of output (**modality, medium**) pairs linked with **redundancy** or **complementarity** properties.
- Example: an incoming call on a mobile phone may be expressed through a multimodal presentation composed of **two** pairs.
  - First pair: ("ringing modality", "speaker medium") indicates a phone call
  - Second pair ("text modality", "screen medium") presents the caller's identity.

Behavioural model

- The behavioural model is probably the most critical part when designing a multimodal presentation.
- It identifies the **best** interaction components (**modes, modalities** and **media**) adapted to the current **state** of the interaction **context**.

- Formalization can be made in different ways:
  - rules
  - automats
  - Petri networks

Maria De Marsico - demarsico@di.uniroma1.it

Extracting the relevant elements

- Different steps are required to extract the required elements.
- A **preliminary** step requires to collect a **data corpus** composed of scenarios / storyboards (referring to normal or degraded situations) but also of relevant knowledge on application field, system, environment, etc.
- The following steps extract relevat information from the corpus
  1. To identify **pertinent** data which **can influence** the output interaction (**interaction context** modelling).
  2. To specify the **interaction components diagram**. Media are often defined in technical documentations and from media it is relatively easy to identify output modes and modalities.
  3. To identify semantic information which should be presented by the system.
  4. To decompose these information into elementary semantical parts.
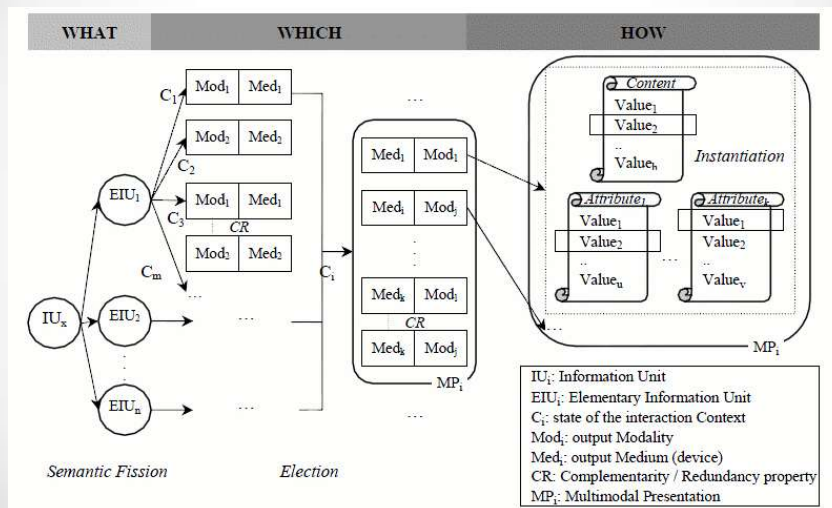- The extracted elements will allow the behavioural model definition.

Maria De Marsico - demarsico@di.uniroma1.it

Extracting the relevant elements
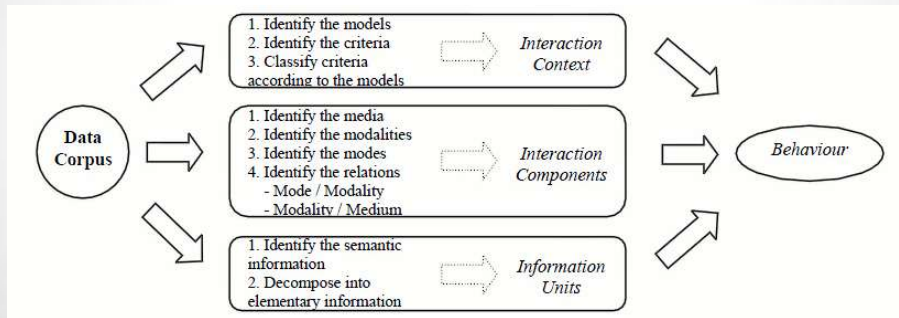
Maria De Marsico - demarsico@di.uniroma1.it

The design process

Maria De Marsico - demarsico@di.uniroma1.it

A further change in terms

- Till now we have used the word "fission" by the opposite to the word "fusion" to name the process of output modalities selection.
- In the work presenting the WWHT model
  - "semantic fission" happens during the decomposition of the semantic information into elementary information
  - "**allocation**" happens during the output modalities selection

Maria De Marsico - demarsico@di.uniroma1.it

Allocation

- The allocation of a multimodal presentation consists in selecting **adapted output modalities**.
- The selection process according to the **interaction context** is based on the **behavioural model**.

