Please insert in the following the information that is relevant to your project

Deliverable Title	State of the Art on current interaction paradigms based on vision and action
Workpackage number	WP4b
Deliverable Number	D4b.1
Edited by:	INPG
Approved by	INPG, NICOD, LUC-EDM, UPS, PERCRO, COSTEH, UNIGE, CEIT, UPPSALA, DIST
Nature of Deliverable	Report
Distribution[1] (as by Technical Annex)	PP
Contractual Delivery Date (DD/MM/YYYY)	30 / SEP / 2004
Actual Delivery Date (DD/MM/YYYY)	01 / OCT / 2004
Authors	Editor Luciani Annie (INPG) , WP4b leader Authors INPG : Luciani Annie, Couroussé Damien, François Thil, Daniela Urma, Sylvain Marlière NICOD : Nicolas Bullot (coordinator, Elena Pasquinelli, N. Gangopadhyay LUC-EDM : Joan De Boeck UPS XI : Le Runigo Cyrille, Benguigui Nicolas, Bardy Benoit UPPSALA : Gunnar Jansson PERCRO : Marcello Carrozzino COSTECH : Fabien PFAENDER, Gunnar DECLERK DIST : Volpe Gualtiero, Antonio Camurri, Barbara Mazzarino UNIGE : Jarlier Sophie, George Papagiannakis, HyungSeok Kim CEIT : Ignazio Mansa, Emilio Sanchez
Abstract	cf. Summary
Keywords	Computer Graphics, HCI, VR, VE, Teleoperation, Telecommunication, immersion, vis-à-vis, geeometrical modeling, light modelight, motion modeling, gesture, motion capture, expressive gesture, haptic interaction, haptic devices, colocation, visuo-haptic conflict, interceptive actions, temporal delay, stereoscopic devices.

[1] Please indicate the dissemination level for deliverables using one of the following codes:
PU = Public
PP = Restricted to other programme participants (including the Commission Services).
RE = Restricted to a group specified by the consortium (including the Commission Services).
CO = Confidential, only for members of the consortium (including the Commission Services).

ENACTIVE

"ENACTIVE INTERFACES"

Project IST-2004-002114-ENACTIVE

WP4b/D4b.1 /DOC/2004

Edited by: A. Luciani, INPG

Nature of Deliverable: Report

Distribution: PP

Contractual Delivery Date: 30 september 2004

State of the Art on current interaction paradigms based on action and vision

(D4b.1)

Actual Delivery Date:
1 October 2004

Abstract: cf. Summary

Keywords: Computer Graphics, HCI, VR, VE, Teleoperation, Telecommunication, immersion, vis-à-vis, geeometrical modeling, light modelight, motion modeling, gesture, motion capture, expressive gesture, haptic interaction, haptic devices, colocation, visuo-haptic conflict, interceptive actions, temporal delay, stereoscopic devices**.**

A.Status Sheet

DOCUMENT TITLE: State of the Art on current interaction paradigls based on vision and action

ISSUE	REVISION	DATE	CHAPTER - PAGE REVISED
1	0	1srt October 2004	n.a.

B. Executive Summary

Documents provided by the participants

Title	Authors and institutions	Reference
Action-Vision fusion in CG, HCI, VR, teleoperation, telecommunication : Cornerstones and trends	A. Luciani, INPG	EI_WP4b_STAR_040906_INPG
Action-Vision fusion in CG, HCI, VR, teleoperation, telecommunication : Novel issues	A. Luciani, INPG	EI_WP4b_STAR_040906_INPG
Geometrical, Light and Visualisation Modeling and Computation of 3D scenes	M. Carozzino, PERCRO Ignacio Mansa, CEIT	EI_WP4b_STAR_Chapter2_FinalDraf EI_WP4b_STAR_020804_MCardoc
Motion Modeling and Computation	A. Luciani, INPG	EI_WP4b_DLV1_INPG3_040930
Action Processing	A. Luciani, INPG D. Couroussé, INPG	EI_WP4b_DLV1_INPG2_040930
Temporal delays in action-vision loop	D. Couroussé, INPG	EI_WP4b_STAR_040826_INPG
Expressive gesture	Volpe Gualtiero (DIST) Antonio Camurri (DIST) Barbara Mazzarino (DIST)	EI_WP4b_STAR_040727_DIST
The HCI point of view:Multimodality in HCI	Joan de Boeck (LUC-EDM)	EI_WP4b_STAR_1_040720_LUCEDM EI_WP4b_STAR_1_040720_LUCEDM v2
Haptic modality. Some definitions and problems of classification Touch. The sense or reality Espitermic seeing Crossmodal attention capture Sensorimotor theories of perception	Nicolas Bullot (NICOD coordinator for this document) Elena Pasquinelli (NICOD) Nivedita Gangopadhyay (NICOD)	EI_WP4b_MNT2_First-Nicod-Proposal_040720_Bullot(coord)
Co-location of visually (stereoscopically) and haptically presented and perceived virtual objects	Gunnar Jansson (UPPSALA)	EI_WP4b_STAR_1(040805_UPPSALA)
Perceptual conflict in visuo-haptic integration. the case of the pseudo-haptic feedback and its implications for haptic interfaces	Fabien Pfaeffer (COSTECH) Gunnar Declerk (COSTECH)	EI_WP4b_STAR_040725_COSTECH WP4b_STAR_part1_Costech_2909

Interceptive actions	Le Runigo Cyrille (UPS)
	Benguigui Nicolas (UPS)
	Bardy Benoit (UPS)
		EI_WP4b_STAR_1(040721_UPSXI)
		EI_WP4b_STAR_1(040921_UPSXI)

STAR in Technologies for action-vision fusion: Gestural devices, stereoscopic devices, real time computer graphics, 3D sounds	Jarlier Sophie (UNIGE) George Papagiannakis (UNIGE) HyungSeok Kim (UNIGE)	EI_WP4b_MNT2_040723_UNIGE STA Mocap 1.7 & 4.3 EI_WP4b_STAR_Section8_3_StereoVision EI_WP4b_STAR_8_3_MiddleWare_040930_UNIGE

Table of contents

Summary

This document is composed of eight chapters.

Chapter 1. "Action-Vision fusion in CG, HCI, VR, teleoperation, telecommunication : Cornerstones and trends"
This part presents the scientific fields that are involved in action-vision fusion by:

- Focusing on their relevant historical evolution,
- Identifying cornerstones and trends,
- Identifying the common concepts and open issues towards which they are converging and that have been solved.

Chapter 2. Novel common issues in Action-Vision fusion in CG, HCI, VR, teleoperation, telecommunication
The common concepts and research issues pointed out in the Chapter 1 are detailed:

- Draftly faced to the main shifts in new technologies foreseen for the future : computational power, networking, new technologies in displays,
- Focusing on the main conceptual difficulties to overcome : immersion vs. vis-à-vis, exploration of large amount of data vs. accurate manipulation, presence,
- Analyzing the discrepencies between the action and the vision in term of technologies : physics for the action vs. geometry for the vision,
- Summarizing these discrepencies through functional diagrams of the action-vision chain.

Chapters 3, 4 and 5, are devoted to the state of the art in the three most important technological developments required for the action-vision computer implementation:
Chapter 3: 3D computer technologies to visualize 3D complex scenes and objects
Chapter 4: Computer modeling of the motion
Chapter 5: Computer processing of action

Chapter 3. "Geometrical, Light and Visualisation Modeling and Computation of 3D scenes"
From the state of the art of these technologies, the chapter adresses the major difficulties to overcome :

- Optimisation techniques in 3D Scenes complexity
- Shapes representations and their impact on the complexity of the scene
- Light representations and their impact on the computational costs
- New trends in real time visualisation
- Stereoscopic visualisation

Chapter 4. "Motion modeling and computation"
This chapter addresses the main technologies to represent and compute motions of virtual objects :

- Phenomenological models : key-frames, evolution functions, Kinematic representations
- Generative models : physically-based models
- Generative models : Artificial life, agent-based models, behavioural models

It stands each of them from an enactive point of view, underlining the conceptual and / or pragmatical adequations or discrepencies (1) with the human manipulation and control, and (2) with the shape and visual representations stated in Chapter 3.

Chapter 5. "Action processing"
This chapter adresses the processing of human action through input-ouput devices and the associated processing technologies.
It starts with a general framework focused on a typology of actions and proposing criteria to categorize them in the context of interaction with computerized environments.
The state of the art in processing of input gestures focuses in :

- Data representation in action and motion capture,
- Gesture recognition.

It points out the specific question of expressive content that is of paramount importance for enhancing usability and communication with computerized machines.
It underlines the conceptual and technological shift brought by the introduction of bilateral action (gesture action and gestural perception), pointing out the properties (energetic consistency, physical reactivity) that the correlated processing have to take into account.

Chapter 6. "The HCI point of view : Multimodality in Human-Computer Interaction"
Chapter 6 grounds the theoretical foundations of multimodality in HCI and the type of tasks HCI is confronted with. It details action-vision metaphors developed in such domain (general, navigation, selection, manipulation).
And finally it addresses the role of haptic interaction in computer interfaces, through:

- Some relevant applications,
- The underestimated case of two-handed interaction,
- The outlining of some problems and solutions in Haptic HCI.

Chapter 7. "Perceptual and cognitive issues in Action-Vision fusion"
A first paragraph grounds some definitions in touch and seeing. It addresses:

- Problems of classification in the definition of haptic modality,
- The main property of touch as to be the sense of the reality,
- The epistemic functionality of the vision
- A brief summary of the main streams in sensorimotor theories of perception

The following five paragraphs outline main perceptual and cognitive problems and properties in action-vision correlations:

- Co-location of visually and haptically presented and perceived virtual objects,
- Perceptual conflict in visuo-haptic integration : the typical case of the perceptual identification of a matter property,
- Properties of cross-modal attention capture,
- Control action mechanisms through vision and mainly the representative case of interceptive actions,
- The critical question of temporal delays in action input and sensory returns (gestural and visual feedbacks)

Chapter 8. "STAR in Technologies for action-vision fusion"
This chapter is survey on current technologies involved in action and vision relationship in computerized environments:

- Gestural devices. Terminology and systems:
  Terminology raises problems of classification in the definition of haptic device. Some commercial gestural devices are presented as examples of type of systems, including haptic devices and motion capture devices.
- Stereoscopic visualization devices.
- 3D sound devices and techniques (commonly with WP4a).
- A state of the Art (STAR) of current commercial middleware and scene graph APIs acting as 3D virtual frameworks for Real-time Graphics simulations
- A survey on 3D sounds in Virtual Environments.

Following this analytical state of the art, there are three annexes. These lists will be updated by the ongoing of the WP4b activity.
Annexe 1: List of scientific fields addressed by the action-vision relationship.
Annexe 2: List of keywords with definitions as the contribution of WP4b to the Lexicon
Annexe 3: List of commented references as the contribution of WP4b to the bibliographic Enactive Data base.

I. Analytical STAR in Action-Vision fusion

1 Action-Vision fusion in CG, HCI, VR, teleoperation, telecommunication: Cornerstones and trends

©Annie Luciani, INPG

1.1 Introduction

First remark:
If we except the technologies for speech communication, the evolution of interfaces is quite completely mapped on the evolution of the technologies for seeing and acting.

Second remark:
To understand and to steer the future of the link between and vision on interface technologies, we assert that:

- We have to point out the relevant changes in its evolution: from a pioneer's phase that set all the basic and still used concepts up to the intensive developments of them and their contemporary reunification.
- We have to ask several scientific and technological fields such as Computer Graphics, Human-Computer Interfaces, tele-operation, tele-communication, Virtual reality.
- We have to put in perspective these two points under the lighting of the contemporary technological breakthroughs.

This state of the art tries to present the relevant issues in each of these three points.

1.2 Pioneer phase: the primary fusion of hand and vision in Computer technologies

In the domain of Computer Aided Design, the first famous Sketchpad system of I. Sutherland (1963) **pioneered** the concepts of graphical computing, was the first GUI (Graphical User Interface) long before the term was coined [1] [2]. For the first time, we were able to display a graphical item on a computer display, and to manipulate it with the hand. Two domains started simultaneously from this pioneer's experiment: the domain of computer graphics and the domain of the computer interactivity defined as the analog control of visual synthetic shapes by hands. Until that pioneering concept appeared, action inputs linked to graphical outputs did not exist, since images were only printed. This means that, since the action appeared as a computer input, it has been considered as "naturally" and closely linked with vision, so much that the action was called « graphical input », (i.e. the computer input provided by position and displacement hand sensors as light pen, tablets caked, and also graphical tablets, etc.).
At the same time, Douglas Engelbart [3] invented the first "mouse" (1963) at the Stanford Research Institute, patented in 1970, and a few years after (1966), I. Sutherland built the first *Head-mounted Three-Dimensional Display* [4]. The 1rst January 1970, Daniel Vickers [5] reproduced the canonical sketchpad experiment (wired visualization of a cube) on a helmet-mounted display controlled by the motion of the head.

These technological acceleration led in 1972 to the set up of the first meeting preparing the creation of SIGGRAPH, Special Group of Interest on Graphics of ACM and the SIGGRAPH Conference in June 1974 by the First annual SIGGRAPH Conference, precisely called Conference on Computer Graphics and Interactive Techniques".

This pioneer phase boosted an intensive economical activity in the field of the computer graphics, leading to a standardization process 1977-1979 [6] [7] [8] [9], aiming at the definition of a standard basic graphic software, from the devices drivers (printers, plotters, display) to the basic graphics library and visualization architecture. It was surprising that the criteria that rose appassionato discussions and that was finally used to choose the winning proposition among others, was precisely the status of the graphical inputs in the graphical basic software: did the graphical inputs be internal or external functions of the graphical software (i.e. opened to the user programming)? The final choice was to be « internal functions » [6] [10], reinforcing the concept of a natural link between hand and eye.

This choice opened the way for another technological and economic revolution on displays. At that time, the market of electronic displays was dominated by companies as Tektronix and Hewlett-Packard, with a CRT (Cathodic Ray Tube) display technology based on scan-line technology (similar for oscilloscopes and plotters). The electronic manufacturers did not understand, and did not agree with this kind of interactivity. Thus, the raster technology, which was mature in research laboratories [11], made a breakthrough (1980's) on the market place, and was immediately used in the implementation of new standards. The concept of "graphics" has been improved by the concept of images. These companies moved out to their native field of instrumentation displays.

Remark: In CRT technology, the accuracy of graphics drawing could very high but the time necessary to display the image depends on the complexity of the scene. In raster technology, all the pixels are displayed whatever the image is; this limits the number of pixels and consequently the quality of the image, but the displaying time is constant, leading to another standardization in the visualization process: the so produced images were able to be seen in a same way whatever the display was. These correlated two revolutions in the hand-eye interactivity and in visual display, leaded to the current concept daily used of the mouse basically connected to the visual display.

1.3 The first post – pioneer phase: from Interactive Computer Graphics Design to Computer Graphics and HCI

A second consequence of this double technological revolution in computer action inputs and computer visual outputs, is the fact that two scientific domains had born, each one improving from the two parts of the pioneer's process: Science and technology in Computer Graphics science and science and technology in Human-Computer Interaction.

- The SIGGRAPH conference focused on research and developments in graphics and image synthesis in all their features: shapes, motions, light modeling and rendering, optimization processes, displaying, etc.
- The ACM SIGGHI started in 1982 with the SIGGHI Conference "Human Factors in Computing Systems".

In Computer Graphics, a huge amount of works have been preformed, leading to an incredible technological and scientific breakthrough in tools and methods of modeling and rendering the spatial shapes and visual effects. They were guided quite unanimously under the banner of the keyword "realism", understood as morphological and photometrical faithful representations of the real shapes and visual effects.

<u>Note of the author:</u>
To avoid misunderstanding about this term, I suggest considering this phase of revolutionary productions in Graphics as the quest of a technological feasibility. In the tri-partition (neutral level, poïetic level and aesthesic level) proposed by the Molino-Nattier's semiology [12a,b,c], it should correspond to the "neutral level". It would correspond to the "neutral level" in the tri-partition proposed by the Molino-Nattier's semiology [12a,b,c] (neutral level, poïetic level and aesthesic level), A similar example is the development of signal processing theories and technologies, which started with the demonstration of Shannon's theorem, used currently nowadays in several fields (computer music or speech synthesis) at the poeïtic and aesthesic level. Thus, we can intend "realism" in a similar sense that we could say Shanon's conditions are "realistic" in a way that they give conditions to produce digital signals identical to real signals, whatever their producing cause and content meanings are.

At the same time, HCI domain developed intensively and quite exclusively the concept of action-vision link. The famous corner stone was the WIMP concept (Windows, Icons, Menus and Pointing device), conceptualized by Xerox in 1972 and implemented in the Alto machine in 1973, which was the first computer to use the desktop metaphor and graphical user interface, leading to the Apple "desktop without paper" in 1985, with the famous "apple mouse" and implementing the "vis-à-vis" concept.

Progressively, each domain extended its developments. Computer Graphics re-introduced the user's interaction in the manipulation of synthetic 3D objects as in the very active flight simulators in which very large 3D landscapes were interactively displayed in real-time by means of large hardware graphical boards [13]. HCI improved the metaphors of interaction with 3D representations and opened the fixed graphical inputs process by adding a lot of input devices, as illustrated in the Super Cockpit project (1986-1989), which fed a lot of development in interactive environments [14a,b,c], composed of a non limited panoply of sensors inputs and a non limited panoply of metaphors of interactions including real and synthetic 3D and iconic presentations, and implementing the "immersion" concept.

As at their beginning, the domains of CG and of HCI are again closely linked and have now to be explored together in the design of new information interfaces technologies. The common point is the "magic carpet metaphor" (fly and see, move and see), used as well in navigation in 3D spaces as in iconic or data spaces.

1.4 The second post-pioneer phase: From CG and HCI to VR

The new convergence between CG and HCI, represented in **Erreur! Argument de commutateur inconnu.**, triggered what can be considered as the third and contemporary stage in the action-vision link in computer tools: Virtual Reality. Started under the name of "artificial reality", coined by M. Kruger in 1983 [15], it appeared with two different orientations discriminated by the choice of the human's position in such tools : the "vis-à-vis" position or the "immersive position":

- Following the Sutherland's approach, J. Foley [16a, 16b, 16c], a renowned researcher in Computer Graphics, adopted the "vis-à-vis" point of view, and introduced force feedback devices in Computer Graphics. (Note that in HCI, the daily used "display and mouse" technology implements the vis-à-vis concept). The meaning brought to AR by J. Foley refers to an instrumental approach, that is the use of an object within reach ("à portée de main") as an instrument to perform an external task and can be currently considered as synonymous of Virtual Worlds.
- J. Lanier [17] coined the term "Virtual Reality" in 1988 and M. Kruger [15] started with the concept of "immersion". Keeping out the understanding of VR initiated by the data glove, data suit and head-mounted display as tools to completely isolate the real human of the real world and as a completely reconstructed world to replace completely the real world[1], this meaning of AR or VR is often synonymous of Virtual Environments in the sense of worlds surrounding the user and being explored by him.

At that point, we are able to assert that the domain of "Information interfaces technologies" and the domain of Virtual Reality are closely linked and have to be questioned together.

The historical evolution presented here shows that the concept of "immersion" and the concept of "vis-à-vis", which are fighting at some economical level, (J. Lanier explicitly presented his data glove as a competitive product of the mouse), are probably conceptually and technically complimentary. The complimentarity is currently not completely elicited. Vis-à-vis and Virtual worlds take place in instrumental applications such as surgery, assisted manual tasks learning, etc… or in tele-manipulated tasks, in which the relevant questions are the manual skills and the complex behaviors of manipulated objects (deformations, transformations as cutting, blending, growing, etc.). Immersion and Virtual Environments trigger the intensive development of tools such as CAVE, in which the involved processes are the modeling and visualization of large scenes with their correlated cognitive and perceptual question of the co-location, which appears to be critical with the un-succeeded attempt to introduce force feedback objects manipulation. The elicitation of the relevant features of each concepts,

[1] Virtual reality as a drug, virtual reality as experimental platform to study the aliened and altered states of the consciousness, etc…

in the aim to draw the future of the technologies and the uses, supposes to ask them in deeper, technologically and cognitively.

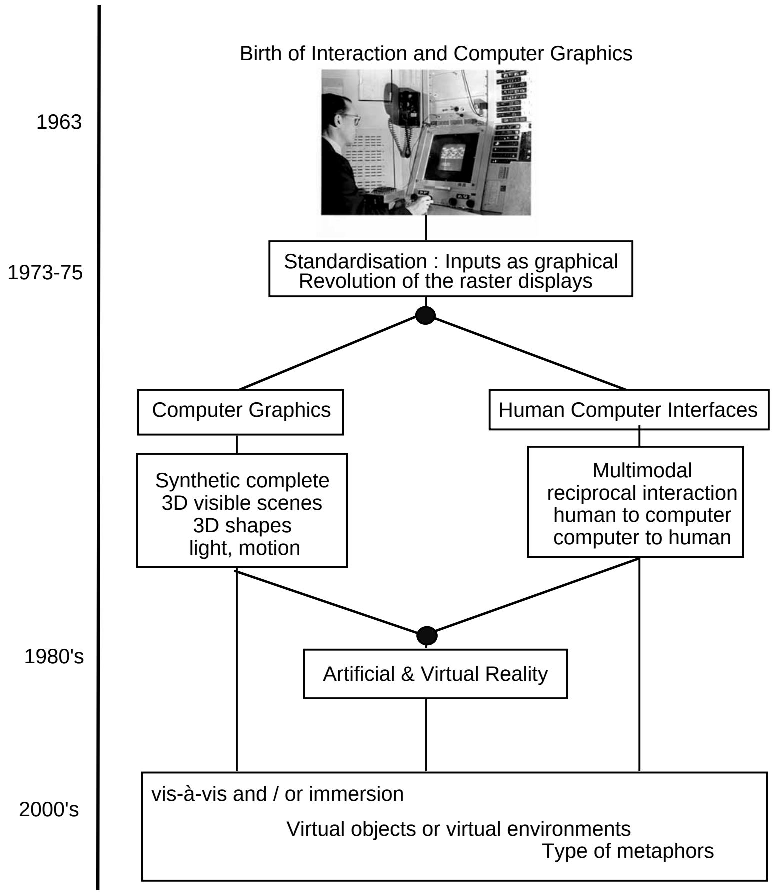

Birth of Interaction and Computer Graphics
1963
1973-75
Standardisation : Inputs as graphical
Revolution of the raster displays
Computer Graphics
Human Computer Interfaces
Synthetic complete
3D visible scenes
3D shapes
light, motion
Multimodal
reciprocal interaction
human to computer
computer to human
1980's
Artificial & Virtual Reality
2000's
vis-à-vis and / or immersion
Virtual objects or virtual environments
Type of metaphors

Figure Erreur! Argument de commutateur inconnu. **– The CG and HCI convergence**

1.5 Parallel evolution: From teleoperation, telecommunication and telepresence to VR

At the same time as the evolution of visual tools of Computer Graphics for representation and the interaction with computers that support these tools, the link between action and vision was natively asked in tele-operation and tele-manipulation and more recently in tele-communication. Teleoperation introduced the separation between the user's space and task's space, two space being "distant" in a large sense of "distant" in space or in nature [17][2].