1. For **each** elementary **information unit**, a **multimodal** presentation **adapted to the current state** of the interaction context is selected.
2. Selected presentations are **merged** into only one presentation expressing the initial information.

Maria De Marsico - demarsico@di.uniroma1.it

Rules

- The behavioural model can be formalized by a base of election rules of the form (If … Then…instructions)
- Pro: limiting the learning cost
- Cons: problems on the scalability (evolution ability), the coherence and the completeness of a rule-based system.

- Example: two rules with equivalent premises must have coherent conclusions

Types of rules

- Three types of rules are defined in WWHT : contextual, composition and property rules.
- The premises of a **contextual** rule describe a **state** of the **interaction context**. The conclusions define **contextual weights** underlining the **interest** of interaction components (according to the context state described in the premises rule).
- The **composition** rules allow the modalities composition and so the conception of multimodal presentation with **several** (modality, medium) **pairs** based on **redundancy** and/or **complementarity** criteria.
- The **property** rules select a set of modalities using a global modality property (linguistic, analogical, confidential, etc.).

Allocation = Election

The application of the contextual and property rules defines the "pure" election = elects the best modality-medium pair

The application of the composition rules defines the "compound" election = election enriches the presentation by selecting new pairs redundant or complementary to the first one.

Maria De Marsico - demarsico@di.uniroma1.it

Example rules

Id	Name	Description in natural language
R1	Visually impaired person	If user is a visually impaired person Then do not use Visual mode
R2	Increased mode	If mobile phone is in increased mode Then use Redundancy property
R3	Speaker unavailability	If speaker is unavailable or audio channel is already in use Then do not use Speaker medium
R4	Low battery level	If current IU is a call reception and battery level is low Then do not use Photography modality and do not use Vibrator medium
R5	Too noisy	If noise level is superior to 80 dB or mobile phone is in silent mode Then Auditory mode is unsuitable
R6	Call event	If current EIU is an incoming call event Then Ringing modality is suitable
R7	Caller identity	If current EIU is a caller identity Then try to express it with Analogical modalities

Seven rules to allocate the "phone call of X" information: five of contextual type, one of composition type (R2) and one of property type (R7).
In a normal situation, only R6 and R7 rules are applied to present an incoming call. The call is then presented though a multimodal presentation composed of two pairs: (Ringing, Speaker) to indicate the phone call event (first EIU) and (Photography, Screen) to present the caller (second EIU).
In a different interaction context such as a low battery level, R4 rule changes the form of the last presentation (stops the use of the photography modality) by choosing the Text modality to present the caller.

Maria De Marsico - demarsico@di.uniroma1.it

How to present information

- The **instantiation** process consists in selecting the lexico-syntactical content and the morphological attributes of the presentation, according to the chosen mode/modality/medium for the Information Unit.
  - o Concrete content expressed through the presentation modalities are chosen .
  - o From this content, the presentation attributes (modalities attributes, spatial and temporal parameters, etc.) are fixed.
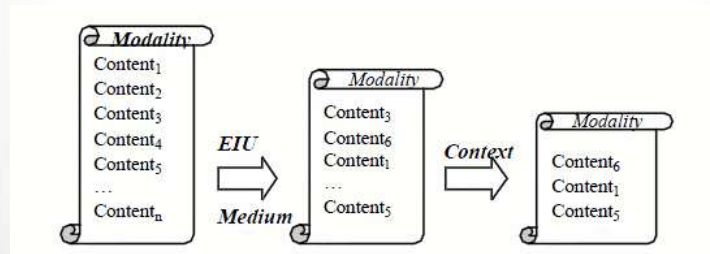
Maria De Marsico - demarsico@di.uniroma1.it