1.5.1 Tele-symbiosis - Telepresence

The tele-operation process was the first to address the question of presence. Vertut [18] coined the term tele-symbiosis in the tele-operation context in 1974.

An explicit problem of Presence occurs whenever human beings manipulate real objects, directly or indirectly through mechanical instruments or when humans communicate in tele-communication through signals provided by real objects, directly or indirectly through sensors (microphones, telephones, cameras, etc.). Since 1950's, the manipulation of dangerous materials, such as nuclear materials, required a distant manipulation setting two different spaces up: the user's space and the task's space. As long as the manipulation remains mechanical, i.e. as long as the two spaces are near in

[2] Tele-operation, tele-manipulation and tele-communication means material operation or symbolic communication between distant worlds, in a piece of time, i.e. with no need of memory in the teleprocess. Usually the term "distant" means distant in space. In Luciani and al. [17], the notion of distance can be enlarge to worlds that are not accessible immediately to our senses: distant in space (far away as a distant planet), in scale (at a larger or upper scale that the world at our scale, called "macroscopic scale"), in nature (with different laws of physics as world under the nanoscale, chimical, electrical, but also mathematical (virtual)).

space, in time and in nature, there is no problem of Presence. The experimenter manipulates the block of nuclear matter through a mechanical pantograph, feeling it mechanically and seeing it through the glass that separates the two spaces. When this direct physical communication is replaced by electrical communication between the two spaces, and when the both spaces become more and more distant, the immediate and trivial presence disappears.

1.5.2 Moorings of teleoperation and VR

With the separation of the manipulation space in both spaces described before, the classical teleoperation instrument has been decomposed in three parts: the part which is in the user‘s space, the part which is in the task’s space and the communication between them. Establishing an appropriate communication between these two different worlds means correctly equipping each part of the communication chain.

As stated by Luciani [19], firstly, both sides were equipped with pairs of actuators and sensors that work together, with sensors on one side and corresponding actuators on the other side (Figure 2) and vice-versa: from microphones to loudspeakers, from cameras to displays, from mechanical sensors to register the user’s actions to mechanical actuators to perform these actions. These pairs of actuators and sensors are dedicated for each basic human sensory and motor apparatus: vision, audition, and action. Thus, the human representation of both realities is split into different pieces that are clearly segregated, according to the transducers used: hearing by means of a specific device, seeing by means of another device, and action by yet another. Once the representation of these two realities is conveyed by separate signals on each side, layers of signal processing are inserted for each part in order to reconstruct one space in the other. As long as we could have a good mental representation of the distant space, as long as it remains an “alter ego” space, this reduced information is sufficient to restore the distant space. But when the real phenomena cannot be sufficiently reproduced, a third module is progressively inserted in order to reconstruct in real time the lost information, that is, typically, a computer synthesis system which handles the creation of the unknown information by inserting virtual entities on each side (virtual objects, virtual humans, etc.).

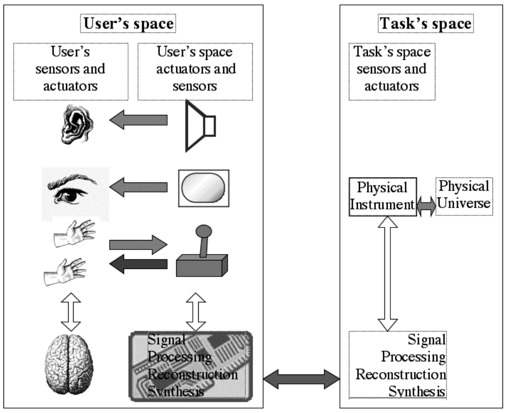

User’s space
User’s sensors and actuators
User’s space actuators and sensors
Signal Processing Reconstruction Synthesis
Task’s space
Task’s space sensors and actuators
Physical Instrument
Physical Universe
Signal Processing Reconstruction Synthesis

Figure **Erreur! Argument de commutateur inconnu. -** The complete Teleoperation -

Telecommunication chain.

At this point, note that we obtain a similar platform on both sides (Figure 2), composed of pairs of sensors and actuators corresponding to all the sensory-motor capabilities (for the human on one side, for the physical object on the other) complemented by real-time simulation systems, including signal processing from and to the alternate distant world and virtual representations that are completely built. We can also remark that this platform is precisely what is usually implemented in Virtual Reality (VR) systems enhanced with Augmented Reality (AR) functionality, and creating thus a Mixed Reality (MR) architecture. This Mixed Reality Architecture can be seen as a generic component that will equip symmetrically the user's space and the task's space, leading to the specification of a general common architecture of all our instruments.

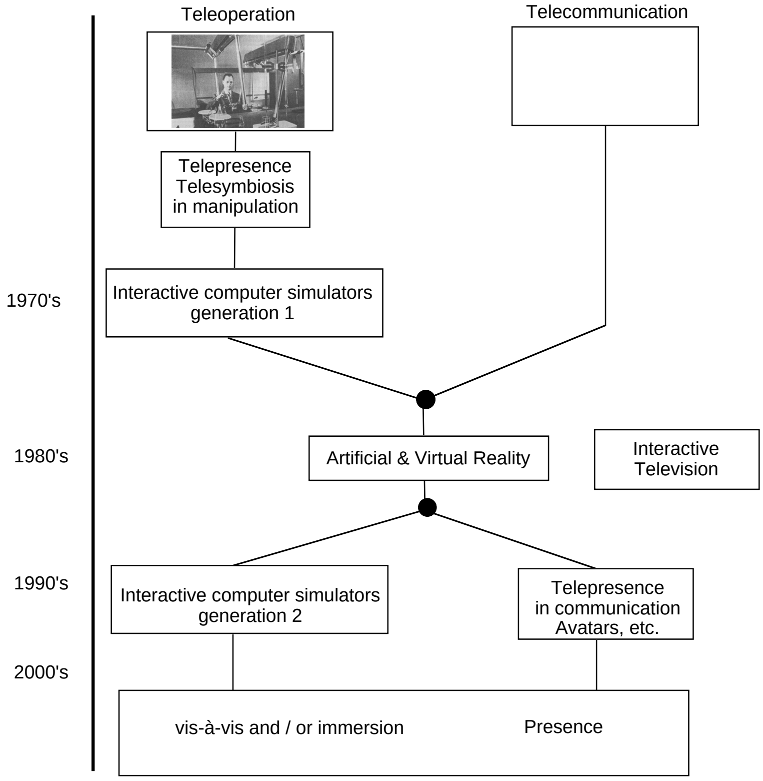


Figure Erreur! Argument de commutateur inconnu. **- Anchorage between teleoperation-telecommunication and VR**

This analysis shows the historical anchorage of tele-operation and tele-communication with the VR (AR, MR). As CG and HCI, these domains exhibited then the two different concepts of immersion and vis-à-vis, which seems more and more a relevant axis of analysis (**Erreur! Argument de commutateur inconnu.**), both stamped by the question of Presence of one world in the distant (physically, sensorially, cognitively) other.

1.6 New issues opened by convergence

The contemporary convergence between the four a-priori separated domains of CG, HCI, tele-operation, tele-communication has been boosted by the very fast development of Virtual Realities technology, which is nothing else but the cooperation between general real-time computation (simulation, signal processing) and a panoply of input-output transducers designed to act on and perceive. It leads to new technological development and uses and to the elicitation of two fundamental questions:

• Immersion and/or vis-à-vis: these two concepts underlie two completely different –but perhaps complementary– approaches of the relation between humans and objective world, virtual or not. They convey two different ways of interaction supported by two types of (1) metaphors and (2) technical objects (i.e. instruments):

- Type of metaphors: Two types of metaphors are associated to the duality immersion / vis-à-vis, “move-&-see” / “take-&-see” metaphors.
- Type of instruments: immersion leads to represent large scenes to be explored and vis-à-vis leads to represent handled objects.

• Presence: the break with the ontological link between the objective and subjective worlds caused by the electrical non-sensorial communication, leading to a generalized mediation of the object’s manipulation and of the human communication, triggers with a new force the renewal of the question of Presence, again shared in two dual meanings of “being there” [20] and “being with” [21].

1.7 References

[1] I.E. Sutherland. Sketchpad: A Man-machine Graphical Communications System. Ph.D. Thesis, 1963. Mass. Institute of Technology

[2] I.E. Sutherland. *The Ultimate Display*, Proceedings of the IFIP Congress 2, 1965

[3] D. Engelbart. To be completed

[4] I.E. Sutherland. *A Head-Mounted Three-Dimensional Display*, Fall Joint Computer Conference, 1968

[5] D. Vickers. To be completed

[6] R.H. Ewald, R. Fryer. Final report of the GSPC. Computer Graphics Quarterly Report of SIGGRAPH-ACM. Vol 12, n°1-2. June 1978.

[7] R.A. Guedj. Seillac Seminars I. Methodology in Computer Graphics. IFIP Workshop. May 1976.

[8] R.A. Guedj. Seillac Seminars II. Methodology of Interaction. IFIP Workshop. May 1979

[9] Satuts report of the Graphic Standards Planning Committee. Computer Graphics Quarterly Report of SIGGRAPH-ACM. Vol 13, n°3. August 1979.

[10] P.J.W. Ten Hagen. Interactive techniques. Eurographics tutorials 1983

[11] Raster displays. To be completed

[12a] MOLINO, Jean, 1975, "Fait musical et sémiologie de la musique", Musique en jeu, 17, Paris, pp. 37-62.

[12b] NATTIEZ, Jean-Jacques, 1975, Fondements d'une sémiologie de la musique, Paris, U.G.E.

[12c] NATTIEZ, Jean-Jacques, 1987, Musicologie générale et sémiologie, Paris, Bourgois.

[13] CG –flight simulators – real time boards. To be completed.

[14a] Furness, T., "'Super Cockpit' Amplifies Pilot's Senses and Actions," Government Computer News. August 15, 1988, pp. 76-77.

[14b] Furness, T., "Helmet-Mounted Displays and Their Aerospace Applications," National Aerospace Electronics Conference, Dayton, OH, May 1969.

[14c] D. Underwood. "VCASS: Beauty (and Combat Effectiveness) Is in the Eye of the Beholder," Rotor & Wing International. Vol. 20, no. 3, pp. 72-73, 107, Feb., 1986.

[15a] M.W. Krueger. Responsive environments - Proc. National Computer Conference, p. 423-433-1977

[15b] M.W. Krueger, Artificial Reality, Addison-Wesley, 1983.

[16a] James Foley, "Les communications entre l'homme et l'ordinateur", Pour la Science, décembre 1987

[16b] Foley, J. D., 1987, Interfaces for Advanced Computing, , (4), 126-35. *Scientific American* **257**

[16c] Foley, 1987. Manipulative interfaces.

[17] J. Lanier. A Vintage Virtual Reality Interview. *Whole Earth Review. 198.8*

[18] J. VERTUT, Ph. COIFFET. Téléopération : évolution des technologies. Hermes éditeur - 1986

[19] A. Luciani, D. Urma, S. Marlière, J. Chevrier. PRESENCE : The sense of believability of inaccessible worlds. Computers & Graphics. 2004. Vol 28/4 pp 509-517

[20] G. Riva, F. Davide, W.A. Ilsselsteijn editors. Being There : Concepts, Effects and Measurements of User Presence In Synthetic Environments. IOS Press. 2003.

[21] IST-2001-38040-FP5. TOUCH-HapSys : Towards a Touching Presence : High definition Haptic Systems. project. www.touch-hapsys.org

2 Action-Vision fusion in CG, HCI, VR, teleoperation, telecommunication : Novels issues

©Annie Luciani, INPG

2.1 New Technological revolution in the basic components

As it happened in the pioneer's phase with the double revolution of "new needs" (Sutherland and Graphics standards) and of the novel technology of displays, we have to take care to take into account? the current three breaking technological (re)evolutions:

• The revolution in the technology of the extra flat and portable displays: a lot of EU projects in FP5 are related to extra flat and portable displays that are not based on the raster technology (plasmas, very low energy, etc.). Nowadays, the panoply of displays goes from very large displays (the size of several bodies) to very small displays (the size of the hand), through the canonical raster or plasma display (the size of the head-arm). As in the seventy's, the main factor of is not in the display itself but in the standardization of the uses and representations.

• The exponential progression in chip integration and in the associated increase of the computational power of chips, as planned by the Moore's law [22a] up to 2020's, has two correlated consequences: the increasing of portability and the increasing of computational power. Available reports on the ITRS web site (International Technology Roadmap for semiconductors) [22b] update all the planned evolution in semiconductors (process integration, technologies for wireless communications, emerging research devices, front end processes, etc.).

• The exponential progression of the networking (terrestrial or Hertzian's networks).

According to three technological shifts, we may envision the possibility to make wearable computers and tangible objects (disappearing computers) equipped with communication and interaction tools so that to be enactive interfaces.

2.2 Immersion vs. vis-à-vis and vis-à-vis vs. prosthesis

As the evolution presented bellows shows that the duality between immersion and vis-à-vis concepts goes accross all the referenced domains (CG, HCI, teleoperation, telecommunication, VR) and can be chosen as a structural axis to analyze the different types of action – vision relationship in information interfaces technologies.

The elicitation of the reasons of this duality and of the means of their cooperation will be probably one of the major challenges for the next years that would spur new technological evolutions and shifts in novel interfaces.

2.2.1 Immersion and vis-à-vis : similarities and differences

2.2.1.1 Immersion: or humans inside a world

Immersion focuses basically on the seeing (or hearing) sense. The related actions are thus spatial actions such as displacements of the own body itself: that is an "observational situation", implemented in the computer by metaphors such as "magic carpet", "fly and see", "move and see", etc. These are exploratory metaphors used in VE navigation as well as in flight or driving simulators, landscapes or cities' navigation, etc. In such cases, the "immersive situation" seems to be a natural and common. Basic correlated questions are similar in the real spatial world and in virtual or abstract worlds. Both of them raise the difficulties (1) to plan step by step the displacements to reach the goals and (2) to memorize a spatial reference to locate at each time where we are and how do we reach. Nevertheless, some drastic discrepancies rose. The most important of them is that in VE, the human body does not move. Movements are instrumented by means of an intermediate real object (stick, wheels, balls, travelators, etc.) assisted by a virtual one (virtual arrow, virtual camera, etc). Thus, a physical transformation between the localization and displacements in real world and their effect in the virtual world is introduced. This transformation leads to the design of adapted metaphors and to study their effects on human's capabilities. One of them are those related to the question of "co-location" (see the G. Jansson's State of the Art and [23a, 23b, 23c], referred by G. Jansson. Furthermore, the immersive situation remains conceptually problematic. From the point of view of manipulation, it is a kind of teleoperation: human manipulates a tool in human space that has an effect in a task's space, i.e. as a kind of vis-à-vis situation. From the point of view of seeing, it is an immersive situation in which the space is moving around the human body.

2.2.1.2 Vis-à-vis: humans in front of the world, or the world within reach ("à portée de main")

The vis-à-vis situation is related to manipulation activities. It refers to objects that are in a local space, i.e. hand or body's attainable objects. It supports the functional transformation from an object to an instrument as a usable object to do something, and further the functional transformation of an instrument as external object or as a prosthesis, i.e. as a part of the body.

In the vis-à-vis situation, the relation between the action and the sight is deeply different compared to the immersive one. During the immersive activity, seeing is mainly (even if it is not only) the aim of the current action (move in order to see). Conversely, in the vis-à-vis situation, seeing is mainly a way of controlling the current action (put here, shock, write, etc…).

This analysis shows that the concepts of immersion and of vis-à-vis have to be considered not only as competitive but also as complementarily operational concepts studying the aim of studying and instrumenting the relation of humans to world. It points three progressive different scales: (1) from "outside-far way", (2) via "close to the body", (3) to "in contact with the human body":

(1) "outside-far way": far in spatial distance with the predominance of the space and the geometry of the space and the predominance of seeing,
(2) "close to the body": defining possible manipulated objects "à portée de main", with a balance between space and geometry on one hand and physics and materiality on the other hand,
(3) "in contact with the human body": with the predominance of the materiality in the experience and the use of such objects to experience the fluent and permanent transformation between objects that remains cognitively external and prosthetic objects playing as a part the body.

2.2.2 Proposal to a categorization between the immersive and the vis-a-vis situations: The tri-partition "Environment / object / instrument"

This progressive transformation of the physical and cognitive status of the external universe can be operationally schemed by the three following proposed words: **environment**, **object**, **instrument**. We call :

- "environment" the set of objects in which "the body is embedded",
- "object" something that can "be taken", i.e. physically and cognitively "à portée de main",
- and "instrument" an handled object used to act on or with.

Examples: the wall, the ceil and the floor are objects of our environment that we cannot (or we are not in situation to) modify them as objects. They surround the body. The door belongs to the environment since we open it. Thus it becomes an "object", and when we slam it to express our anger, it becomes an "instrument". The pencil on the desk of another people belongs to the environment (it can be behind us). When it is on our desk, it becomes an "object" that can be used as an "instrument". The piano in a room belongs to the environment, being a furniture of the room for the visitor that is not a pianist. It becomes an object for the pianist before playing and a prosthetic instrument for the confirmed pianist during his play.

The two last stages of the status of an external material thing of the world correspond to the distinction between the action preparing the acting on (for example the pre-grasping, reciprocally the moving to intercept, the pre-percussive gesture for a percussionist when he positions his stick and approaches of the thumb with a given velocity) and the action during the "acting on", the grasping, the shocking of the ball of the surface of the thumb.

In such progressive transformation, the frontier between the purely immersive phase and the purely manipulation phase is the "close to the body" phase, which appears as a bi-faced phase (Figure BBB): objects can be considered as a part of the environment or as to be manipulated. This stage is cognitively an interface between the two other.

Cadoz and al. in [24a] [24b] proposes an operational criterion which defines precisely the third phase, (operational means: usual to steer the design of technological interfaces). This criterion is expressed as following: if the action is not necessarily encoded in the final perceptual (visual or auditory) result of the action, then the situation can be considered as a non-manipulatory situation. Let to take to examples: in the pointing gesture ("look at here"), the visual result of the action "the pointing" for who is pointing and "the look at" for who is looking at, does not depend on the effective cinematic of the motion used to point the target. Conversely, when we mould a past or when we "drive in a nail in a wall", the result (visual, auditory and, more, physical) depends on the evolution of the action (of its dynamics). The entire action (the way in which it is performed at each instant) is encoded in the result.

Putting face-to-face the clear definition of the two extreme situations, the "purely immersive" one ("fly and see") and the so defined "purely manipulatory" one, Cadoz in [24a] [24b] proposes a tri-partite typology of actions, called by him "gestures". Despite action is more general than gestures, action can address the goal of the performance. Gestures address only the performance:

- purely "ergotic gesture": this word has been invented to be more precise that "manipulation gesture" to avoid misunderstandings with the polysemy of that term. Ergotic gesture is the gesture during which a physical energy is exchanged between the two bodies (human and object). It needs contact and it results to the correlated physical modification (more or less durable) of both.
- modification gesture: it is the gesture with which we modify the conditions of the first during its performance.
- and selection gesture: it is the gesture that modify the conditions of the two first before their performance.

Cadoz called "instrumental situation", the interaction situation including necessarily a manipulatory phase as the aim of the action.

Examples of instrumental situations:

- during manual writing: the writing is of the first, the positioning of the sheet by the other hand under the pen is of the second and the selection of the pen or the selection of the location of the writing in the page is of the third.
- During the play of a violin: the bowing gesture (the friction of the bow on the string) is of the first, the modification of the length of the string to change the pitch of the string is of the second and the selection of the string among the four is of the third
- in a tennis playing : the shocking of the ball is of the first, the motion to position the body to intercept the ball is of the second, and the choice of the racket and of the ball is of the third.

And he defines an "instrumental situation", first as being necessarily a vis-à-vis situation, second as it exhibits necessarily an ergotic gesture, and third as composed by these three types of interactions.

This categorization can be confronted to Guiard analysis of bi-manual tasks, [25a] referenced by Joan De Boeck and al, in their WP4b State of the Art, [25b] or J.P. Gaillard [25c].

Conversely, non-instrumental situations have no need of actuation ergotic gestural interaction. They are only composed by modification and selection gestures. Immersive situations are necessarily non-instrumental situations.

2.2.2.1 Examples of non-instrumental situations

Immersive situations: the relationship with a surrounded environment, since it remains "surrounded", such as "moving your body and see", the free body motion is of modification motion: we modify our position or the position of the landscape. The choice of the landscape or of the part of the landscape to be explored is a selection gesture. There is no ergotic gesture: the landscape has not to be physically

modified. To modify the landscape, we have to change the cognitive status of the landscape to consider it as in an object in vis-à-vis.
In his WP4b State-of-the-art, G. Jansson [26], states that some participants had problems to visualize a 3D stereo object and focused on the front wall instead of the 3D position of the stereo model. A planned solution was to place the haptic interaction as close as possible to the projection wall. This observation suggests that such people prefer an instrumental vis-à-vis situation. The 2D physical screen is considered as an ”object a portée de main”, as a paper sheet with which we can have physical interaction. And it leads to a novel question: is the cognitive style of people is defined with regards to immersive or vis-à-vis situations? Shifted according to the Cadoz’s analysis: is the cognitive style of people is defined regarding instrumental or non-instrumental situations? And the correlated question of co-location can be asked in another ways: Why and when is visuo-action co-location needed to avoid discrepancy in interaction comparing virtual world and real world or to improve the performance of the task? Are these questions similar in immersion and vis-à-vis situations? Are these questions similar in instrumental and non-instrumental situations?

Several people and several teaching methods in manual learning (instrumental musical playing, animation of objects, sports), some of them being conducted (but non published) by Luciani-Cadoz-Florens, relate that in the ergotic relation, the gesture is more accurate and fast if we don’t look continuously our hands performing the task. This means that vision is used for the modification and selection gestures during the performance of instrumental activity.

From the point of view of human-computer interfaces and more generally of electromechanical interfaces, this categorization in instrumental and non-instrumental interaction is operational, at least because it recovers two different scales of temporal delays between action and vision. Le Runigo and all state in their WP4b State-of-the-Art [26b] that the interceptive actions supports an occlusion delay of about 100-200ms. Florens and al evaluated empirically from force feedback real time simulation of physical objects that the manipulatory phase in instrumental situations requires to sample the motions of the bodies at least at 1 ms (1Khz) or less: 100Hz in Berberyan’s PhD Thesis in 1882 [27a], about 700-800 Hz in Cadoz 1990 [27b], 200 to 1500Hz in Florens 1991 [27c] and Luciani 1991[27d] and 200 Hz to 4 Khz for Uhl 1995 [27e].)

We can associate these two different temporal scales to another result related to the cyber-sickness [28]. For the vision only, (or for the non-manipulatory interaction between action and vision as when we manipulate the mouse moving on the display on a computer desktop), a refreshing rate of 25-30 Hz for the visualization as well as for the sampling of the motion of the mouse are commonly implemented without any noticeable disease for the user. But this sampling rate seems not to be sufficient in simulators, causing the cyber-sickness and requiring to increase of the refreshing rate of the gestural input (towards 1 Khz) and of the visual output. Currently the refreshing rate in computer displays used in VR and simulation is from 70 to 120 Hz and the delay between hand control and visual display is at the visual rate (cf. paragraph “Temporal delays inaction-vision loops”, [Allison 2001][Frank 1988]).

According to this, it seems that this frontier between purely manipulatory and purely non-manipulatory interaction has to be explored as a critical criteria to specify new interfaces, to specify technological requirements of interfaces, as well as to understand human cognitive features in the human-world interaction.

2.3 The disturbing arrival of force feedback device

The force feedback devices have been introduced first in teleoperation [18][29b][29b] and in clearly defined instrumental applications as musical playing and animation [33a,b]. Both are explicit instrumental situations including necessarily direct a physical manipulation of a physical object. No major questions rose, except those of Presence, telesymbiosis [18], or telepresence, discussed after.
Conversely, they are the last components that have been integrated in human-computer interaction and in Computer Graphics, rising several new questions, the main of them being precisely: When does the

force feedback be considered as a necessary component in the human - (virtual or real) world interaction?

Compared as stated in the beginning of this paper, the Computer Graphics and HCI domains started with the middle stage of the vis-à-vis situation, considering “objects” rather than “environments” or “instruments”. Computer Graphics started to model objects rather than large surrounded scenes and HCI started to define interactive icons, (i.e. as possibly manipulated objects). Both cases are implemented without ergotic interaction. Conversely, tele-operation started earlier with force feedback and ergotic interaction, even if it was unfaithfully rendered in the first simulator generation. Before 1995’s, force feedback devices ran currently at 50-100 Hz [18] [29a].

Since its beginning up to now, the manipulation of objects in Computer Graphics remains under the form of displacements (translation and rotation). Obviously, for HCI icons the manipulation is also restricted to displacements.
In both cases, from this intermediate state, works have been naturally extended to:

- on one side : manipulation tasks and instrumental situation
- on the other side : non manipulation and non instrumental situation leading to the development of immersive systems.

The arrival of the force feedback devices in CG and HCI [30][31], causes waves of enthusiasm accompanied by passionate clearly-cut positions. About a half of the scientific actors consider that they have to be integrated in interfaces and work to find arguments for and the other half consider that they are not or they cannot be used and work to find arguments for.
For the first ones, it is a new channel of perception among both existing channels of audition and vision [29] and it thus introduces a new feeling of objects in non sufficiently sensible computerized worlds. Atkinson [32] untitled his visionary paper “Computing with feeling”. Cadoz, Luciani, Florens, [27d][33a, b, c, d] started by considering force feedback as a necessary (impossible to avoid) component in expressive and dexterous tasks as musical performance and animation control.
The seconds argue of:

- the supplementary non negligible costs (financial, technical in terms of new developments and in terms of know-how in their uses, computational)
- the user needs and the possibility (or not) to encode what it is conveyed by mechanical feedback (the information perceived by the mechanical human body through all his apparatus of mechanoreceptors, whatever they are) in non mechanical ones (visual or auditory).

These clear-cut and well-justified positions indicate that force feedback it is not only a new way of interaction, - even if it obviously is. It forces to leave the unclear intermediary level of object and compels to understand the cognitive and technological transformation occurring in traversing the frontier: from (and to) the distant immersive environment to (and from) the object as instrument and forward as a prosthesis: from the object as stimulating the exteroceptive senses only to the object as mechanical linked to the body. And the correlative question is: is force feedback always necessary, in immersive of non-immersive environments? or not? and then, for what situations?

Cadoz’s answers is [24a,b]: the force feedback is a necessary (but not sufficient) and a non-avoidable component in purely ergotic phase in the instrumental interaction. Consequently to his typology, he proposes to consider the gestural computer inputs space needed for the instrumental vis-à-vis interaction (the interaction with the proximal space) as composed of a panoply of devices, one type per category of interaction: force feedback devices for manipulation actions, analog sensors for modification gestures and discrete sensors for selection gestures.
A subsequent conclusion is that there is no fundamental needs to introduce force feedback in immersive situation. And a subsequent issue is: how, from a technological point of view, is it possible to introduce instrumental situation needing as a limit, force feedback manipulation immersive environments, i.e. With what kind of concepts and techniques, can we build systems in which objects are sometimes immersive and sometimes in closed vis-à-vis?

2.4 The disturbing arrival of physically-based modeling

The introduction of force feedback systems in the action sensors – visual actuators loops, leads to introduce a physically - based modeling stage in this loop (and vice-versa). Force feedback needs force generation. Force generation needs physically - based modeling (or metaphor of physically-based modeling). In Computer Graphics, physically - based modeling are the last type of techniques to appear, after the geometrical modeling and the light modeling.

Physically - based modeling is a kind of way to model dynamic systems, efficient to calculate forces but obviously also to calculate the movements as the effects of the forces. This means that the arrival of physically - based model (or physical modeling) in Computer Images hugely improved the motion synthesis. Before that, animation was only based on cinematic restitution of the observed motion. Cinematic methods are based on the representation of the motion at a phenomenological level. Their goal is the reproduction of what is observed, directly from the observation itself. There are not generative methods and they belong to the analysis-synthesis methods. The basic techniques are: mathematical description of motion through evolution functions in which the time is an explicit variable (explicit key-frames with automatic interpolation running with explicit-time, temporal functions, etc…). With these kinds of techniques, complex kinematics (dynamic changes, dynamically correlated motions, etc…) cannot be produced. In 1985's, generative models appear in motion synthesis that can be classified in two majors categories: physically based models (Luciani and al. 1984 [33c, 35a, 35b], Terzopoulos and al. 1987 [34]) and agent-based models (Reynolds 1987, [36]. Thus a breakthrough in motion models occurred and a lot of works have been achieved, due to the development of generative approaches, among phenomenological approaches.

A first conclusion is that motion modeling and synthesis are strongly correlated to the computation of non-geometrical models.

A second conclusion is that among all the generative models that produce motion, physical modeling takes a specific place. Physical models are related to the physical properties, which are the properties of the matter. These properties have to be necessarily taken into account for the rendering of dynamics (dynamics of collision, dynamics of deformation, dynamics of physical cooperation, etc.) by using exchanged forces to compute the motions and to calculate forces to render to the external physical user the feeling of the materiality of the physically manipulated objects.

Thus, physical modeling and force feedback devices are intrinsically linked: the first allows to compute the material behavior of the mechanical object and the second plays the role of a transducer which transmits this behavior to the human who manipulates.

From this point, the relation between hand and vision cannot be longer considered from a spatial and geometrical point of view. Two components come to be inserted between the passive spatial displacement of the hand and the geometrical computation for the visual rendering, aiming at the representation of the materiality of the manipulated objects: physically-based models for the mechanical gestural feedback (materiality for the hand), and physically-based models for the visual motion (materiality for the visual motion). Motion is necessarily produced by physical objects and hand is necessarily manipulated physical objects. Or reciprocally, hands never manipulate non-physical objects and motion is never produced by non-physical objects. These two modeling components support two functionalities: (1) they improve the skill of acting on and (2) they improve the visual believability of the represented objects by introducing relevant motions and relevant dynamics. Luciani argues in [19] that the feeling of presence, understood as the sense of "being with" cannot be reached without any clue, in virtual objects of computer representations of sensorial events, of dynamics, any evocation of the matter, any clue of physical energetic consistency. This is called "the concept of evoked matter".

This points the fact that the loop hand – vision, firstly considered and envisaged as natural since the 1990's, is mainly based upon the priority of the spatial on the physics and upon the vision on the manipulation, which is too simple and incomplete. The two complementary extreme situations (1) immersive & non-manipulatory and (2) vis-à-vis & manipulatory", improve the complexity of the link between action and vision by adding the role of the matter, of the body's interaction and of the motion.

2.5 Summarizing the complete technological vision-action chain via some figures:

Since the action-vision relationship has been considered at the intermediate level of objects "à portée de main", being an ambivalent state between instrument and environment, the action-vision loop in man-environment interface can be described by the following assembling of technological components (figure 4):

- ➔ sensors which sense the displacements of the body (hand, arm, etc…)
- ➔ geometrical representation of the objects, which is called in computer graphics "geometrical modeling".
- ➔ visual representation of such objects by placing them in a light field and representing the interaction between the light field and geometrical objects, called in CG light rendering.

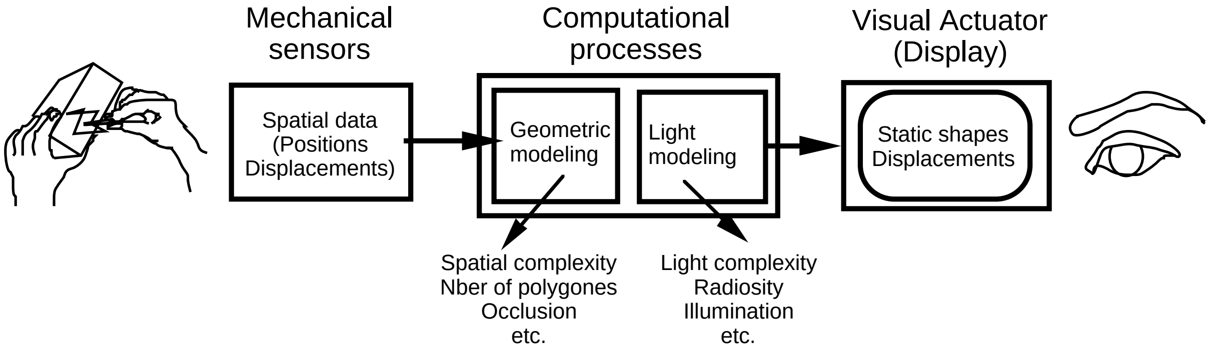


Figure 4. The hand-vision chain in conventional interaction

At the beginning stage, this set of components was the same for vis-à-vis situation (objects "near the hand") and in the immersive situation (surrounded objects). Nevertheless, the recent developments of the immersive situation push to describe scenes more and more larger compared to the local vis-à-vis standard situation, increasing the weight of the computational problems (qualitatively and quantitatively). These problems have been well identified in the Carrozino and al. in their state of the art [37].

Conversely, the orientation towards the instrumental tasks, as needed in manual learning, simulators for manipulation tasks (surgery, driving, sculpting, etc.) VR realities systems for computer music and computer animation, etc…. needs the correlated introduction of new layers of force feedback and physical modeling and hard real time simulation of physical models (figure 5).

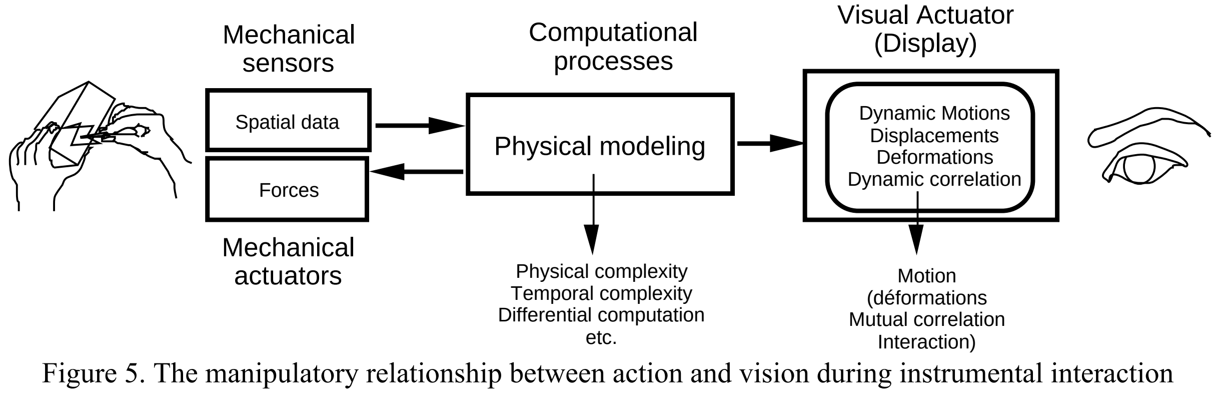


Figure 5. The manipulatory relationship between action and vision during instrumental interaction

2.6 The duality between space and matter

The figure 6 integrates all the technological components of the action-vision cooperation and points out the complexity of the chain between hand and eyes, which could be summarized as: from mechanics to optics, or from graviton to photon.

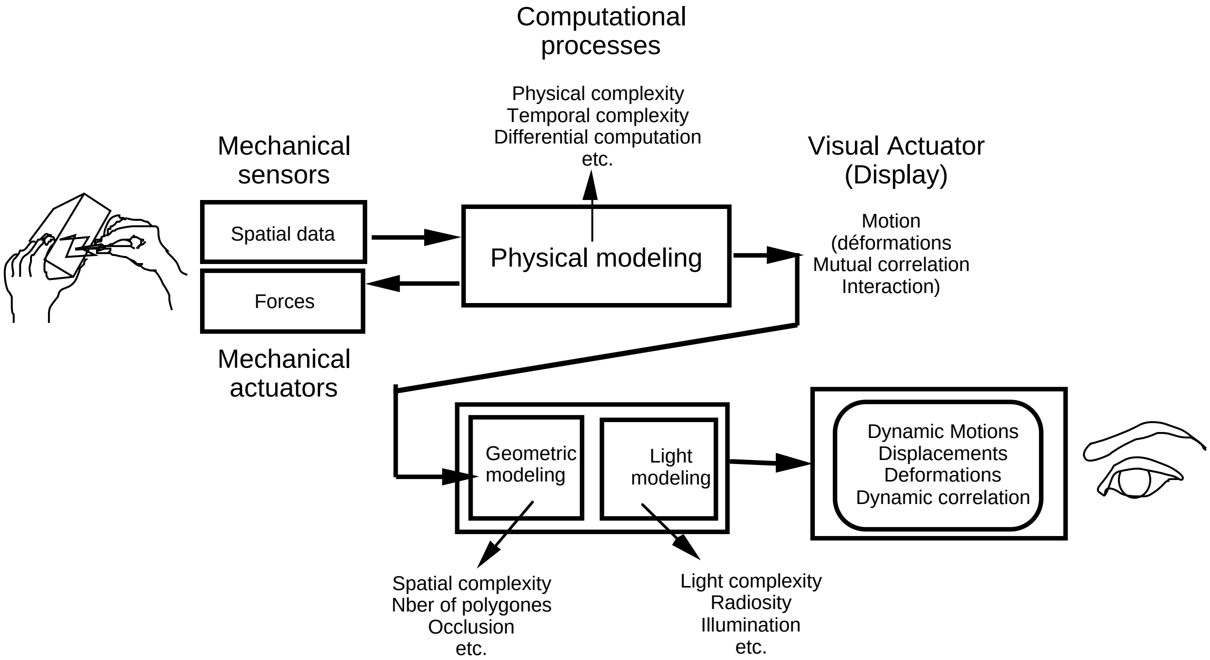


Figure 6. The complete action-vision chain

In [38], Luciani addresses the complexity of this chain by pointing out the paradoxical ambivalence of the notion of "shape", by writing "shape do not exist as single pattern affected to an object". Shape has two faces, one looking to the physical materiality of the object, one looking to optical property (Figure 7).

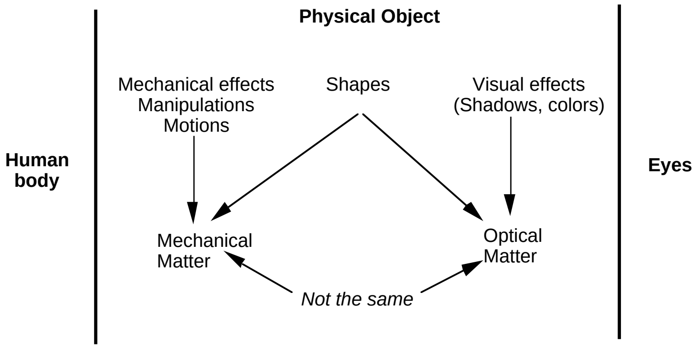


Figure 7. Geometry vs. Physics: the ambivalence of the shape.

Keeping in mind that the computational power will stop its exponential increase in about ten years, the computational complexity of the processes put between the hands (the body) and the eyes will reach its high ceiling. Drastic choices must probably be done: Have we put the emphasis on the geometrical computations (i.e. to the spatial visual features)? Have we put the emphasis on the physical computations (i.e. to the manipulatory features)?

This enlightens the passionate research choices, which are clearly separated at the moment in two lanes: Computer Graphics trends to choose the first, trying to extend the geometric approach to solve physics problems, in the following of partial differential formulation, variational methods, etc. Dynamics systems (automation, robotics, etc…) trends to choice the second, in the following of interaction, regulation and control processes and cellular dynamic automata point of view.

The knowledge on human representations seems to follow similar categorization:
• Lots of works are dedicated to the perception of shape (but what shape?). Only in very specific cases, the visual and mechanical shapes are "physically" the same.
• Lots of works are dedicated to the passive touch.
• Lots of works point out the fact that textures are encoded as well by touch and by vision.
Texture can be considered as microscopic physical shapes as well microscopic optical shape. As microscopic physical shapes, it conveys very low forces, and consequently it does not play the main role in the motion (immobility, motion and deformations) of objects during manipulation. According to this point of view, it has similar properties than optical features, leading to the definition of touch (that is of the contour surface state) as the "eye of the hand". According to a "scaling point of view, both in space and forces", texture is like a scale on which mechanics and optics can be superimposed, as the frontier between mechanics and optics. The only, but not physically negligible role of texture is that it supports friction and thus it is a main component to an object to be grasped. Mechanically speaking, texture is a macroscopic parameter that emerges from microscopic features. That is the same for the optical properties of the surface: reflectance, etc.

2.7 Conclusion

Despite the huge quantity of works related to the hand – eye interaction, in the technological fields as well as in the perception and cognitive fields, the fundamental questions are not solved. They just start to be understood and to be raised. The Enactive project is at the core part of this questioning. The Enactive project has to explore in deep the action-vision co-operation, and find paths to render it operational in Human-Computer (and machines) interaction.

2.8 References

[22a] G.E. Moore. Cramming more components onto integrated circuits. Electronics. Vol 38, n°8. April 19, 1965.
[22b] ITRS : International Technology Roadmap for Semiconductors. http://public.itrs.net/
[23a] G. Jansson.WP4b State of th Art.
[23b] Jansson, G. & Öström, M. (2004). The effects of co-location of visual and haptic space on judgements of form. In M. Buss & M Fritschi (Eds.), *Proceedings of the 4th International Conference Eurohaptics 2004* (pp. 516-519). München, Germany: Technische Universität München.
[23b] Wann, J. P., Rushton, S. & Mon-Williams, M. (1995). Natural problems for stereoscopic depth perception in virtual environments. *Vision Research, 35,* 2731-2736.
[23c] Messing, R. (2004). Distance perception and cues to distance in virtual reality. Poster at First Symposium on Applied Perception in Graphics and Visualization, co-located with ACM SIGGRAPH, August 7-8, 2004, Loa Angeles, CA.
[24a] C. Cadoz. Le geste, canal de communication homme/machine : la communication instrumentale». Technique et science de l'information. Hermes Editeur. Volume 13 - n° 1/1994, pages 31-61
[24b] C. CADOZ C., M. WANDERLEY. Gesture and Music. in Trends in Gestural Control of Music. IRCAM Editeur. 2000. avec CDROM.
[25a] Y. Guiard. Asymetric division of labor in human skilled bimanual action: The kinematic chain as a model. In Journal of Motor Behaviour, volume 19, pages 486–517, 1997.
[25b] Joan De Boeck and al. WP4b State of the Art
[25c] J.P. Gaillard - "Organes de commande en téléopération" Janv. 1990. English version ???
[26a] G. Jansson. WP4b State-of-the-Art
[26b] Le Runigo and al.. WP4b State-of-the-Art.
[27a] T. Dars-Berberyan. Etude et réalisation d'un calculateur spécialisé pour la synthèse sonore en temps réel par simulation de mécanismes instrumentaux", Thèse de Docteur Ingénieur Spécialité Electronique - I.N.P.G. - Grenoble 1982.
[27b] C. Cadoz, L. Lisowski, J.L. Florens. A modular Feedback Keyboard design. Computer Music Journal, 14, N°2, pp. 47-5. M.I.T. Press, Cambridge Mass. 1990.
[27c] J.L. Florens, C. Cadoz. The physical model: modeling and simulating the instrumental universe. Book Chapter in Representation of Musical signals. G. de Poli, A. Piccioli, C. Roads editors. MIT Press. 1991.

[27d] A. Luciani, S. Jimenez, J.L. Florens, C. Cadoz. Computational physics : a modeler simulator for animated physical objects. Proceedings of the European Computer Graphics Conference and Exhibition. Vienna, Austria, September 91, Editeur Elsevier

[27e] UHL(C), FLORENS JL, LUCIANI(A), CADOZ (C) - «Hardware Architecture of a Real Time Simulator for he Cordis-Anima System :Physical Models, Images, Gestures and Sounds» - Proc. of Computer Graphics International '95 - Leeds (UK), 25-30 June 1995 - , Academic Press. - RA Ernshaw & JA Vince Ed. - pp 421-436

[28] K. M. Stanney, R. R. Mourant, R. S. Kennedy. Human Factors Issues in Virtual Environments: A Review of the Literature. Presence, Vol 4, No. 4, August 1998, 327-351

[29a] BATTER, J.J. and BROOKS, F.P., Jr. - GROPE-I - A computer display to the sense of feel - *Information Processing, Proc. IFIP Congress 71, 759-763.* 1971

[29b] BEJCZY, A.K. and SALISBURY, J.K. - Controlling Remote Manipulators Through Kinesthetic Coupling - Computer in Mechanical Engineering, July 1983, pp. 48-60. 1983

[30] M. MINSKY and al - "Feeling and seeing : issues in force display". Computer Graphics. Vol 24 - n°2 - March 1990

[31] IWATA (H) - "Artificial reality with force-feedback : dev. of Desktop Virtual Space with Compact Master Manipulator" - Computer Graphics, vol.24, n°4, August 1990, p.165-170.

[32] W.D. ATKINSON, K.E. BOND, G.L. TRIBBLE, K.R. WILSON - "Computing with feeling" - Comput. and Graphics, Vol 2 – 1977

[33a] FLORENS (JL), 1978 - "Coupleur gestuel interactif pour la commande et le contrôle de sons synthétisés en temps réel" - Thèse Docteur Ingénieur - Spécialité Electronique - I.N.P.G. - Grenoble 1978.

[33b] CADOZ, C., LUCIANI, A., FLORENS, J.L. – Responsive input devices and sound synthesis by simulation of instrumental mechanisms : The CORDIS system - Computer Music Journal - N°3 – 1984

[33c] LUCIANI (A.) & CADOZ (C.), "Modélisation et animation gestuelle d'objets - Le système ANIMA", CESTA - 1er Colloque Image, Biarritz 1984.

[33d] LUCIANI (A), CADOZ (C), FLORENS (JL), 1994 - "The CRM device : a force feedback gestural transducer to real-time computer animation" - Displays , Vol. 15 Number 3 - 1994 - Butterworth-Heinemann, Oxford OX2 8DP UK, pp. 149-155.