Example continued

- We suppose that the presentation of an incoming call is composed of two (modality, medium) pairs.
- A first pair (ringing modality, speaker medium) indicates the **phone call event** (first EIU) and the second pair (photography modality, screen medium) presents the **caller** (second EIU).
- An example of the presentation content may consist in using the pink panther **music for the ringing modality** and a **portrait of the caller** for the photography modality.
- For presentation attributes, we may use an **average volume** for the ringing modality and a full screen **resolution** for the photography modality.
- We can also decide to maintain this presentation during 15 seconds.

Maria De Marsico - demarsico@di.uniroma1.it

Selection of content

- For each modality, a set of all possible contents can be defined during the outputs specification.
- The elementary information units to express, the elected medium as well as the current state of the interaction context **reduce** the possibilities allowing the selection of the most suitable content for the modality.

Maria De Marsico - demarsico@di.uniroma1.it

And Then? Context evolution

- The evolution has to pay attention (**only**) to the **context elements** which may influence the **presentation**. These elements can be deduced from the **premises** of **contextual** rules which led to the design of the current presentation.
- A context **evolution** on one of the **criteria** which appears in these **premises** may change the application of the behavioural model.

Maria De Marsico - demarsico@di.uniroma1.it

Types of evolution

- An evolution factor is any element which requires an **update** of the presentation:
  - a context evolution
  - a user action on the presentation
  - a time constraint.
- Two types of evolution have been defined.
- The first evolution type called "**refinement**" changes the presentation **instantiation** (**local** state change).
  - Example: the increase of the vibration level or the ringing volume is a refinement of the presentation.
- The second evolution type called "**mutation**" changes the presentation **modalities** and/or media (**global** state change).
  - Example: the evolution from vibration modality to ringing modality is a mutation of the presentation.
- These two evolutions can be then used in sequence to strength in a progressive way the presentation of an incoming call event.

Maria De Marsico - demarsico@di.uniroma1.it

Readings

- Foster, M.E. (2002). **State of the art review: Multimodal fission**. Public deliverable 6.1, COMIC project.

- M. A. K. Halliday. **Notes on Transitivity and Theme in English: Part 1**. Journal of Linguistics, Vol. 3, No. 1 (Apr., 1967), pp. 37-81 Available at http://www.jstor.org/stable/pdfplus/4174950.pdf

- M. A. K. Halliday. **Notes on Transitivity and Theme in English: Part 2**. Journal of Linguistics, Vol. 3, No. 2 (Oct., 1967), pp. 199-244. Available at http://www.jstor.org/stable/10.2307/4174965

- Dey, A.K., Salber, D., & Abowd, G.D. (2001). **A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications.** In T.P. Moran, & P. Dourish (Eds.), *Human-Computer Interaction*, *16*(2-4), 97-166. Available at http://cleo.ics.uci.edu/teaching/Winter10/231/readings/8a-DeyAbowdContextToolkit.pdf

- Schilit, B., Adams, N., & Want, R. (1994). **Context-aware computing applications**. *Proceedings of the 1st InternationalWorkshop on Mobile Computing Systems and Applications*. Los Alamitos, CA: IEEE. Available at http://graphics.cs.columbia.edu/courses/mobwear/resources/schilit-mcsa94.pdf

- Rousseau, C., Bellik, Y., Vernier, F., & Bazalgette, D. (2006**). A framework for the intelligent multimodal presentation of information**. *Signal Processing, 86*(12), 3696-3713. Available at http://perso.limsi.fr/Individu/bellik/publications/2006_SIGNAL_PROCESSING.pdf

Maria De Marsico - demarsico@di.uniroma1.it

Readings

- M. Johnston and S. Bangalore, **Finite-state Multimodal Integration and Understanding**, Natural Language Engineering 11 (2) (2005) 159-187. Available at http://www2.research.att.com/~johnston/papers/johnstonbangalore2005jnle.pdf

Maria De Marsico - demarsico@di.uniroma1.it