[34] D. Terzopoulos, J. Platt, A. Barr, K. Fleischer. Elastically deformable models. *Computer Graphics*, **21**(4), 1987, 205-214, *Proc. ACM SIGGRAPH'87 Conference,* Anaheim, CA, July, 1987. Translated to Japanese by Nikkei-McGraw-Hill and published in *Nikkei Computer Graphics*, **3**(18), 1988, 118-128.

[35a] CADOZ (C), LUCIANI (A), FLORENS (JL), LACORNERIE (P) & RAZAFINDRAKATO (A), "From the Representation of sounds towards a Integral Representation of Instrumental Universe", International Computer Music Conference - Venise 1984.

[35b] LUCIANI (A), JIMENEZ (S), FLORENS (JL), CADOZ (C) & RAOULT (O), "Computational physics : a modeler simulator for animated physical objects", Proceedings of the European Computer Graphics Conference and Exhibition. Vienna, Austria, Septembre 91, Editeur Elsevier

[36] Reynolds C.W. . *Flocks, herds and schools: A distributed behavioral model.* Proc. of 14th Conf. on Computer Graphics and Interactive Techniques, pages 25–34. ACM Press, 1987.

[37] Carozzino and al. WP4b Percro State-of-the-Art.

[38] A. Luciani. Dynamics as a common criterion to enhance the sense of Presence in Virtual environments. Presence 2004 conference. Valencia, October 2004. To be published.

3 Geometrical, Light and Visualisation Modeling and Computation of 3D scenes

© Marcello Carrozzino (Percro), Ignazio Manza (CEIT)

3.1 Introduction

Virtual reality (VR) is a multi-sensorial experience. The basic idea is to build artifical computerized environments which can provide sensations and stimuli able to make it verisimilar every possible imaginary situation, so that the user which experiences VR is able to be in such virtual scenes and to act on and with such virtual object, with the same skill as they were real.

In particular, the 3D rendering of the scenses, which includes the modeling and the computation of the geometrical features, of the light behavior and of the real time visualisation makes the user able to perceive the three-dimensionality of such virtual scenes and objects, called in the following Virtual Environments (VE), in the large meaning of Virtual Worlds, immersive or not imersive.

To pursue these objectives, and in particular to allow a natural interaction between the user and the VE, we need computers able to execute in real-time all the calculations needed to produce the response to user actions without perceivable delays.

The rate at which these images are generated must be high enough to create the illusion of continuous movement. Images should be recomputed with a frequency of over 24 Hz to be realistic. As stated in [1] "*At one frame per second (fps), there is little sense of interactivity; the user is painfully aware of the arrival of each new image. At around 6 fps a sense of interactivity starts to grow. An application displaying at 15 fps can be certainly considered as real-time. There is a useful limit, however. From about 72 fps and up, differences in the display rate are effectively undetectable.*"

The necessity of producing such a high frame rate, however, limits the complexity of the virtual environment which, in a first approximation, can be meant as a measure of the realism of its representation.

The VE complexity can be expressed by means of the number of geometrical entities (often polygons) and texture maps which it is composed of. In spite of the increasing performances of graphics hardware, its power is often not enough for a fluent and interactive visualization of environments often consisting of millions of polygons. Therefore, too complex environments, i.e. built of too many polygons, are displayed intermittently; this damages the VR experience. The art of developing convincing VR applications is to keep the complexity of the environment as simple as possible, while conserving the impression of realism.

Actually nowadays the performances of the graphics hardware allow visualizing environments more and more realistic, making it possible to raise the degree of complexity of VEs as to be very close to the realization of the real-time photorealism. However in many cases it is still necessary to use software algorithms that reduce the number of the polygons processed in each frame, without perceptibly altering the main visual features of the VE. These optimization techniques can operate on VEs acting on the visibility or on the complexity of its components.

3.2 Data management for complex VEs

Data related to VEs, in order to be effectively processed, must be expressed in a well-organized structure so that all the calculations needed for their management can be performed very quickly. This kind of structures is known in literature as ***Scene Graph***.

The Scene Graph (SG) is often organized as a hierarchical structure in order to maximize its efficiency. This structure can be a *tree* or a *DAG* (Direct Acyclic Graph). In the first case every node of the SG can only directly descend from one parent, in the DAG case a node can descend from one or more parents. In both cases, anyway, the relationships between nodes don't allow the creation of cycles. Every node of the SG encapsulate information between geometrical data (polygonal meshes), points of view (cameras), materials etc.

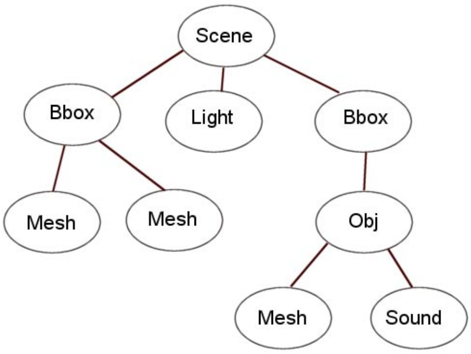


Figure Erreur! Argument de commutateur inconnu. **– Scene graph example**

The way a hierarchical structure can improve the SG efficiency is easily explained. If the scene has to be processed, for instance, to test its components visibility, a hierarchy based on the space occupation of the nodes (a parent node occupies a space greater or equal to the space occupied by its descendants) allows a quick exploration of the tree. In fact the non-visibility of a node implies the non-visibility of the whole sub-tree starting from that node. Therefore it is possible to avoid processing all the nodes owned from that sub-tree saving, this way, precious computational resources.

SGs are often built in a pre-processing phase in which all the subdivision operations are performed once for all. In other cases it can be preferable to build, or update, the tree during the execution of the application.

The hierarchical structures commonly used for the SGs are the BSP-trees (Binary Space Partition) and the BV-trees (Bounding Volumes). The BSP trees operate a top-down subdivision of the whole space, recursively partitioning it in two subspaces. The BV-trees follow a bottom-up approach: objects are enclosed in Bounding Volumes (simple elementary entities, such as spheres or boxes, completely containing them and easier to manage) that, in their turn, are enclosed in BVs of higher lever, till a BV for the whole environment is built.

Once the SG is ready, in the real-time phase it is traversed verifying the nodes properties: if, for instance, a visibility check has to be performed, the visibility of a higher level BV can be checked then, if needed, its sub-tree is traversed recursively.

3.3 Optimization techniques

The optimization techniques used to reduce the complexity of the 3d VE can be divided in:

- Culling techniques, if they operate on the visibility of its components
- Simplification techniques, if they operate on the complexity of its components

3.3.1 Occlusion culling techniques

The ***culling techniques*** identify a subset of components of the VE that are not visible from the current point of view and, for this reason, don't need to be processed from the rendering pipeline of the graphical hardware. These techniques can be subdivided in:

- ***View-frustum Culling*** techniques, which exploit the fact that the view volume is finite and it can be approximated by a frustum. The components which lie outside the frustum are not visible and they can avoid the pipeline process, saving time and resources. The hierarchy of the SG is exploited testing first the higher level nodes, passing to the lower level nodes only if the higher ones are classified as "visible" or "partially visible". If they are not visible, the related sub-tree does not need further tests. A lot of work has been made to refine this technique, operating on the SG side, on the type of BV to use and on the efficiency of the visibility test [2]
- ***Backface Culling*** techniques, which identify a set of polygons not facing the observer and, therefore are not visible. As the facing of polygons can be determined by observing their associated normal vectors, in these techniques the partitioning is made on the normal space

rather than on the location. This means that polygons are clustered on the basis of their associated normals [3]

- ***Occlusion Culling*** techniques, which identify a set of components of the VE that are hidden from other components. This is the most challenging field of research because, whilst in the first two cases the state of the art seems to be already satisfactory, there is still a big work in progress in this latest case. Several methods exist and only recently some of these have been implemented directly in the graphics hardware [4].

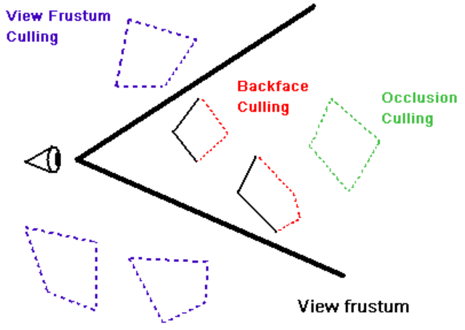


Figure Erreur! Argument de commutateur inconnu.**a – Culling Techniques**

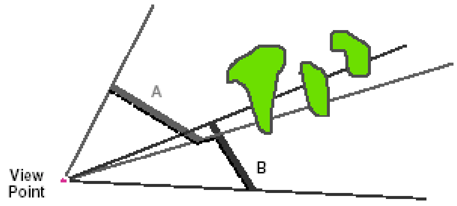


Figure 2b – Occluder fusion

The exact computation of the visible objects is very complex, that is why these algorithms try to find a set of objects that includes at least all the visible set plus some additional invisible objects. This set is known as the **Potentially Visible Set** (PVS), proposed by [16] and it is based on a **conservative visibility** concept. The goal is to be as close to the visible set, and be easy to compute.

A fundamental fact that must be taken into account in occlusion determination is that the cumulative occlusion of multiple objects can be far greater than the sum of what they are able to occlude separately. This is illustrated in 2-D in figure 3 with two occluders A and B. Neither A nor B can individually occlude any of the green shapes. However, A and B do occlude the green shapes if their occluding effects are combined to produce the fusion occluder, which is equivalent to the occlusion of the dotted line from the particular view.

Another important concept proposed by Airey [16] is scene **densely occluded**. Scenes densely occluded stand for objects that can hide a large number of elements in the scene

The occlusion culling methods can be classified by two philosophies:

- Point vs. region: For point-based methods the computations of visibility are made for the current viewpoint only. By opposite, region-based methods perform complex computations that are valid in a given region of the space.

- Image space vs. object space: In object space, the computations of visibility are made in the three-dimensional space while in image space, the methods work with the plane where the scene is projected onto.

The following section presents several occlusion culling algorithms.

3.3.1.1 Visibility from region using Cells and portals [16, 5]

This technique performs a visibility culling on some higher-level entity, such as rooms, doors etc. It subdivides the space in ***cells*** and ***portals***, the firsts being any kind of space volume (usually rooms), and the seconds being connections between adjacent cells. For each cell is determined, in a pre-processing step, a PVS, which is a set of the cells potentially visible from any point of view lying inside the cell. This can be made as the topology of the environment, i.e. the cells and how they are

connected each other by portals, is know. In real-time the ***active cell***, which is the cell where the observer currently is, is determined and only the geometry of the cells in the PVS of the active cell is managed, while the rest of the VE is not processed. The time and storage space required by the pre-processing tend to be excessive for large models.
This technique is not suitable for general purposes and is used in architectural/indoors environments.

3.3.1.2 Hierarchical Z-buffer

This technique was proposed by [17] and is an extension of the hidden surface removal using z-buffer. It replaces the z-buffer with a z-pyramid. The z-pyramid is a layered buffer with different resolutions. The lowest level is a copy of the z-buffer. Higher levels are created by halving the resolution in each dimension. The pixels of the coarser levels correspond to further z values of the 2x2 square of the lower level (figure 4). The hierarchy ends with a 1x1 buffer.

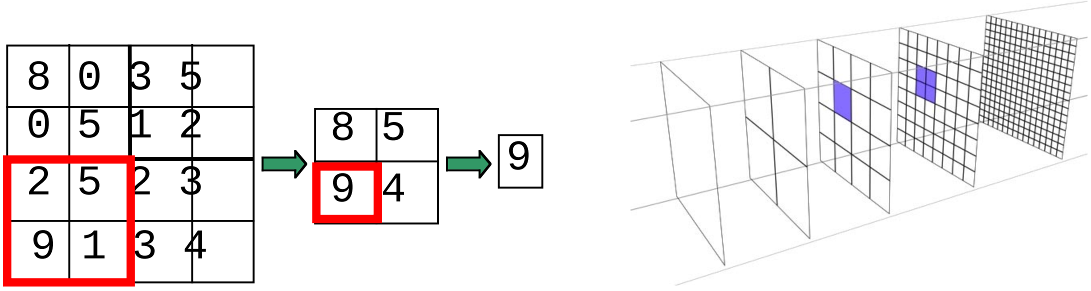


Figure Erreur! Argument de commutateur inconnu. **– Z-Pyramid values propagation and layered buffers**

While the scene is rendered, if the values of the z-buffer change, the z-pyramid must be updated.
The Hierarchical z-buffer method exploits object space coherence subdividing the scene with an octree. And it also takes advantage of the coherence among the layers of the z-pyramid.
The main idea is to test the polygons against highest level first. If the polygon is further than distance stored in the z-pyramid the polygon is occluded. Otherwise the test continues with lower levels until its visibility can be determined. An optimization can arise rendering the faces of the cube containing the node before the polygons of the node itself. If the containing cube is occluded, no more tests are needed and the algorithm ignores the entire node.The main handicap is its requirement of a special hardware.

3.3.1.3 Hierarchical occlusion map

This method proposed by [18] follows the same philosophy as the Hierarchical Z-Buffer but designed to work in current graphic hardware. The visibility is verified through two tests: An **overlap test** that analyses if an occluder object hides others elements. And a **depth test** to resolve the proximity of the occluder.
Occlusion maps are the projection of several elements into a plane (image space) without texturing, lighting and z-buffering. This algorithm manages a hierarchy of occlusion maps with multiple resolutions called Hierarchical Occlusion Map (HOM). In figure 5, a representation of this hierarchy is shown. The coarser levels are created by averaging blocks of 2x2 pixels and halving the resolution in each dimension. The hierarchy ends with 1x1 pixel map. Black pixels stand for the transparency level of the region while white pixels mean full opacity. Unlike the Hierarchical Z-Buffer the HOM stores only opacity information and not z-values.
In a pre-process step, the potential occluders are selected.
Then in run-time the algorithm performs for each frame the next steps.
From the current viewpoint and with the selected occluders, the algorithm proceeds to create the HOM.
The second step corresponds to the overlap test. It consists in projecting the bounding box of each object and comparing the projection calculated with the occlusion map of lower resolution. If the box overlaps pixels of the HOM which are not opaque, it means the box cannot be culled. Otherwise the depth condition must be fulfilled.
The final step determines whether an object is behind the occluders. The simplest method for testing the depth makes use of a plane placed behind the occluders. If the object is behind this barrier, it is occluded.

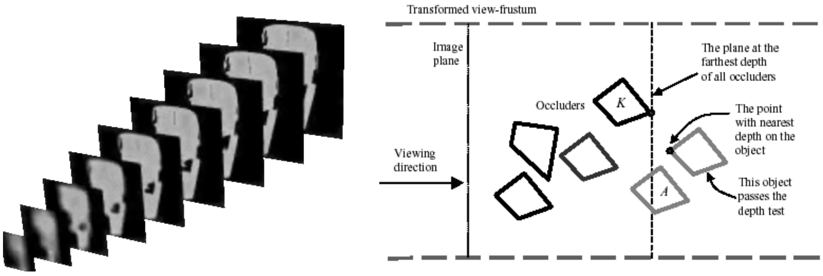
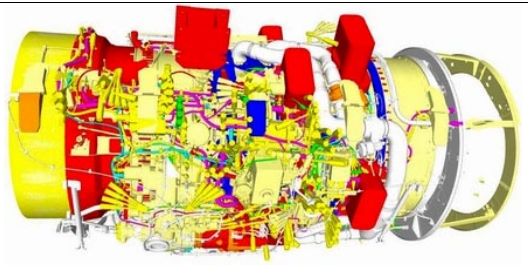
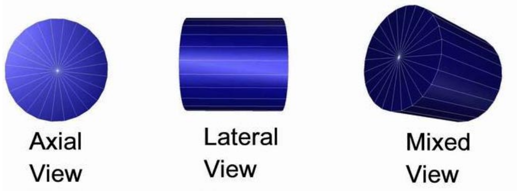

Transformed view-frustum
Image plane
Viewing direction
Occluders
K
A
The plane at the farthest depth of all occluders
The point with nearest depth on the object
This object passes the depth test

Figure Erreur! Argument de commutateur inconnu. **– Occlusion maps hierarchy and depth test using a plane**

The author proposes others depth test methods like the **depth estimation buffer**, which provides reasonable depth estimation.
This algorithm is designed to work with generic models.

3.3.1.4 *Spatial division occlusion culling*

This method [19] is applied in the visualization of aeronautical engines digital mock-ups and customizes the HOM algorithm [18]. Such environments are **non-densely occluded** where less of 50% of the scene can be occluded. Furthermore, the elements of the model have small dimensions compared to the complete model. Contrary to architectonical walkthrough, the navigation within the scene never penetrates inside the model. Another feature is that the model has axial symmetry with a cylindrical shape.

 Figure **Erreur! Argument de commutateur inconnu.**a – DMU of an aeronautical engine	 Figure **Erreur! Argument de commutateur inconnu.**b – Views of a cylindrical shape

Before the explanation of the method, it is necessary to mention that according to the viewpoint and the cylindrical shape of the model, three views can be defined:

- An axial view where the side of the model is seen.
- A lateral view where the front of the model is seen.
- A mixed view composed by an axial and a lateral view.

In a pre-process step, the algorithm partitions the model into cylindrical sectors. For each cell, the method identifies two occluders sets: axial and lateral. A cylindrical projection is used for lateral occluders and an orthographic projection for the axial ones (figure 8).

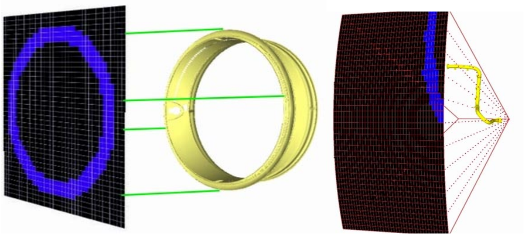


Figure Erreur! Argument de commutateur inconnu. **– Example of axial and lateral occluders**

The first step of the algorithm, determines the visible cells from the current viewpoint.
In a second step, the method divides the scene into two volumes: **Region of Probable Visibility** (RPV) and **Region of Probable Occlusion** (RPO). In figure 9 the two regions are divided by means of a pair of **cutting planes**. This separation must guarantee that the projection of the RPO elements is behind the projection of the RPV elements.

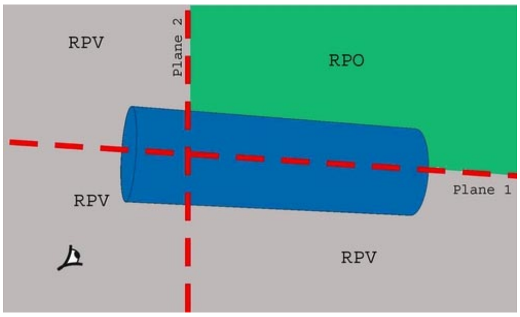


Figure Erreur! Argument de commutateur inconnu. **– RPO vs. RPV**

The third step corresponds to the construction of the HOM. The selected occluders are simplified and then rendered into a buffer where its resolution is lower than the image (512x512). Like in [18] the texturing and the lighting are turned off. In fact, the only region that must contribute to the occlusion map is the RPV.
In order to obtain this result the cutting planes are rendered into an auxiliary buffer with the color
buffer disabled. Their depth information will be stored in the z-buffer and objects inside the RPO will not be taken into account. At this point, the color buffer is enabled and the occluders are rendered into the auxiliary buffer. Only the polygons of the RVP will pass the z-buffer test. Once the occlusion map is obtained, the algorithm calculates the hierarchical structure of occlusion maps. The goal is to minimize the amount of operations in the overlap test.

The next step determines by the depth test, the region the objects belongs to. The idea is to compare the bounding volume of the object with the cutting planes. The objects inside the RPV can be sent to the graphic pipeline.

Finally only the objects inside the RPO must be checked with the overlap test. The projection of an especial bounding volume of the objects is compared with the occlusion map. The evaluation begins with the maps with less resolution. If the test is not conclusive then the test pursue with a finest level of the HOM. At this point, the method guarantees that the objects of the RPO who passes the overlap test are hidden.

3.3.2 Simplification techniques

The simplification techniques can be subdivided in:

1) Level of Detail (LOD) techniques, which exploit the principle that from a certain point of view not all the components of the VE need a detailed representation, as the small details may become undistinguishable because of the distance or because of the orientation. Therefore it may be useful to have more distinct representation of the same component, with an increasing level of complexity. In real-time, according to the distance from the observer, the orientation and the dimension of the component, the most appropriate (i.e. the one with the best compromise between quality and computational load) of these representation is chosen. Among these techniques the ***progressive meshes*** [6] have particular relevance: instead of creating manually several level of detail, they are created automatically from the original representation (the most detailed one), producing rougher and rougher representations of the starting one. The key improvement is that, to optimize bandwidth and memory occupation, every LOD can be stored as the difference from the previous one. This result is particularly important in the case of distributed VEs in which network considerations have to be carefully taken in account.

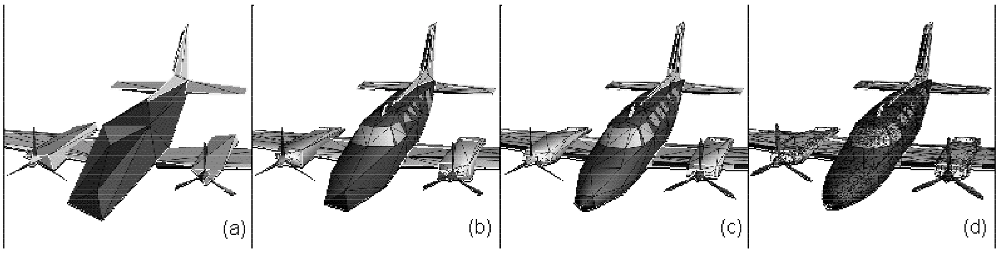


Figure Erreur! Argument de commutateur inconnu. **– Different levels of detail**

2) ***Image Based Rendering*** (IBR) techniques, in which the so-called *frame-to-frame coherence* is exploited. The graphic pipeline, for each frame, makes heavy calculation to process geometrical data in order to create the final image. Anyway, if the point of view changes continuously in time (and this is exactly the case of VR in which it is tied to a human being, rather than to one or more cameras - which can make it possible a discontinuity) each final image will differ very little from the previous one. In other words, wide portions of the final image will remain unchanged. An interesting option is not to render all the geometry from scratch at each frame, rather to re-use this unchanged portions avoiding processing the related geometrical entities (***image caching***).

Another option is to directly use images instead of geometry to simulate details. The ***texture mapping*** is the most common example: an image is added to a simpler shape like a decal pasted to a flat surface, to simulate a particular material or to add details hardly representable by means of geometry. The ***bump mapping*** [7] is based on the same principle but, rather than "stick" an image onto a polygon, it relies on light-reflection calculations to create small bumps on the surface of the object. The bump map "perturbs" the normals of the polygons it is mapped on, producing the required effect. Another technique, ***normal mapping***, operates in a similar way with the difference that the normal map replaces, rather than perturbing, the normals associated to the polygons.

Finally, on the extreme of mapping techniques, the ***impostors*** (or ***sprites***) technique [8] consists of completely substituting a component of the VE with an image, representing the whole object as seen from a certain point of view (POV), mapped on a polygon. Obviously these images are valid only in a limited range of POVs, while it needs perspective corrections the farther we go away from that particular POV. Different approaches are available: it is possible to store a fixed number of pre-computed images of the object from a significant set of POV, or to dynamically generate the images as soon as they are needed. In the first case a pre-processing phase is needed but no additional overload is requested in real-time. In the second case an overload is unavoidable (the time needed to generate images) but much less memory is needed, as there is no pre-computed representation to store.

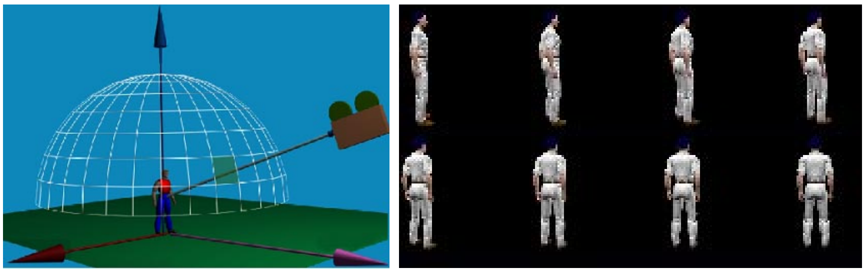


Figure Erreur! Argument de commutateur inconnu. **– Impostors creation and visualization**

3) ***Compression*** methods, which reduce processing time by explicitly removing redundant data to be processed. There have been approaches to reduce number of bytes to be transferred through a bus by encoding information using offset or by reducing significant range of values. The compressed data is decompressed in later processes such as the geometry processor or the pixel rasterizer. Along with the geometry compression, texture images [20] are also compressed to reduce the transfer cost. Some of these methods are implemented in current consumer hardware.

The compression method reduces the transfer bandwidth and it is effective for rendering systems with a bus bottleneck.

All these techniques operate on the performance side, in the sense that they allow to manage complex VEs reducing their computational weight without apparently produce a decrease of the visual quality of the images.

Currently, the noticeable advances of the graphics hardware performances make it at last possible to move the accent on the qualitative aspects of the virtual experience. Therefore the VEs become more and more detailed and faithfully reproducing reality. Two interesting fields can be explored; the first related to the modelling of the VE, the second related to the visual rendering.

3.4 Shape Modelling

Modeling can be defined as the process that brings to the description of an object (model) which can be used for different purposes (simulation, rendering etc.). Each modelled object can be characterised by several properties, regarding geometry, material, physical behaviour and so on.

The existing techniques for modelling the geometrical properties of objects are usually subdivided according to morphological features:

- Surface modelling
- Solid modelling
- Point Modelling

3.4.1 Surface Modeling

The most common representation of a 3d object is a polygonal mesh describing its surface. A ***mesh*** is a surface composed of polygons (usually triangles) that approximate a portion of the surface. The polygons are connected among them so that they never intersect, but they at least share a vertex or an edge. In formal terms, the geometry of a mesh can be expressed as a tuple (K,V) in which:

- V is the set of the vertices composing the shape of the model (geometrical information)
- K specifies how the elements of V are connected (topological information)

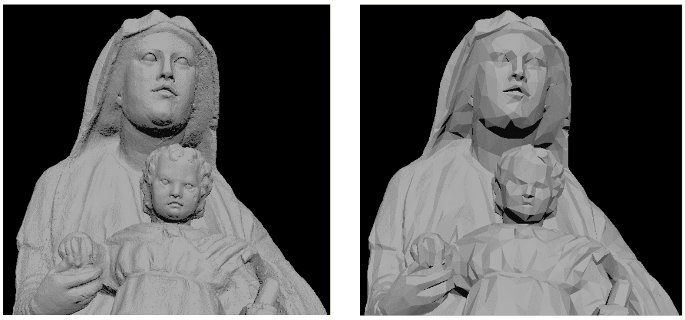


Figure Erreur! Argument de commutateur inconnu. **– Polygonal meshes with different resolutions**

The reason why triangular meshes are widely used is because triangles are the simplest polygons and they can be easily represented and manipulated. In addition to this, every modern graphical architecture has hardware-accelerated optimized capabilities of rendering triangles.
A drawback in mesh modeling is that a great number of triangles is needed to provide a detailed representation, even if this problem can be solved using the techniques described in the section **Erreur! Argument de commutateur inconnu.**.

Surfaces can be also mathematically specified, rather than geometrically approximated. Each surface can be seen as a 2-dimensional subset of R^3. There are several methods of expressing a surface:

- ***Implicit representation***, in which a surface is defined as the set of the points satisfying the condition F(x,y,z) = 0. The drawback of the implicit surfaces is that not every function produces a surface, and not every surface can be defined by a function.
- ***Parametric representation***, in which a surface is defined as a collection of ***patches***. A patch ψ is a subsurface characterized by a grid of control points hat determine its position and its shape.

In this case the difficulties are related to ψ, not always easy to determine.
The most commonly used patches are the ***Bezier Cubic Patches*** and the ***B-Splines***.

Basically the Bezier Patches use cubic polynomial blending functions (n=3), and a grid of 4x4 control points, with the condition that the value of the function and the value of its partial derivatives in the control points is the same of the grid control points one. This allows to have only a global control on the shape of the curve.
The B-Splines use different blending functions that allow to have also a local control on the shape. In other words, the B-splines allow curves to be a better approximation of the control points. The used blending functions make use of additional points, name knot-points, which give a smoother control on the surface behaviour: they can be equally spaced or not. In the last case they are usually referred as ***NURBS*** (Not-Uniform Rational Based Splines). NURBS are very precise for defining conic sections and other analytic functions, but they are at the same time flexible to define also other shapes. The drawback is that their rendering is computationally demanding, therefore they are commonly used for modelling the shapes and then they are usually approximated by polygonal meshes in order to be rendered.

3.4.2 Solid Modeling

Solid modelling provides a description of objects that represents solid interiors of them: in this case the surface may not be described explicitly. One of the simplest case of solid modelling is the ***CSG*** (***Constructive Solid Geometry***), in which an object is seen as the result of composing simple 3d primitives (cubes, spheres, cylinders etc.) with boolean operators. The primitives can be modified by the usual transformations (translation, rotation, scaling, deformation).
This is the usual approach of CAD modelling programs that, often, allow the modeller to combine also other typologies of modelling (i.e. mesh modelling etc.).

Voxel representation assumes an increasingly importance in the field of solid modelling. In this technique, the space is uniformly portioned in a regular grid, whose elements are called ***voxels***. In other words a voxel is the 3d correspondent of a 2d-pixel. In each voxel are stored the main properties of a given object (colour, density, temperature, etc.).

The voxel representation is the most natural and practical one in some applications, as in the case of some scanning devices and the FEM simulations, and it offers the advantages to be a very simple, uniform and intuitive representation. Voxels are easily manageable and they can be manipulated with trivial operations. Of course it is a discrete representation of the space, therefore it is only an approximation. Moreover it requires a huge amount of memory for storing, even if advanced compression techniques are available to effectively reduce this requirement.

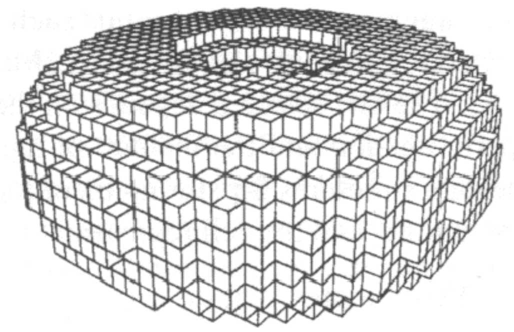


Figure Erreur! Argument de commutateur inconnu. **– Voxel representation**

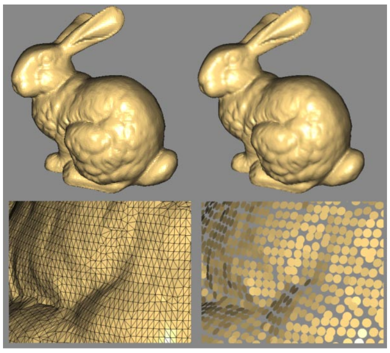


Figure Erreur! Argument de commutateur inconnu. **– Polygonal mesh Vs. Surfels**

3.4.3 Point-based Modeling

A relatively new technique is the ***Point-based Modeling***, which can be considered as the evolution of ***particle systems***. Particle systems consist of points that are moved according to different rules, typically for rendering fuzzy phenomena such as fire, smoke, snow, etc. Though this technique is not recent, only today it can be effectively used thanks to the advancements in graphics hardware and CPU development, as the handling of large point sets is a computationally demanding task.

In Point-based Modeling [21], the basic building blocks are points, also known as surface elements (***surfels***), whose basic properties are their 3-d coordinates, colour, material and a normal vector that is used for the computation of lighting.

This technique is very handy when used on models obtained by 3d-scanning procedures, as they natively produce a cloud of points. This allows bypassing the common operation of producing a polygonal mesh from the obtained points, and therefore eliminates the problems related to the connectivity of points. In comparison to voxels, points can carry more details, have less memory requirements and don't suffer from the discretization problems. On the other hand currently there are not high-speed implementations of point rendering and, besides, the transformations and calculations required for point clouds are computationally demanding. Moreover, polygonal meshes are still more effective when dealing with flat surfaces, as they can result in simple geometry compared with the one obtained with point-based technique.

3.4.4 Procedural Modeling

In order to reach a high complexity of the VE it is less and less practical to model all the details by hand. An alternative to manual modelling is represented by algorithmic techniques [9] to generate texture maps, surfaces and 3d objects.

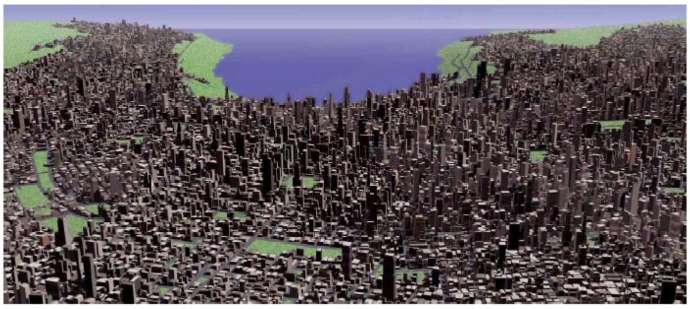


Figure Erreur! Argument de commutateur inconnu. **– A city procedurally modeled**

These ***procedural*** techniques allow creating very complex structures, animated if needed, at low costs [10]. The information needed are a set of parameters, thus relatively simple inputs, and a procedure that uses these parameters to generate the desired structures.Usually a modelling language is implemented, which provides a set of algorithms and procedures to code and abstract the details of a model, allowing a high-level control and freeing the modeller from the necessity to insert detailed specifications.

Moreover, the procedural approach allows also a strong amplification of the results, by opportunely controlling the parameters: with few input data it is possible to generate surprisingly refined models.

The typical example of procedural modelling is the generation of fractal terrains. In general, the procedural modelling is particularly advisable when generating several objects of the same type that have to be similar but not identical. A forest is a good example: if it was consisting of the same tree repeated several times, the sensation of realism will be weak, while a set of trees of the same kind that differ even for few details, will appear much more verisimilar. As it is not conceivable to manually model these differences, a procedure able to automatically produce them seems to be much more effective.

3.5 Lighting and shading

Once the model of the VE has been created, and after all the management techniques, it has to be graphically rendered. Ideally the VE should be undistinguishable from a real environment and one of the key features is the ***photorealistic rendering***, which means that the produced images can be mistaken for real photographs. During the last years photorealistic rendering was mostly an expensive non-realtime process: it took hours, or even days, to produce high-quality photorealistic images. Nowadays real-time photorealistic rendering is becoming more and more of a reality, also thanks to programmable graphics boards. The introduction of CG, a language for the low-level programming [11] of graphics hardware, has in fact reduced development efforts for photorealistic rendering effects.
Most of the techniques used for this purpose are related to the simulation of the way real light behaves. This is a very demanding process and, in order to be performed in real-time, a compromise must be found between realism and computational costs. One of the most accurate methods to model the light behaviour is the ***BRDF*** (bi-directional reflection distribution function [12]), a function that describes the directional dependence of the reflected energy. Although very accurate, the BRDF-related calculations are still too heavy to be used in real-time for every kind of model.
The most commonly used technique to deal with real-time lighting is the ***Phong*** lighting, that can be considered as a particular case of BRDF and gives an inexpensive formulation to calculate the specular component of light (the most computationally demanding component, responsible of the reflective behaviour of objects). Phong lighting, although it is not physically coherent and it is not accurate for several materials, is still in any case the most reasonable compromise for real-time lighting and shading of objects.
Once the lighting properties are defined, the model has to be shaded accordingly. The most inexpensive ***shading*** algorithm is the ***flat shading***: in this case the whole triangle has the same colour, calculated through the normal vector associated to the polygon. Therefore no interpolation has to be made; on the other side, the results of this kind of shading are often non-realistic.

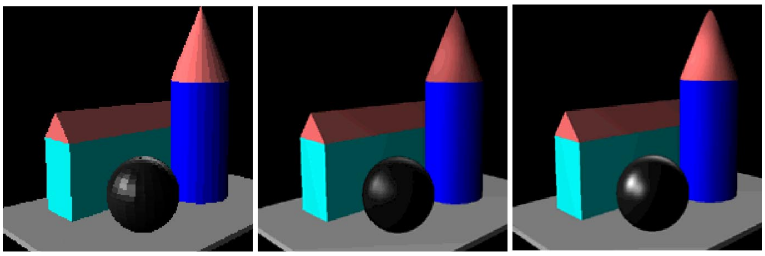


Figura Erreur! Argument de commutateur inconnu.**6 - Flat, Gouraud and Phong shading**

A compromise between quality and speed, commonly used till nowadays in real-time computer graphics, is the ***Gourad shading***: the colour is calculated for the vertices of the polygon, using the related normal vectors. For the other pixels an interpolation is made among the computed colours.
The most accurate technique, computationally demanding and performable in real-time only in the last years thanks to the current generation of graphical hardware, is the ***Phong shading***: it is similar to the

Gourad shading, but the interpolation is made on the normal vectors, rather than on the computed colours. The Phong shading represents the basis of the per-pixel shading (see section 2.6).

One of the ways to deal with more realistic light effects is to sample the objects after an accurate non-real time lighting process, and obtain ***lighting maps*** that can be mapped like textures on the objects, or ***radiosity meshes***, which are meshes more complex than the starting ones, this added complexity incorporating also lighting information stored as per-vertex colours. Disadvantages are that specular reflection can not be modelled.
The best-known non real-time lighting technique is probably ***Ray-Tracing***, which is a global illumination based rendering method. It traces rays of light from the eye back through the image plane into the scene. Then the rays are tested against all objects in the scene to determine if they intersect any objects. If the ray misses all objects, then that pixel is shaded the background color. Ray tracing handles shadows, multiple specular reflections, and texture mapping in a very easy straight-forward manner. The major drawback is that even in this case the long calculations make it impossible to use this technique for real-time rendering, even if a lot of research is presently carried out and good performance are achieved: at now frame rates of about 5 fps are obtainable from hybrid ray-tracing algorithms that, therefore they can be considered near real-time rendering performances.

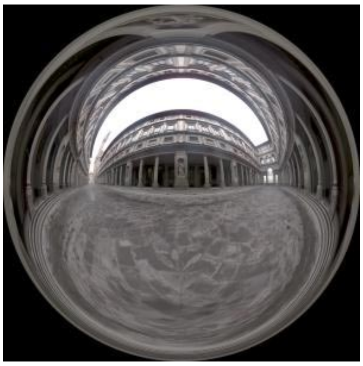


Figure 17 – A light probe for the Uffizi Gallery in Florence

The most complex interaction between lights and VEs happens at the surface, and it is a combination of the surface properties (specified from the BRDF) and of the nature of the light. To recreate this interaction mathematically takes a long time and artistic skills.

However, instead of using a formula to approximate the ***illuminance*** (i.e. the lighting energy that the observer indirectly receives reflected from the objects surface), it can be measured, stored in a ***radiance map***, and read in real-time. It can be considered that a surface gathers light from a sphere surrounding it.

In order to build a radiance map, a possible methodology consists in taking a series of pictures of a
light probe (i.e. a mirrored sphere [13] that reflects the entire surrounding environment) under certain well-specified exposure conditions. ***High Dynamic Range*** (HDR) radiance maps can be built with pictures taken at a range of exposures [14].
Once these maps are available, they are mapped on a fictitious object (a sphere or a cube) that surrounds the whole VE. Using hardware-accelerated texturing it is possible to render in real-time the components of the VE as they actually reflected the environmental light.

3.6 Real time visualisation: new trends

Until 1999 the graphical hardware had a well-defined boundary line between the professional devices and the consumer devices. Thanks to the strong success of computer games, consumer devices experienced a quick increase of performances on low-cost basis, if compared with the extremely expensive multi-pipeline architectures. After the appearance of the first 3d hardware-accelerated graphic boards, it was in 1999 that NVIDIA came out with the first consumer-level GPU (Graphics Processing Unit). This GPU was the first to feature a hardware “transform and lighting” engine, and it allowed an impressive increase of performances.
As a result of this technical advancements in graphics cards technology, some areas in 3D graphics programming had become quite complex. To solve the problem, new features were added to graphics cards to simplify the process, including the ability to modify the rendering pipeline of the graphics card using ***vertex and pixel shaders***.

A vertex shader is a graphics processing function used to add special effects to objects in a 3D environment by performing mathematical operations on the objects' vertex data. Each vertex can be defined by many different variables. For instance, a vertex is always defined by its location in a 3D environment using the x-, y-, and z- coordinates. Vertices may also be defined by colors, textures, and lighting characteristics. Vertex Shaders don't actually change the type of data; they simply change the values of the data, so that a vertex emerges with a different color, different textures, or a different position in space. Before the introduction of the GeForce3, vertex shading effects were so computationally complex that they could only be processed offline using server farms. Now, developers can use Vertex Shaders to breathe life and personality into characters and environments, such as fog that dips into a valley and curls over a hill; or true-to-life facial animation such as dimples or wrinkles that appear when a character smiles.

Pixel shaders are based on the same principles, but they work on 2d elements (pixels) rather than on 3d vertices. The availability of this instrument allow to perform in real-time advanced shading effects, such as the ***per-pixel shading***, a shading technique that operates in the pixel space and results in a more accurate lighting model and therefore in realistic effects, even if it is more computational demanding.

In the beginning, vertex and pixel shaders were programmed at a very low level, using the assembly language of the graphics processing unit. Although using the assembly language gave the programmer complete control over code and flexibility, it was pretty hard to use and non-portable. A higher level language for programming the GPU was needed, and a high-level programming language (***Cg***) was created to overcome these problems and make shader development easier.

There are still new features that can be implemented to add realism, both on the hardware side and on the software side. ***Depth of field*** is an example: real cameras (and eyes too) have depth of field: they are focused at one plane that has particular distance and everything else is more or less fuzzy. 3D rendering has somewhat different internal workings (simulating an infinitely small lens) and as a result the entire scene is in focus. With a pixel shader it is possible to simulate depth of field and achieve more photorealistic scenes by selectively blurring parts of the image

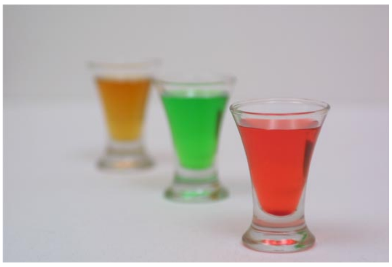


Figure 17 – Depth of field effect

There are still new features that can be implemented to add realism, both on the hardware side and on the software side. ***Depth of field*** is an example: real cameras (and eyes too) have depth of field: they are focused at one plane that has particular distance and everything else is more or less fuzzy. 3D rendering has somewhat different internal workings (simulating an infinitely small lens) and as a result the entire scene is in focus. With a pixel shader it is possible to simulate depth of field and achieve more photorealistic scenes by selectively blurring parts of the image.

On the hardware side, new visualization devices are claiming an increasing attention. For instance ***HDR Displays*** allow displaying images with a dynamic range much more similar to that encountered in the real world. In the field of portable displays, ***retinal*** (fig. 18) ones may represent the near future of HMD see-through displays. They use scanned-beam technology, and then optically guide the computer generated image directly to the user’s eye. The specially-coated ocular piece is optimized to allow the image to be reflected into the user’s eye, while simultaneously allowing the user to continue to see the outside world unhindered. Users can adjust the optical focus of the image, placing it precisely at the user’s working distance. The result is a good clarity of combined image data and the real-world.

Additionally, new revolutions are announced on the customer level graphical architectures: after the advent of ***AGP*** in 1997 (an architecture that allowed a significant speed-up of the communication between the GPU and the central memory, resulting in high-performance graphics boards), the new architecture today available, ***PCI Express***, delivers a new level of PC performance for graphics and networking, by doubling the bandwidth of the most powerful AGP boards and offering scalability for the future.

3.7 Stereoscopic visualization

Finally, the rendered images must reach the user's eyes. This can happen in several ways: by using a simple desktop monitor, projection screens, of stereoscopic devices, like the Powerwall, the CAVE [15] or a Head Mounted Display (HMD), a wearable helmet provided with two small displays put just in front of the user's eyes.

The principle of stereoscopy is that humans own two eyes, each one of them giving a different perspective vision of the surrounding world. Combining these two perspective bi-dimensional images, it is possible to recover information on the missing third dimension (depth). The perception of the depth is made possible by a series of visual "depth cues". If an accurate simulation of the reality has to be provided in a VR experience, these cues must be replicated in order to "cheat" the observer and let him perceive a 3d world. The most important cues are: shading, parallax, perspective, eyes separation are focus. Some of these cues can be effectively replicated on a computer system, while some others are very difficult to implement. There are also non-visual cues that provide additional information, like the ones coming from the vestibular apparatus or kinaesthetic data from the neck etc.

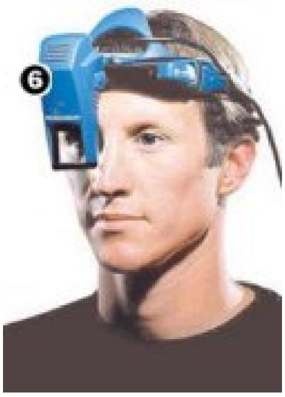


Figure 18 – Retinal Display

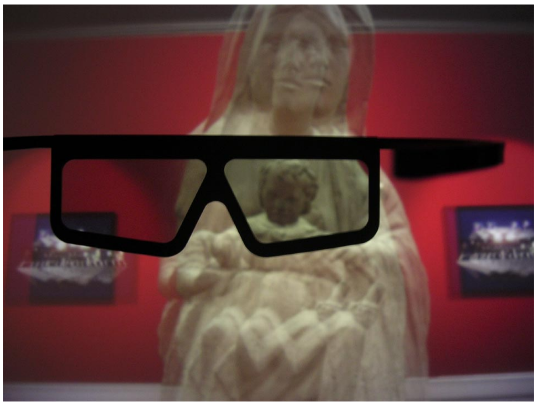


Figure 19 – Stereo image with polarization filters

A system for the stereoscopic visualization must be composed of:
- Software able to generate two monoscopic bi-dimensional images, one for each eye, created and synchronized in order to give back the opportune depth cues
- Hardware able to let each eye perceive only its correspondent image

Basically two different hardware technologies are available for the perception of stereo images: active stereo and passive stereo. In ***active stereo***, for each frame two images are projected *sequentially*; therefore there's a continuous hi-frequency (about 120Hz) switching between the images for the right eye and the images for the left eye. Users wear a special active device, ***shutter glasses***, which are synchronized with the image switcher and able to make lenses opaque or transparent. When the image for the right eye is present, the left lens is completely opaque, otherwise it is transparent. The same happens for the right lens. The human brain, actually, receives a sequence of images but they are so quickly presented that it believes to perceive them at the same time. In other words the brain merges the images and can reconstruct depth from them

In ***passive stereo*** both images are projected at the same time but, thanks to a system of polarization filters that exploits the phenomenon of light polarization, only the correct image (fig. 19) reaches each eye. There are advantages and disadvantages for both technologies: active stereo is more expensive and requires dedicated hardware, passive stereo presents the problem of ***ghosting*** (or stereo crosstalk), which means that one eye perceives also a small fraction of the image presented for the other eye.

The ***HMD*** does not suffer from this problem because each eye has a LCD panel directly in front of it, however it is expensive and it has a limited field of view.
In order to produce the right feedback to the sensorial system and allow a natural interaction with the VE, the stereoscopy must be combined with a ***tracking*** system that at every instant read the position of the user's head. It is necessary to know the position of the observer in the VE in order to calculate the correct perspective and the direction of the eye separation. A tracking system, able to perform a complete motion capture, is also desirable to provide to the VR system information about the location of every component of the user's body, particularly the ones most involved in interaction, i.e. limbs.

3.8 References

[1] Real-Time Rendering, Tomas Akenine-Möller and Eric Haines, A.K. Peters Ltd., 2nd edition, ISBN 1568811829
[2] Hierarchical geometric models for visible surface algorithms, James H. Clark, Communications of the ACM, 19(10): 547-554, October 1976
[3] Fast backface culling using Normal Masks, H.Zhang e K.Hoff, , ACM Interactive 3D Graphics Conference, 1997
[4] Occlusion (HP and NV Extensions), Ashu Rege, NVIDIA website, www.nvidia.com/developer
[5] Visibility preprocessing for interactive walkthroughs, Seth J.Teller, Carlo H.Sèquin, Proceedings of the 18th annual conference on Computer graphics and interactive techniques 1991
[6] Progressive meshes, H. Hoppe, proceedings of SIGGRAPH '96, pp.109-118 (1996)
Highly detailed geometric models are rapidly becoming commonplace in computer graphics.
[7] Simulation of Wrinkled Surfaces, James F. Blinn, Computer Graphics, Vol. 12 (3), pp. 286- 292
[8] Dynamically generated Impostors, Gernot Schaufler,MVD '95 Workshop "Modeling – Virtual Worlds – Distributed Graphics" (Nov.95) D.W.Fellner (ed.) Infix pp. 129-136
[9] Texturing & Modeling: A Procedural Approach, 3rd Edition, David S. Ebert, F. Kenton Musgrave, Darwyn Peachey, Ken Perlin, and Steven Worley
[10] Artificial Evolution for Computer Graphics, Karl Sims, July 1991, Computer Graphics, Vol. 25,
[11] GPU Gems: Programming Techniques, Tips, and Tricks for Real-Time Graphics, Randima Fernando, Addison-Wesley Professional;ISBN: 0321228324
[12] Texture and Reflection in Computer Generated Images, J.Blinn and M. Newell, Communications of the ACM, volume 19, 1976, pp. 542-546
[13] Illumination and Reflection Maps: Simulated Objects in Simulated and Real Environments, G. Miller and R. Hoffman, SIGGRAPH '84 Course Notes Advanced Computer Graphics Animation, 1984
[14] Recovering high dynamic range radiance maps from photographs, P. E. Debevec and J. Malik, Proceedings of SIGGRAPH'97, pp. 369 – 378
[15] The CAVE--Audio Visual Experience Automatic Virtual Environment, Cruz-Neira, C., Sandin, D., DeFanti, T., Kenyon, R., & Hart, J. (1992). Communications of the ACM, 35, 6, 65-72.
[16] Towards Image Realism with Interactive Update Rates in Complex Virtual Building Environments, Airey, John M., John H.Rohlf, and Frederik P. Brooks Jr., Computer Graphics (1990 Symposium on Interactive 3D Graphics), vol. 24, nº 2, pp. 41-50, March 1990.
[17] Hierarchical Z-Buffer Visibility, Greene, Ned, Michael Kass, and Gavin Miller, Computer Graphics (SIGGRAPH'93 Proceedings), pp. 231-238, August 1993.
[18] Visibility Culling using Hierarchical Occlusion Maps, Zhang, Hansong, D. Manocha, T. Hudson, and K.E Hoff III, Computer Graphics (SIGGRAPH'97 Proceedings), pp. 77-88, August 1997.
[19] Virtual reality for aircraft engines maintainability, A.Amundarain, D.Borro, A.Garcia-Alonso, JJ.Gil, L.Matey and J.Savall, Mécanique & Industries 5, pp.121-127 (2004)
[20] Multiresolution Colour Texture Synthesis, M. Haindl and V. Havlicek in proceedings 7th International Workshop RAAD'98, ASCO Art & Science, pp. 297 - 302, 1998
[21] Point-Based Modeling, Markku Reunanen, Helsinki University of Technology

4 Motion modeling and computation

©Annie Luciani
INPG, September 2004

In this part, we will examine the main types of models to represent motions in the aim of positioning them in the action-vision chain and of analyzing their strengths and weaknesses from an enactive point of view.

During the first phases of the Computer Graphics development, until the 80's, the main feature brought by computer animation, compared to the conventional cinematographic approach, was the automation of interpolation processes between key-frames [Lasseter 1987].
In the middle of 80's, with the apparition of deformation and transformation models, the computer models for motion synthesis followed various directions and the research in computer modeling and synthesis experienced a new boom.
Up to now, computer models for motion synthesis has been usually classified in three main categories :

- Phenomenological models of motion by means of cinematic evolution functions
- Generative models of motion by means of physically-based models
- Generative models of motion by means of behavioral models of artificial life and artificial intelligence.

In the following, we present the properties of each of this type of models, positioning them with the shape modeling.

4.1 Phenomenological models of motion

4.1.1 Principle

Key-frames and interpolation methods (including morphing techniques) are related to a phenomenological representation of the motion. These methods have been improved by cinematic representation in which the evolution of morphological (positions, shapes) or visual (color, intensity) parameters is represented by temporal evolution functions where the time is an explicit variable: Motion = f(t).
Such representation supposes that the morphological and visual parameters are predefined and their evolution is applied on them. It can be considered as a mapping of motion on predefined object. Then, the methodology consists in :

- Modeling the 3D non evolving object by means of geometrical modeling (whatever the representation is - as described in the previous paragraph "3D modeling). In these representations, the time is absent.
- Applying evolution on the geometry of visual parameters of such models.

4.1.2 Types of phenomenological models

There are three techniques in the phenomenological approach of motion:

- Key-framing and morphing
- Explicit interactive evolution functions
- Kinematics models

4.1.2.1 Key-framing and morphing

Assuming that particular states are relevant in a continuous motion, key-framing and morphing technique is based on the idea that only particular states are relevant in a continuous motion. Thus, the modeling process consists in establishing the definition of such "key-frames" and in reconstructing the continuous motion by interpolation methods. In the 80's, methods to automatically generate families of key-frames from existing motions are developed, mainly for the human motions from biomechanical information. A lot of interpolation methods have been developed. Mainly, they are based on mathematical functions (linear, spline and Beziers functions, etc.) as spline functions and they take into account various criteria such as:

- the presence of singularities of the complex motion for rendering smoothness, transients, breaking points, etc.,
- the localization of the interpolation process between the control points,
- the control of the derivatives in order to generate motion families (proximity, tension, etc),
- the control of the velocity variation including effects like anticipation, stretching, squeezing,
- the smoothing in specific points (starting and end points), etc. [Burtnik 1976] [Steketee 1985] [Reeves 1981] [Friedrich 1998].

4.1.2.2 Explicit interactive evolution functions:

The motion applied to a 3D shape can be defined explicitly by evolution functions, mathematically defined [Coquillart 1990] [Coquillart 1991] [Lamousin 1994] [Frish 2002] [Milliron 2002] or provided by sensors (mouse, sticks, motion capture devices). Even they can be controlled, the mathematical functions offer a limited set of expressive motions, with too much regularities. Signals provided by sensor inputs that acquire human gestures corresponding to real motions do not have this type of limitations.

Motion captures are techniques widely used to profile free-motions. Nevertheless, the signals provided by such motion sensors input have to be processed by signals filtering in order to obtain a sufficient signal to noise ratio, respectively by applying extraction features processes in order to obtain data usable for the motion design. The signals provided by sensors are generally noisy and characterized by a lack of clear information. They have to be improved by signals filtering and reconstruction processes.

In terms of low-level control, this type of representation corresponds directly to devices that are pure sensors (non-retroactive sensors).

In term of high-level control, the signals representing the motion have usually the form provided by the sensors, for example position sensors. Usually, a high-level animation control needs to transform these signals in others, like: velocities (t), direction changes (t), etc…

4.1.2.3 Kinematics models

Kinematics representations of motion are generated by mathematical descriptions of the object kinematics. These types of techniques are well adapted to the representation of the displacement of solid, rigid or articulated rigid objects.

The main limitations of such approaches are situated in the representation of complex motions like:

- Complex dynamic non-linear features (fractures, states changing, etc.),
- Motions that are highly correlated and interdependent, for example deformations allowing the representations of the non-penetration during collisions, the temporal relation between displacements and shape deformation, the coupling between two deformations,
- Evolving scenes in which there is a huge quantity of different motions, as in collective interactive phenomena

In term of control, there are two ways to model an expected motion: direct and inverse kinematics. In direct kinematics, temporal evolutions are calculated by mathematical functions in which time is explicit: $\Delta X(t) = J \Delta\_(t)$, Jacobian matrix $Jij=dXi/dQj$. If the action is designed in term of task to be performed (for example to reach a target) and not in term of movement performance, the modeling process consists in defining targets and in calculating motions by means of inverse kinematic methods. These types of methods are related to the modeling of constrained motions.

4.1.3 Main properties from an enactive point of view

The two main properties of such phenomenological models of motion, related to the action-vision fusion and seen from an enactive point of view are:

1. Such phenomenological representations are well adapted to the free-hand control. We call free-hand control, the control of motion through human gestures that are not constrained by the object manipulation. This kind of control corresponds to the “pure semiotic action” in the Cadoz’s typology of gestures [Cadoz 1994, 2000] and they may be conveyed by non-

retroactive transducers as pure sensors. These sensors work through extensive variables [Luciani 2004], as displacements, positions, velocities, that follow the principle "observable variables". At a first glance, they seem to be well adapted to the direct hand manipulation. The "hand-eye" chain, underlaid by such representations is a direct chain: from extensive variables given by gestures to geometrical computations and representations, which process also extensive variables (positions, shapes, etc.). All the processed data are extensive variables, i.e. of the same nature.

2. The motion is applied on a shape. This means that the representation of the world is a geometrically - based representation. The geometrical features are "given", they only have to be modeled. This means that the approach from an enactive point of view differs (1) from the ecological cognitive approaches for which the space is built from psycho-cognitive experiences and (2) from the emergent dynamic approaches for which the shapes derive from the motion. In such phenomenological approaches, the spatial properties (geometrical shapes) of the space are predetermined and precede motion. In ecological cognitive approaches, shapes are built. From dynamics approaches, shapes are not predetermined but they emerge from movements. Only the infinitely rigid objects are compatible with such phenomenological approaches.

4.2 Generative models of motion

The strength of phenomenological models consists in the fact that they correspond to an explicit description of possible observed motions and performances, to be phenomenological. They are underled by an analysis-synthesis method. Theoretically, any motion can be represented by such methods. Nevertheless, their complexity increase dramatically in representation of high level qualities as "softness", "hardness", "rhythm changes", "dynamic complex correlations of complex shapes motions", "emergent non predictable evolutions.

This led to the development of generative models of motion. Here generative means processes in which the time is implicit and the calculation process produces families of evolution functions according to the parameters of the generative processes. There are two types of generative models:

- The physically-based models, modeling the dynamics.
- The agent based models, developed in artificial intelligence and artificial life.

4.2.1 Physically-based models

4.2.1.1 Principle

Physically-based modeling implements dynamics. They are basically motion generative due to the fact that Dynamics computes the kinematics of motions. In ancient Greek, "Dynamo" means forces, and "Kinema" means movement. Dynamics is based on the use of intermediate intensive variable – namely the force – to compute co-evolution of extensive variables. The intensive variable, with its native action-reaction principle, represents the mutual influences between two observed extensive variables. In computer simulation, the processes are:

1. a couple of extensive variables produce intensive variable through interaction laws,
2. intensive variables produce extensive variables through behavioral rules,
3. extensive variables are used to display motion.

In computer animation, physically-based models have been used, in a first time, as computational method to solve minima problems as the penalty method in collisions algorithms, and more generally, as a constrain solver used locally in kinematics (direct and inverse) approaches. Afterwards, they have been used to synthesize natural evolving phenomena. A lot of works were devoted to the simulation of realistic evolving natural phenomena by implementing dedicated physics models (calculation of Navier-Stokes equation for turbulences, heat equations, tridimensionnal elasticity, etc.). They have been rarely used as a generic method to model dynamics behaviors. More recently, physically-based particle methods [Greespan 1973, 1997] [Luciani 1991] have been developed, allowing the modeling of a huge variety of dynamic effects and motions, with minimal models easy to compute.

4.2.1.2 Types of models

There are three main types of physically-based models:

- Continuous models
- Mesh-based discrete models
- Particle-based discrete models

Continuous models

The phenomenon (for example, deformation of an object) is represented in a continuous formulation. Each model corresponds to a specific phenomenon: rigid and flexible objects [Terzopoulos 1988a] [Terzopoulos 1988b] [Baraf 1992] [Terzopoulos 1993], Metaxas 1996], tridimensionnal elasticity, Navier-Stokes equation for turbulent fluids, matter transport for granular material, friction models [Baraf 1991] etc. Thus, such differential partial equations are solved according to various methods for : finite difference method, implicit and explicit resolution etc. These models are mainly "one-shot" models, able to produce highly realistic motion, but less reusable, and less easy to design and to manipulate.

Mesh-based discrete models

The most known method is basically the finite elements method (FEM), used to calculate the dynamic behaviors of objects. FEM was widely used in mechanics to compute deformations of compact mechanical bodies. It is also widely used to solve problems as variational problems in physics. This method is a geometrically-based physical one in the sense that the geometrical features (shape, volume) of the body are given and discretised in space by geometrical basic elements, constituting a mesh. The forces applied within the elements are contact and cohesive local forces representing the contiguity of the matter. A lot of works that use this type of method exist in the field of mechanics, that leads computer graphics researchers to use it. Since the shape is predefined and maintains its cohesion, this method provides complex deformation of complex body.

Another mesh-based methods are those used to simulate behaviors of continuous medium (fluids, gas, etc.), called lattice methods (Lattice Gas Method, etc.). They were rarely used in Computer Graphics. The reason is probably that computer graphics focuses mainly on object exhibiting a clear shape, even if this shape is deformable.

Particle-based discrete models

Particle-based models appear more recently in Computer Graphics. They were used since the beginning of computer calculations in physics, as in the Los Alamos laboratory to compute complex behaviors of turbulent fluids. There were stopped due to the low computation power of computers at that time. The exponential increasing of the computational power of computers renders newly attractive this approach. The usual understanding of this method, often named mass-spring, (physical objects being modeled by a set of punctual masses linked by elasticity) reveals that, in computer graphics, physical modeling is understood only as an implementation of the rules of physics, rather than as a generic method of modeling, at the same level of abstraction that neural or cellular automata networks. Used in such way, its modeling power as well as its computational efficiency is obviously limited. As described by their founders [Greespan 1967, 1997] [Luciani 1991], physically-based particle modeling is a modeling concept based on the explicit duality of variables, extensive variables (EV) and intensive variables (IV) , and the basic action-reaction principle.

It allows to understand the physical modeling, not only as a representation of nature, but as an abstract representation system by which we describe algebraically the dynamic correlation between two (and further any number of) dynamic phenomena, whatever they are, this algebra being based on two dual variables: one (EV) describing the intrinsic evolution of the phenomenon from the influences (IV) of all the others phenomena, and one (IV) describing the mutual influence between each pair of them from the evolution of extensive variables (EV). All the rules that are involved to model a dynamical system are rules that links EV and IV. These rules can be called "physical rules". We can notice that natural phenomena are obviously represented (modeled) in Physics (Mechanics, Electricity, etc.) by these types of abstract rules.

From this abstract point of view, a physically-based particles model is a network of dynamic automata, similar to the well-known Kirschoff's network in Electricity, in which behavioral differential components producing extensive variables are linked by differential interaction components producing intensive variables. This type of network can be seen as a type of cellular automata calculating real states instead of logic states. Several works show that such methods are able to produce any types of motions: from displacements of rigid or articulated bodies [Nouiri 1994] [Chanclou 1995] [Chanclou 1996] [Jimenez 1993] to complex emergent dynamic phenomena (as crowd behaviors [Luciani 2003]), including all types of deformations, complex motions as chaotic or non-linear evolutions (avalanches, collapses, fractures, etc.) [Luciani 2000] and all the various states of the matter (fluids, gas, solid, pastes, etc.) [Luciani 1995b] [Luciani 1995a]

Note : Direct and Inverse Dynamics
In Computer Graphics, literature states frequently on direct and inverse dynamics. Direct dynamics calculates extensive variables as positions (variables to be displayed) from intensives variables as forces. Inverse dynamics calculates the intensive variables from extensive variables, and it is mainly used as a control process to calculate the forces to be applied to reach a given position. As physically-based models integrates the two stages and are controllable either by positions or by forces, we don't examine here these two particular cases of computation as general method to produce motion.

4.2.2 Behavioral models

Models based on artificial intelligence or artificial life approaches are the second type of generative models used in computer motion modeling and synthesis. They were initially used to model mainly behaviors of living organisms. Similarly with physically-based models, they can be used as an abstract representation to models evolutions, whatever these evolutions are. Physically-based models are used to model physical phenomena but also the dynamics of non physical phenomena. Behavioral models are used to model autonomous evolutions, for living or non living beings. The complementarity between these two types of models has been well expressed by Newman and Comper [Newman 1990] in the case of living tissues modeling. They called the first type of model (physically-based model) "generic systems" and the second "genetic systems". Generic mechanisms are defined as those physical processes that are broadly applicable to living and non-living systems, such as adhesion, surface tension and gravitational effects, viscosity, phase separation, convection and reaction-diffusion coupling. They are contrasted with 'genetic' mechanisms, a term reserved for highly evolved, machine-like, bio-molecular processes. Generic mechanisms acting upon living tissues are capable of giving rise to morphogenetic rearrangements. Many morphogenetic and patterning effects are the inevitable outcome of recognized physical properties of tissues, and generic physical mechanisms that act on these properties are complementary to, and interdependent with genetic mechanisms. Major morphological reorganizations may arise by the action of generic physical mechanisms, that could be stabilized and refined by subsequent evolution of genetic mechanisms.
There are two major types of behavioral models: Genetic algorithms and Agent-based models.
Genetic algorithms, as L-systems or cellular automata, aim to model developmental processes based on genetic evolutions. Agent –based models are mainly based on implementation of perception-decision-action processes, to model autonomous behaviors. In computer graphics they are widely used in modeling of living growing [Prusinkiewicz 1993, 1999, 2002] [Lindenmayer 1992], living behaviors, evolutionary processes, morphogenesis processes, autonomous behaviors [Sims 1991, 1992, 1994] [Musse 1999], and emergent cooperation between actors [Panatier 1998] [Heguy 2001] [Sanza 2000].

4.2.3 Main properties from an enactive point of view

As said before, Computer Graphics was mainly oriented, during its first stages, in geometrical and light modeling and computing. The main features of scenes and objects that have been taken into account are morphological and visual features. Physically-based models introduced the physical properties of objects in virtual objects and scenes, with new variables as intensive variables (forces). In addition to their ability to produce easily higher quality in complex motions, the introduction of physically-based models cause two shifts, at the conceptual as well as at the pragmatic level:

• Firstly, they allow to introduce mechanical retroactive transducers for the gestural manipulation. In order to be felt, the matter has to be modeled and simulated. Geometrical models are not able to generate forces. The only case in which geometrical models can be used to generate force is to render an infinitely rigid shape by predetermined potential fields. Thus, physically-based modeling is the core component for retroactive action and the physical manipulation of virtual objects. The ergotic interaction [Cadoz 1994, 2000], that needs force feedback devices and force computation, i.e. physically-based modeling, has been the last to be introduced in computerized environments. From the enactive point of view, enactive interfaces cannot avoid such interaction, which plays a complementary role of pure epistemic-semiotic interaction loops between pure non retroactive action and vision (and audition).

• Secondly, up to now, the main stream in computer graphics and VR is devoted to geometrical modeling, the motion being mapped on predefined shapes. The shapes were represented only through their geometrical (no-matter) features, that are basically static representations. Physically-based modeling introduces the matter as a core component of object representation. The complementarity of the two parts (geometry and matter) of the objects modeling are not obvious to define and implement. Only physical models that are based on geometrical features, as meshes or lattices, allow a direct and compatible link with the geometrical representations of shapes. But, as we said, they are limited to continuous matter and they do not allow a wide use of non-continuous behaviors (collisions, fractures, manipulation of pastes or fluids, etc.). These types of physical objects cannot be avoid neither in VR systems nor in telemanipulation or telecommunications. In addition, physically-based models are able to produce more complex unformed and highly evolving shapes as in such complex behaviors, than those produced by pure geometrical modeling and kinematics.

- ➔ Faced to the available computational power and its foreseen evolution, will these different processes, one focusing on the shape and the vision, another focusing on the matter, the motion and the manipulation, pragmatically compatible in enactive interfaces?
- ➔ Faced to the underlying concepts – one focusing on the predetermined geometrical properties of space, another focusing of the dynamic construction of such properties – will there different concepts conceptually compatible in enactive interfaces?
- ➔ What types of typology of tasks, can we define in which an optimal use of each of them could be designed and implemented?

4.3 References

[Lasseter 1997] J. Lasseter "Principles of Traditional Animation Applied to 3D Computer Graphics," SIGGRAPH'87, pp. 35-44.

4.3.1 Phenomenological models of motion

[Steketee 1985] S. N. Steketee, N. I. Badler . Parametric keyframe interpolation incorporating kinetic adjustment and phrasing control. 12th SIGGRAPH. 1985. Pages: 255 - 262

[Burtnik 1976] N. Burtnyk , M. Wein, Interactive skeleton techniques for enhancing motion dynamics in key frame animation, Communications of the ACM, v.19 n.10, p.564-569, Oct. 1976

[Reeves 1981] William T. Reeves, Inbetweening for computer animation utilizing moving point constraints, Proceedings of the 8th annual conference on Computer graphics and interactive techniques, p.263-269, August 03-07, 1981, Dallas, Texas, United States

[Friedrich 1998] Axel Friedrich , Konrad Polthier , Markus Schmies, Interpolation of triangle hierarchies, Proceedings of the conference on Visualization '98, p.391-396, October 18-23, 1998, Research Triangle Park, North Carolina, United States

[Coquillart 1990] Sabine Coquillart. Extended free-form deformation: a sculpturing tool for 3D geometric modeling. SIGGRAPH 1990. Pages: 187 - 196

[Frish 2002] Norbert Frisch , Thomas Ertl, Deformation of finite element meshes using directly manipulated free-form deformation, Proceedings of the seventh ACM symposium on Solid modeling and applications, June 17-21, 2002, Saarbrücken, Germany

[Coquillart 1991] Sabine Coquillart , Pierre Jancéne, Animated free-form deformation: an interactive animation technique, ACM SIGGRAPH Computer Graphics, v.25 n.4, p.23-26, July 1991
[Lamousin 1994] Henry J. Lamousin , Warren N. Waggenspack Jr., NURBS-Based Free-Form Deformations, IEEE Computer Graphics and Applications, v.14 n.6, p.59-65, November 1994
[Milliron 2002] Tim Milliron , Robert J. Jensen , Ronen Barzel , Adam Finkelstein, A framework for geometric warps and deformations, ACM Transactions on Graphics (TOG), v.21 n.1, p.20-51, January 2002

4.3.2 Generative models of motion : Physically-based models

[Greenspan 1973] Greenspan, D. Discrete Models. Reading in Applied Mathematics. Addison-Wesley, 1973
[Greenspan 1997] Greenspan, D. Particle Modeling. Birkhauser Ed., 1997.
[Metaxas 1996] Dimitris Metaxas. Physics-based deformable models: Applications to computer vision, graphics and medical imaging". Kluwer Academic. 1996.
[Terzopoulos 1991] D. Terzopoulos, J. Platt, K. Fleischer. "Heating and melting deformable models”. *The Journal of Visualization and Computer Animation,* **2**(2), 1991, 68-73.
[Terzopoulos 1993] D. Terzopoulos, D. Metaxas, "Dynamic 3D models with local and global deformations: Deformable superquadrics," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **13**(7), 1991, 703-714. Special Issue on Physical Modeling in Computer Vision. Reprinted in *Model-Based Vision*, H. Nasr (ed.), SPIE, 1993.
[Terzopoulos 1988a] D. Terzopoulos, K. Fleischer. "Modeling inelastic deformation: Viscoelasticity, plasticity, fracture,", *Computer Graphics*, **22**(4), *Proc. ACM SIGGRAPH'88 Conference,* Atlanta, GA, August, 1988, 269-278.
[Terzopoulos 1988b] D. Terzopoulos, J. Platt, A. Barr, K. Fleischer, "Elastically deformable models," *Computer Graphics*, **21**(4), 1987, 205-214, *Proc. ACM SIGGRAPH'87 Conference,* Anaheim, CA, July, 1987. Translated to Japanese by Nikkei-McGraw-Hill and published in *Nikkei Computer Graphics*, **3**(18), 1988, 118-128.
[Baraf 1992] D. Baraff and A. Witkin. Dynamic simulation of non-penetrating flexible bodies. *Computer Graphics* 26(2): 303-308, 1992
[Baraf 1991] D. Baraff. Coping with friction for non-penetrating rigid body simulation. *Computer Graphics* 25(4): 31-40, 1991
[Luciani 1991] LUCIANI (A), JIMENEZ (S), FLORENS (JL), CADOZ (C) & RAOULT (O), "Computational physics : a modeler simulator for animated physical objects", Proceedings of the European Computer Graphics Conference and Exhibition. Vienna, Austria, Septembre 91, Editeur Elsevier
[Nouiri 1994] NOUIRI (J) , CADOZ (C), LUCIANI (A), "The physical modelling of complex physical structures. The mechanical clockwork". - 5th. Eurographics Workshop on Animation and Simulation. Ed. Gérard Hégron and Olov Fahlander. SINTEF, Oslo, Norway Sept 17-18 1994
[Luciani 1995a] LUCIANI (A), HABIBI (A), MANZOTTI (E) - "A Multi-scale Physical Models of Granular Materials", Proc. of Graphics Interface '95, 16-19 May 1995, Quebec City, Canada - pp136-146
[Luciani 1995b] LUCIANI(A), HABIBI (A), VAPILLON (A), DUROC (Y) - «A Physical Model of Turbulent Fluids", 6th Eurographics Workshop on Animation and Simulation- Maastricht 1995 - Springer Verlag Ed. "Computer Science Series - pp16-29
[Chanclou 1995] CHANCLOU (B), LUCIANI (A) - «Physical models and dynamic simulation of planetary motor vehicles with a great number of degrees of freedom»- Proc of IAS 4 Conf. - IOS Press - 1995 - pp 465-472
[Chanclou 1996] CHANCLOU (B), LUCIANI (A), HABIBI(A), "Physical models of loose soils marked by a moving object" - Proc. of Computer Animation 96 - IEEE computer Soc Press - 1996 - pp36-46
[Luciani 2000] LUCIANI A., "From granular avalanches to fluid turbulences through oozing pastes : a mesoscopic physically-based particle model. Proceedings of Graphicon Conference. Moscow. Sept. 2000.
[Habibi 2001] HABIBI A., LUCIANI A.. « Dynamic Particle Coating ». IEEE Transactions on Visualisation and Computer Graphics. Dec 2001.

[Jimenez 1993] JIMENEZ (S), LUCIANI (A), "Animation of Interacting Objects with Collisions and Prolonged Contacts", Proceedings of the IFIP WG 5.10 Working Conference. Tokyo April 1993
in Modeling in Computer Graphics - B. Falcidieno & T.L. Kunii Ed., Springer Verlag Pub. – 1993
[Luciani 2003] LUCIANI A, CASTAGNE N. A Physically-based Particle Model of Self-organized Emergent Effects in Collective Phenomena. Proceedings of Graphicon Conference. Moscow. Sept. 2003.

4.3.3 Generative models of motion : Behavioral and agent-based models

[Newman 1990] S.A. NEWMAN, W.D. COMPER - "'Generic' Physical Mechanisms of Morphogenesis and Pattern Formation." p1-17. Development, Vol 110, Issue 1 1-18, Copyright © 1990 by Company of Biologists
[Lindenmayer 1992] Lindenmayer A, Jürgensen H. 1992. Grammars of development : discrete-state models for growth, differenciation, and gene expression in modular organisms. In: Rozenberg G, Saloma A, eds. Lindenmayer systems. Berlin: Springer-Verlag, 4-21.
[Sims 1994] "Evolving Virtual Creatures"
[Prusinkiewicz 1993] P.PRUSINKIEWICZ, M.S. HAMMEL, E. MJOLSNESS - "Animation of Plant Development." p351-360, Computer Graphics Proceedings, annual Conference Series1993 (SIGGRAPH'93 Conference Proceedings).
[Prusinkiewicz 1999] P. Prusinkiewicz: Modeling Plants and Plant Ecosystems: Recent Results and Current Open Problems. Computer Graphics International 1999: 110-
[Prusinkiewicz 2002] P. Prusinkiewicz: Geometric Modeling without Coordinates and Indices. Shape Modeling International 2002: 3-6
[Sims 1994] K.Sims, "Evolving Virtual Creatures" Computer Graphics (Siggraph '94 Proceedings), July 1994, pp.15-22.
[Sims 1991] K.Sims. "Artificial Evolution for Computer Graphics" Computer Graphics (Siggraph '91 proceedings), July 1991, pp.319-328.
[Sims 1992] K.Sims, Towards a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life , MIT Press, 1992, pp.171-17
[Heguy 2001] Olivier Heguy, Yves Duthen, Alain Berro . *Learning System for Cooperation in a Virtual Environment* . Dans: *SCI?2001 The 5th World Multi-Conference on Systemic, Cybernetics and Informatics* , *Orlando, Florida USA*. 22 juillet 25 juillet 2001.
[Musse 1999] Soraia Raupp Musse , Marcelo Kallmann , Daniel Thalmann, Level of Autonomy for Virtual Human Agents, Proceedings of the 5th European Conference on Advances in Artificial Life, p.345-349, September 13-17, 1999
[Panatier 1998] Cyril Panatier , Hervé Luga , Yves Duthen, Collective learning for spatial collaboration, Proceedings of the fifth international conference on simulation of adaptive behavior on From animals to animats 5, p.477-482, September 1998, Univ. of Zurich, Zurich, Switzerland
[Sanza 2000] Cédric Sanza , Cyril Panatier , Yves Duthen, Communication and Interaction with Learning Agents in Virtual Soccer, Proceedings of the Second International Conference on Virtual Worlds, p.147-158, July 05-07, 2000

5 Action processing

5.1 Typology of Actions and link with the vision (common with WP4a)

©Annie Luciani
September 2004

5.1.1 Action. Haptic. Gesture

The term « action » covers different meanings. It states the result of a physical task performed by the human body as well as the way through which this task is performed. Mary M. Smyth & Alan M. Wing [Smyth, Wing 1984] distinguish three levels in performing an action : action refers to what it is done (« drink a glass, pick a pencil »), movements refers to how it is done (the movement, with what movement the glass is drunk) and skill refers to the quality of the movement (how the movement is). A specific action can be acted by several movements. A movement can be colored by several skills. These precise definitions do not correspond easily to the current uses of such terms. Action is often used to name all these features, movement is more general that the movement in the action, and skill refers also to the ability to do something.

Thus, confronted to the multiple and sometimes contradictory meanings of « action » and of « haptic » [Pasquinelli 2004], we will called :

- « gesture » in its usual general meaning, to name all what it can be done by the human physical body, whatever the performed objective, rather than « action » or « movement ».
- « gestural channel » all the sensory-motor apparatus composed of all the physical means, through which the human physical body interacts with the physical external universe: hand, body equipped with all its mechanoreceptors and all its actuators. Gesture is then these human biomechanical sensors-actuators during a physical performance.
- "gestural action" as the motor part of the gestural channel involved in the gestural performance. It involves all the physical components (articulated skeleton and muscles) of body.
- "gestural perception" as the sensory part of the gestural channel. These terms (gestural channel, gestural perception, gestural action) are used to avoid the detailed description of each sub-means (subset of sensors, subset of motor capabilities) as well as the human perceptual and/or cognitive results of the use of these means.

5.1.2 Human-world interaction : three basic functionalities

Having in mind that sensorial events are necessarily produced by physical objects, the sensory-motor relation between humans and environments can be splitted in three basic functions :

- Epistemic function : the function to know the enviroment, from which humans are informed on the environment , by the environment,
- Semiotic function : the function from which humans inform the environment.
- Ergotic Function : the term "ergotic" has been coined by C. Cadoz, [Boissy 1992][Cadoz 1994] to state the relation during which the humans ands the physical worlds are physically interacting, caracterised by the fact that there is a exchange of physical energy between them. The term « Haptic » is often used to state this function. Unfortunately, as stated by E. Pasquinelli [Pasquinelli 2004], this term covers several meanings underlying several different points of view. Conversely, the term « ergotic » coming from « ergos », meaning « physical work, energy », represents clearly the principal property of such function.

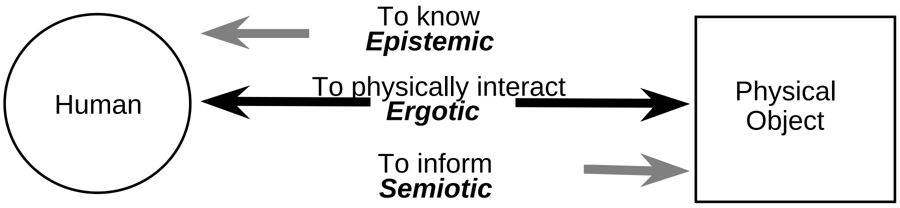


Figure 1. The three functions of the human – environment relationship

Generally, the epistemic function is conveyed by the perceptual apparatus. It is supported as well by the proprio-tactilo-kinesthetic apparatus (mechano and tactile receptors), as the vision and the audition apparatus. Thus, we can speak on the epistemic function of the audition, similarly than the vision as Nivedita Gangopadhyay [Gangopadhyay 2004] proposed with her formal definition of epistemic seeing, and the epistemic function in the haptic sensory modality considered only in its perceptual side, as stated by Hatwell and al. in Touching for knowing [Hatwell, Streri, Gentaz, 2003].

The semiotic function is conveyed necessary by the human channels that are able to emit information to the world. Humans are equipped ONLY BY TWO emitting channels : his/ her mechanical body producing gestural actions (body, arm, hand, face, etc…) and his/her vocal apparatus producing aero-acoustical motions. Some gestural actions aim to transmit pure information (and not energy) to the environment. That is the case of gestures that accompany the speech, the sign language of the deaf-mute, the musical conductor gesture, the gesture of pointing a target with the finger or to point an icon with the mouse, of moving around an object (walking, etc), the cutaneous touch without movements of muscles and joints, of pulling a infinitely light object. Thus the semiotic function is composed of the speech and all the gestures (or motions) the body is able to produce freely (i.e. without any exchange of energy with the external world).

The ergotic function intervenes when physical energy is exchanged during tne interaction between humans and physical world. A specific ability of the gestural channel is to handle directly the matter: to mould it, to transport it, to break it, to cut, to rub, to hit, etc. The hand, (and the all physical body) is in contact with the matter, transmits physical energy to the matter. It applies (and respectively causes) forces, displacements, deformations to the objects and these one react to the human body, resisting to its energetic transfer, and retroacting a part of it. This type of ergotic interaction, aims not only to inform the external world and to be informed by it, but to transformed it. That is possible through a specific property of the gestural channel to be intrinsically bi-lateral: to act on and to perceive, in a inseparable way. During ergotic interaction, and simultaneously of the energetic exchange, humans know (epistemic function) and inform (semiotic function).

The following figure (figure 2) draws all the human – environments interaction channels according to the proposed typology of interaction functions.

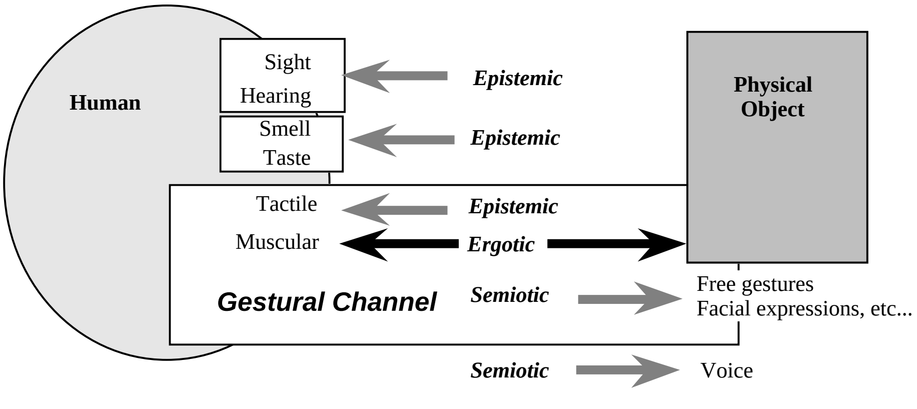


Figure 2. The human interaction channels

In the following, we don't consider the epistemic senses of taste and smell (ambient and localized chemical contact) because of their low relationship with gestures.

5.1.3 The action – perception interaction loops

From the typology of functions in the human-world interaction, several ways of action – perception loops can be declined according to a non redundant criteria of : the existence or not of an energy continuum (or energetic consistency) between the gesture and the perceived phenomena.

5.1.3.1 Pure epistemic-semiotic loops

There are loops linking two grey arrows in figure 3 (sub components of figure 2), in which emission of information from the human subject (to the world) and reception of information by the human subject (from the world) are correlated but without energy exchanges:

- from the semiotic gestural action to the epistemic seeing,
- from the semiotic gestural action to epistemic hearing.
- From the semiotic gestural action (free gestures, facial movements, etc.) and epistemic gestural perceptions, as in cutaneous touch in which there is any noticeable muscular energetic activity in the result of the performance the result of the action. In such actions, the muscular energetic activity can be neglected, or mediated by tools that decrease it without no noticeable loss in the performance of the task.
- From voice to seeing and hearing

Examples are: pointing an object, moving to see or to hear, reading, navigating in a data base or in a virtual environments by means of non retroactive sensors as sticks, mouse, triggering a sounding signals from a non retroactive sensor, selecting an object or an icon, conducting an orchestra, etc. In such action-perception loops, the perceptual result depends on the action but the physical states of the interacting bodies are not modified during and by the interaction process. There are not action-perception loops aiming to act on the world. Mainly there are rather « exploratory activities » oriented to the knowledge of the world or symbolic activities oriented to symbolic constructions.

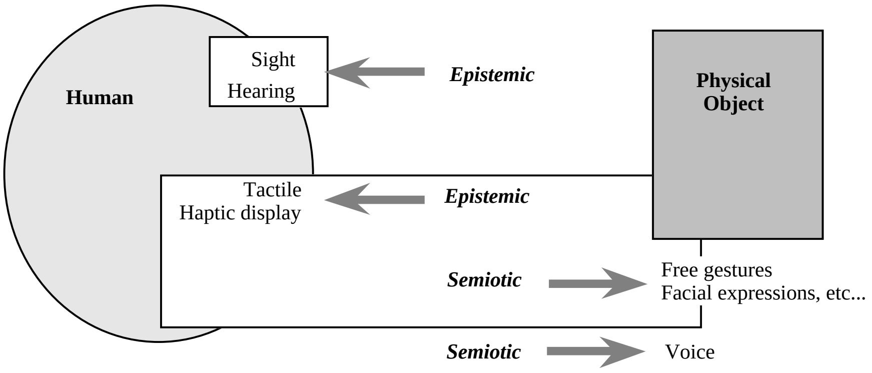


Figure 3. The pure semiotic-epistemic interaction loops

5.1.3.2 Instrumental loop: ergotic interaction and multisensory epistemic feedbacks

It is not sufficient to loop grey arrows as in the pure epistemic-semiotic loops described above to obtain the characteristic property of the ergotic function that is to underlie energy' exchanges (Figure 1, black arrow) between the interacting bodies during the interaction. This means that the ergotic interaction can be clearly-cut distinguished among the others. This points out the operationality of such term to categorize interactions, and consequently to categorize tools and media supporting interaction between a human and his/her external universe. More precisely, the criteria is not the energy spent by an individual during an action, but the energy exchanged between the two interacting bodies, or from the point of view of the human, the energy transferred from (resp. to) human to (and from) object, that is a necessary condition to physically modify the world. As said before, all the handled activities fall in such category: grasping, pushing, pulling, cutting, throwing, carrying, molding, hitting, rubbing, breaking, displacing an infinitely heavy object, writing.

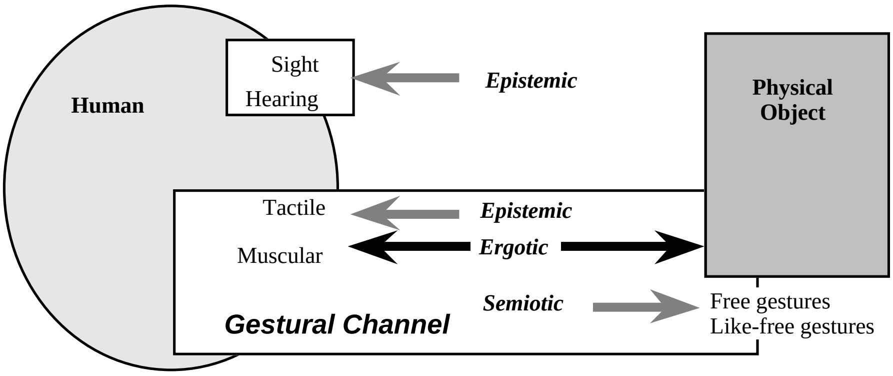


Figure 4. Instrumental Interaction

When one manipulates in such ergotic interaction an object, the physical states of objects (and of humans) are modified by the interaction, exhibiting new mechanical behaviors depending on the interaction (sounds, deformations, fractures, etc…). Thus the sensory epistemic returns (mainly sight and hearing) inform the human of the behavioral answers of the object to the gestural actions. This means that sensory stimuli are not considered by themselves (as conventionally considered by multimodality), but as physical responses that objects do not exhibit without energetic manipulations. These responses encode the coupled system "human body-physical object". They inform the humans of the physical objects and of its physical coupling to the human body. For example, the sound is the encoding of the human / object system during the performance. The visual motion (displacements and deformations) is the encoding of the human / object system during the manipulation, etc. We can state that the physical object transforms the gesture space in auditory (resp. visual) space.

The gestural perception informs about the human/object systems that are in contact, etc. The sensorial space is then seen as:

- intrinsically multisensorial : a priori and at least composed of ergotic interaction (with its action and perception part) and acoustical and/or visual returns.
- aiming to know the coupled system object – human.

This means that the object is known (1) through the answers of the matter (that can be sensed) to the gestural actions and (2) all these sensory answers have to be considered a priori as a system encoding the couple human-object and proposing invariants of this system (if they exist) at our cognition and our instrumental ways of control.

We called this typical very common and frequent situation « **instrumental interaction** », with all its declensions (instrumental situation): to dig over the ground, to mould the paste of the bread, to crumple a paper sheet, to play violin, etc.

The instrumental situation is characterized by two necessary features:

- The interaction presents an ergotic component
- The relation between the sensory returns and the gesture exhibits an energetic consistency.

5.1.3.3 Conclusion

The proposed typology is operational in the sense that it proposes one criterium that organizes the different ways of interaction between humans and external world, in non-overlapped categories. This criteria is the existence, or not, of an energy continuum (energetic consistency) between the gesture and the perceived phenomena.

5.1.4 References

[Smyth 1984] Smyth, Wing. « The psychology of Human movement. » Academic Press, 1984.
[Boissy 1992] Jacques Boissy. Cahier des termes nouveaux. Institut National de la Langue Française, Conseil International de la Langue Française (CILF) and CNRS Editions. 1992. page 52.
[Cadoz 1994] Cadoz C. Le geste, canal de communication instrumental. techniques et sciences informatiques. Vol 13 - n01/1994, pages 31 à 61. 1994.
[Cadoz, Wanderley 2000] Claude Cadoz, Marcello M. Wanderley (2000). Gesture-Music, in Trends in Gestural Control of Music, M. M. Wanderley and M. Battier, eds, ©2000, Ircam – Centre Pompidou, pp. 71-94
[Pasquinelli 2004] Elena Pasquinelli. Some definitions and problems of classification. Enactive WP4b State of the Art
[Gangopadhyay 2004] Nivedita Gangopadhyay. Epistemic seeing. WP4b Enactive State of the Art.
[Hatwell, Streri, Gentaz, 2003] Hatwell Y., Streri A., Gentaz E.. "Touching for knowing : Cognitive psychology of haptic manual perception". John Benjamins Ed.. 2004.
[Luciani 2004a] Luciani A., Urma D., Marlière S., Chevrier J. (2004). PRESENCE : The sense of believability of inaccessible worlds. Computers & Graphics. Elsevier Ed.. Vol 28/4 pp 509-517.
[Luciani 2004b] Luciani A. (2004). Dynamics as a common criterion to enhance the sense of Presence in Virtual environments. Conference Presence 2004. Oct. 2004. Valencia. Spain. To be published.
[Luciani 2004c] Luciani A. (2004). Interaction as exchanged actions and their role in visual and auditory feedbacks. Enactive Virtual Workshop. Enative project.

5.2 Input gesture processing

© Damien Couroussé
September 2004

5.2.1 Motion capture

The Motion Capture systems have been developed to record the movements of human beings. The motion capture technique is essentially used by the animated computer graphics field, either for differed time applications such as the animation of virtual humanoids in movies or video games, or in real time applications, such as artistic performance [Gualtiero, Antonio Camurri, Barbara Mazzarino 2004, WP4b Enactive State of the art] or motion analysis for research purpose.

The motion capture technique corresponds to the recording of the *movements* of an object, not its visual appearance [Menache, 1999]. Therefore, the hypothesis is made that the movement of such an object may be obtained only by the observation of specific points of the recorded object. hence sensors are fixed on these points of the object considered as relevant for this purpose. In the general case of the human body, these relevant points are the joints, such as knee, shoulder, elbow, and extremities of members, such as the top of the head, hands and feet.

5.2.1.1 Main techniques used in Motion Capture

Optical systems

In that kind of system, which is one of the most used today, the data are provided by (mostly infrared) cameras which record the movements of reflective markers. With the help of two cameras, it is possible then to reconstruct in three dimensions the recorded movements. In most cases, in order to ensure a good reconstruction of the geometrical positions, and to avoid 'phantom' points that come from reflections, it is necessary to use between 7 and 24 optical recording systems. However, the main drawbacks of that kind of system are the fact that the visual field before the cameras has to remain free, and the fact that the amount of post-processing that have to be done after recording may interfere with a real time performance.

Electromagnetic systems

Magnetic data, if they are not so often used as optical ones are, have some advantages. The main is that orientation of the recorded points may be obtained, and the measure of position is absolute. Furthermore, the magnetic field created by the sensor system defines a zone in which every movement is possible, without the inconvenient of masking the field of one the sensing systems. Therefore, the number of magnetic sensors is often reduced to three. At last, there is no post-processing of data, since the position and orientation of each points are immediately obtained. This allows for real-time applications.

Mechanical

The performer wears a human-shaped set of straight metal pieces (like a very basic skeleton – MetaMotion) that are hooked onto the performer's back. As the performer moves, this kind of exoskeleton is forced to move as well and sensors allocated to each joint provide of measure of the rotation. This technique is interesting when it is not possible to avoid with light or magnetic interference, but it does not allow for an absolute measure of movements, and the displacements of the performer are not immediately obtained. Moreover, the equipment may hinder the performer in its movements.

Other types of mechanical motion capture sets have been developed for specific parts of the body: gloves, mechanical arms, or articulated models such as monkeys or face puppets (PuppetWorks), ("key framing" technique), with which the performer interacts to mimic movements.

5.2.1.2 Post-processing

Once the motion data are obtained, it is necessary to export them in a suitable format in order to have them manipulated in computer animation softwares. A *structure* is associated with the motion data, which actually represents the skeleton of the actor, and gives the way the points are related one to each other.

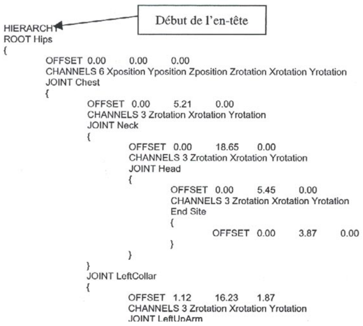
Since the motion capture technique was especially devoted to the recording of human movements, the structure of file formats commonly used permits mainly (and maybe only) the representation of human beings movements.

BVA and BVH

BVH is the acronym for Biovision Hierarchical Data. This file format, which is one of the most commonly used nowadays, has been developed by Biovision. The BVA file format appeared later as a refund of BVH.

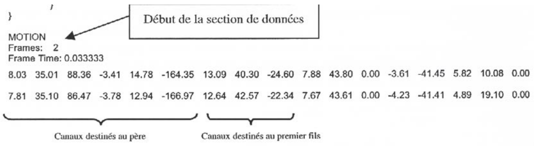
The BVH and BVA file formats were especially designed for the movement representation of humanoid forms. BVA files are filled with ASCII information, and contain two sections. The first one is dedicated to the description of the structure of the data. In that section are defined the hierarchy of the skeleton with simple keywords such as OFFSET or CHANNEL. In this section are defined the positions of each *segments* to the origin or to the upper segment is the skeleton hierarchy. The second section is composed of the raw data describing the decomposed coordinated of each segment declared in the first section. The data are described frame after frame, which means that it is possible to obtain the whole configuration of the skeleton at a given time.

One of the main drawbacks of that file format is that it does not allow for an absolute position, nor it allows for explicit rotation information. It is therefore devoted to inverse cinematic processing, but does not allow for a direct interpretation of the data.


Début de l'en-tête

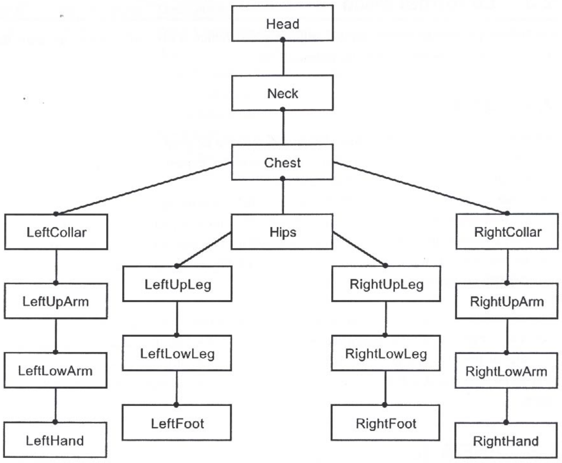
```
HIERARCHY
ROOT Hips
{
        OFFSET  0.00        0.00        0.00
        CHANNELS 6 Xposition Yposition Zposition Zrotation Xrotation Yrotation
        JOINT Chest
        {
                OFFSET   0.00        5.21        0.00
                CHANNELS 3 Zrotation Xrotation Yrotation
                JOINT Neck
                {
                        OFFSET   0.00        18.65       0.00
                        CHANNELS 3 Zrotation Xrotation Yrotation
                        JOINT Head
                        {
                                OFFSET   0.00        5.45        0.00
                                CHANNELS 3 Zrotation Xrotation Yrotation
                                End Site
                                {
                                        OFFSET   0.00        3.87        0.00
                                }
                        }
                }
                JOINT LeftCollar
                {
                        OFFSET   1.12        16.23       1.87
                        CHANNELS 3 Zrotation Xrotation Yrotation
                        JOINT LeftUpArm
```

The BHV file format - beginning of the first section


Début de la section de données
Canaux destinés au père
Canaux destinés au premier fils

```
}
MOTION
Frames:    2
Frame Time: 0.033333
8.03   35.01   88.36   -3.41   14.78   -164.35   13.09   40.30   -24.60   7.88   43.80   0.00   -3.61   -41.45   5.82   10.08   0.00
7.81   35.10   86.47   -3.78   12.94   -166.97   12.64   42.57   -22.34   7.67   43.61   0.00   -4.23   -41.41   4.89   19.10   0.00
```

The BHV file format - beginning of the first section



An example of hierarchical structure for motion capture cata

ASK/SDL

This file format is a variant of the BVA file format. The ASK (Alias Skeleton) file contains only the information related to the hierarchy of the skeleton, with absolute coordinates.

The motion information is included in the SDL file. It allows many supplementary information, for example for the description of the scene.

AOA

Adaptative Optics Associates is dedicated to the creation of hardware support for motion capture. The file format created by AOA is written in ASCII, and allows for very simple manipulation.

The first section is a simple header composed of two lines, which includes comments, the number of frames, the number of markers per sample, and the sampling frequency. The second section contains the motion data. Each line contains one sample of each of the markers. The interest of this format remains in its simplicity: in the second section, the association of a data to a marker is implicit.

ASF/AMC

This file format was developed by Acclaim Inc., which is involved in video games. It has then been redeveloped by Oxford Metrics (Vicon Motion Capture Systems) when it was put in the public domain.

This format is composed of two files, the first one (ASF file format) for the description of the skeleton, and the second one (AMC –Acclaim Motion Capture– file format) for the raw data information. Its interest remains in the fact that it is possible to associate one skeleton files to many collections of motion capture data describing the same performer in as many motion capture sequences.

BRD

This file format is dedicated to the unique usage of the Ascension Technologies "Flock of Birds" motion capture system, developed by LambSoft. It only allows the recording of data arising from a magnetic system.

HTR

HTR is the acronym for Hierarchical Translation-Rotation. This file format was developed as a native format for the skeleton of the Motion Analysis software. It was made as an alternative to the BVH file format in order to make up for its main drawbacks.

The HTR file format contains four sections contained in one single file: header, hierarchy and name of the existing segments, intial position and motion data. In addition to the information available in the BVH file format, the HTR first section allows to define the number of segments, the order of disposition of the Euler angles, the calibration unities, the rotation unities, the gravity axis, the default axis along which are aligned the skeleton segments, and a globa scale factor.

The interest if this file format, compared to the BVH, is that the hierarchical information is provided independently from the characteristics of each segment, which simplifies its reading and treatment.

National Institute of Health 3D

The C3D file format is born from a common need of the research laboratories working on Clinical Gait, Biomechanics and Motion Capture studios to have a universal format for the exchange of motion information. It is thus a public domain file format.

The required properties of this format were the following:

- The ability to store 3D and analog data in an unprocessed form. It is not essential that data is stored without processing, but the format needs the ability to support raw coordinate and analog sample data.
- Preserve information that describes the physical design of the laboratory such as EMG channels used, force plate positions, and marker sets etc.
- Store Trial information relating to the circumstances of the test session such as sample rates, filenames, dates, EMG muscles recorded etc.
- Store Patient information - name, age at trial, with physical parameters such as weight, leg length etc.
- Store calculated analysis results such as gait timing, cycle information and related information.
- Flexibility and compatibility - it must provide the ability to store new information without making older data obsolete.
- A public specification and format description so that anyone can access data without depending on a manufacturer for information.

The C3D file format is therefore a standard that we should classify apart from the other formats, because of its ability to stock a great amount of data and its capacity to comply to very different needs and skeleton structures.

5.2.1.3 Towards standards for motion formats

The MPEG-4 norm, especially developed in the aim to provide an object oriented coding format for the audiovisual scenes, proposed a relatively elaborated model of humanoid objects. This kind of model was optimized in order to allow low transmission costs but a good fidelity for the reconstruction of the scene (i.e. a good replica of human bodies and of their movements) at the reception side.

The MPEG-4 protocol of transmission is pretty inspired by the general form of most of the file formats presented above: the structure and the initial position of the animated bodies is transmitted at first; this first phase provides information about the skeleton composition, surface and texture information for the

reconstruction of the whole animated scene, etc. The second phase of the transmission then consists in an update of the pre-defined objects.
The standard MPEG-4 definition of the humanoid structure has the following characteristics:

- 6 degrees of freedom (dof) for the whole skeleton
- A total of 62 dof for the internal movements of the skeleton
- The hand structure, optional, is defined apart, and contains 25 dof
- The skeleton is divided into 29 segments without hands, or 59 segments if we include the hands.

This illustrates the great complexity and the great specificity of the human skeleton structure, regarding to most of the objects included in animated movies, video games, etc.

Therefore, because of the great specificity of human movements and the great difficulty to render them correctly, motion capture techniques are especially intended in the aim of representing human movements. The motion capture hardware is designed for the recording of human movements, and the software formats reflect this in the way they are conceived.
Furthermore, the relation between the techniques for capturing movements and rendering tools (computer animation software) is based *only on a geometrical purpose*: it means that motion capture techniques deal only with *extensive variables*, i.e. *displacements*, and force variables are less considered in that particular field of computer animation.

We may state that this particular action-vision chain is proceeding of a kind of mapping between the representing humanoid form and the represented object, which is the result of the motion capture process; therefore, the relation between action and vision is opposed to the enactive one since it somehow considers sensed motion as a determined data independent of the behavior of the controlled object, *and therefore do not admit interaction in which such data are influenced by the motion of the manipulated object.*

5.2.1.4 References

[Menache, 1999] A. Menache. Understanding Motion Capture for Computer Animation and Video Games. HandBook. Morgan Kaufmann, 1999.

Optical systems:

Adaptive Optics Associated:	http://www.aoainc.com
Mikromak GmbH:	http://www.mikromak.com/
Motion Analysis Corporation:	http://www.motionanalvsis.com/
Vicon Motion Systems:	http://www.vicon.com

Magnetic systems:

Ascension Technology Corporation:	www.ascension-tech.com
Euclid Research:	www.euclidres.com
Data glove iReality:	www. genreality. com
Polhemus:	www.polhemus.com

Hybrid systems:

MetaMotion:	http ://www.metamotion.com
Digits'n Art:	http://www.DnAsofl.com/
PuppetWorks:	http://www.puppetworks.com/index.htm

File formats:

Biovision Motion Capture Services:	http://www.biovision.com
Format BVIH:	http://www.es.wisc.edu/ginphics/Courses/cs-838-l999/Jeff/BVH.html
Format C3D:	www.c3d.org

Format IHTR, GTR: http://www.cs.wisc.edu/graphics/Courses/cs-838-1999/Jeff/{HTR.html, TRC.html}
LIVE Motion Control Platform: http://wscg.zcu.cz/wscg2001/Papers 2001 /R240.pdf
INRTA: *A Global Framework for Motion Capture* http://www.inria.fr/iTrt/rr-4360.html
A proposal for body animation: http://ligwww.epfl.ch/mpeg4/
http://tigwww.epfl.ch/~boulic/mpeg4_snhc_body/body_prop.html

5.2.2 Gesture recognition

We must care the proofreader that this chapter is intended only to give a general approach of gesture recognition, and not to give a detailed view of this complex and still growing field of research, as it does not concern immediately the enactive understanding of gesture. This analytical part might lead to future developments.

In the first 80's, Christopher Schmandt and Eric Hulteen developed at the Architecture Machine Group at MIT a system which allowed the user to control displayed objects on a screen with the complementary help of voice and gesture [Schmandt & al., 1982]. A voice recognition system was combined with a gesture device (ancestor of MoCap devices), which could record the position of the user's finger is space. Sitting on a chair, the user was able to control the display of objects on the screen by pronouncing key-words: "put-that-here"; as each keywords was pronounced, the user had to point its finger at a location on the screen.
With the design of that new kind of device, a user was then able to use speech to control a computer interface, but was also able to interact with a computer interface thanks to a semantic coding associated with gestures.
This opened to new kind of interfaces commonly currently called gesture recognition interfaces.

The primary goal of gesture recognition research is to create a system which can identify specific human gestures and use them to convey information or for device control. The most common applications of gesture recognition are sign language recognition [Ong et al., 2003] [Vogler et al., 2003], device control or interactive approaches of computer interfaces or large VR environments [de la Rivière, 2003]. The general approach of gesture recognition is rooted into the idea of bringing semantic interaction with computer thanks to gesture. Therefore, gesture recognition systems are designed to mainly recognize hand (and arm) and head movements.

The seductive general tendency is to leave the user free of its movements, i.e. contrary to the motion capture, the aim of gesture recognition research is to use human motions as an input of a computer interface without obliging the user to wear a suit or to hold a tracking-tool. Therefore, it is necessary to design systems that have the ability to recognize human presence (static postures) or motion in a video frame.

Although there is no common definition and meaning under the word *gesture*, it is generally admitted in that it encompasses [Brockman, 2003]:

- Static gestures, which are called *postures*.
- Dynamic gestures, which are called to be *motions*.
- Pointing gestures, which are based on a specific location of a limb.

A posture – of the hand for instance – is generally understood as a specific position, or a characterized deformation of the hand and bending of the fingers. In other words, a given specific position of the hand will be said to be a posture if the hand's position can be recognized as one previously defined.

A general overview of a gesture recognition system (Figure 1) might be the following [McGlaun et al., 2003]:

- Video frames are loaded into computer from the recording of a video camera.

- The *segmentation* phase then consists in localizing candidates in the video image for the position of the user's head or hand.
- The gesture features are extracted from the continuous video frame
- Once the motions are recognized, a kind of classification is made in order to determine which gesture did perform the user.

In some applications, the camera is moving as well in order to localize the position of the user's hand [Brockman, 2003].

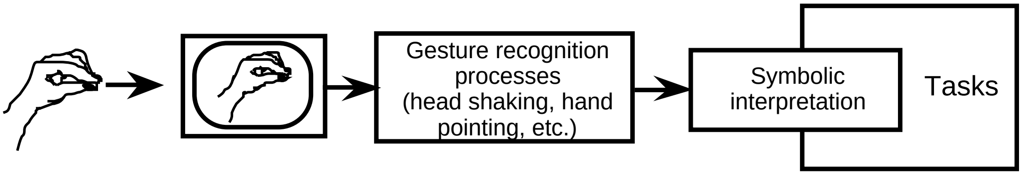

Gesture recognition processes (head shaking, hand pointing, etc.)
Symbolic interpretation
Tasks

Figure 1. General overview of gesture recognition framework

Many extraction techniques are available for the extraction of gestures in a video frame [Yang, 2002]. *Knowledge-based top-down method*, *bottom-up feature-based method* and *template matching* are based on correlation methods: given a standard face pattern (usually frontal), the correlation values are obtained from an input image and the standard patterns. The existence of a face is determined based on the correlation values. Some other applications, which focus on applying analytical methods for breaking down motion sequences and recognizing patterns, use stochastic algorithms techniques combined with Hidden Markov Models [McGlaun et al, 2003], or Bayesian networks [Ong and al. ,2003]. [Schmidt and al., 2003] based their approach of motion filtering method on a dynamic model with a control system for the arm: the idea is that recognition of human arm motion can be improved by the knowledge of the dynamic control performed by the human motor system. This new approach actually deals with motion recognition technologies mixed with models of human motricity.

The first studies on gesture recognition tried to extract simple components of human movements, such as in sign language recognition or in gesture communication. Such a typical example is [McGlaun et al., 2003], where a corpus of "gesture-words" is constructed by simple head-movements along one degree-of-freedom. The obtained movements are mainly purely rotational movements: moving head on the left, right, up, down, bending left and right, and at last head nodding and shaking. [Brockman, 2003] combines dynamic gestures (i.e. movements) of the hand with "static hand gestures", i.e. hand postures and pointing gestures, i.e. arm or hand predefined locations.

Some recent works succeed now in addressing grammatical inflections of gesture speech [Wong, 2003]. Once the recognition of the "basic blocks" is achieved, it is possible to detect small inflections in gesture, which are more relevant of a subtle meaning. In other words, that is to create relations between the gesture performer and the computer interface that go beyond the simple but now well-established interaction with elementary gestures.

References

[1] C. Brockmann and H. Müller. Remote vision-based multi-type gesture interaction. In A. Camurri and G. Volpe, editors, *5th International Gesture Workshop, GW 2003, Genova, Italy*, pages 198–209. Springer-Verlag Heidelberg, April 2003.

[2] A. Camurri and G. Volpe, editors. *Gesture-Based Communication in Human-Computer Interaction, 5th International Gesture Workshop, GW 2003, Genova, Italy, April 15-17, 2003, Selected Revised Papers*, volume 2915 of *Lecture Notes in Computer Science*. Springer, 2004.

[3] J.-B. de la Rivière and P. Guitton. Hand postures recognition in large–display VR Environments. In A. Camurri and G. Volpe, editors, *5th International Gesture Workshop, GW 2003, Genova, Italy*, pages 259–268. Springer-Verlag Heidelberg, April 2003.

[4] A. J. Howell, K. Sage, and H. Buxton. Developing task-specific rbf hand gesture recognition. In A. Camurri and G. Volpe, editors, *5th International Gesture Workshop, GW 2003, Genova, Italy*, pages 269–276. Springer-Verlag Heidelberg, April 2003.

[5] G. McGlaun, F. Althoff, M. Lang, and G. Rigoll. Robust video-based recognition of dynamic head gestures in various domains – comparing a rule-based and a stochastic approach. In A. Camurri and G. Volpe, editors, *5th International Gesture Workshop, GW 2003, Genova, Italy*, pages 180–197. Springer-Verlag Heidelberg, April 2003.

[6] S. C. Ong and S. Ranganath. Classification of gesture with layered meanings. In A. Camurri and G. Volpe, editors, *5th International Gesture Workshop, GW 2003, Genova, Italy*, pages 239–246. Springer-Verlag Heidelberg, April 2003.

[7] C. Schmandt and E. A. Hulteen. The intelligent voice-interactive interface. In *Proceedings of the 1982 conference on Human factors in computing systems*, pages 363–366. ACM Press, 1982.

[8] G. S. Schmidt and D. H. House. Model-based motion filtering for improving arm gesture recognition performance. In A. Camurri and G. Volpe, editors, *5th International Gesture Workshop, GW 2003, Genova, Italy*, pages 210–230. Springer-Verlag Heidelberg, April 2003.

[9] C. Vogler and D. Metaxas. Handshapes and movements: Multiple-channel american sign language recognition. In A. Camurri and G. Volpe, editors, *5th International Gesture Workshop, GW 2003, Genova, Italy*, pages 247–258. Springer-Verlag Heidelberg, April 2003.

[10] R. Watson. A survey of gesture recognition techniques. Technical report, Departement of Computer Science, Trinity College, Dublin, 1993.

[11] M.-H. Yang, D. J. Kriegman, and N. Ahuja. Detecting faces in images: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(1):34–58, 2002.

5.3 Expressive gesture

©Volpe Gualtiero, Antonio Camurri, Barbara Mazzarino
DIST, September 2004

In Mixed Reality (MR), Virtual Environments (VEs), and Inhabited Information Spaces (IISs) analysis, processing, and synthesis of expressive content are of paramount importance for enhancing usability and communication with machines. Multimodal interactive systems employing expressive information coming from several channels to build an application designed with a very special focus on the user, and in which interaction with the user is the main aspect and the main way through which the objectives of the application are reached, can have a relevant role in the development of Enactive Interfaces fusing several modalities (and in particular action and vision). The increasing importance of the information related to the emotional, affective sphere in the design of multimodal interactive systems and Enactive Interfaces is demonstrated by lot of studies that in the last ten years focused on emotional processes and social interaction. Consider for example research on Affective Computing at MIT (Picard, 1997), research on KANSEI Information Processing in Japan (Hashimoto, 1997), the EU-IST funded MEGA project (Multisensory Expressive Gesture Applications; Camurri et al. 2004c; www.megaproject.org).

In this framework, performing arts demonstrated to be a key research and application field for multimodal interactive systems (see for example Rowe, 1993, 2001). From the point of view of research, performing arts are an ideal test-bed for works concerning mechanisms for non-verbal communication of affective, emotional, expressive content. For example, in (Camurri et al., 2004a) music and dance performances have been employed for studying expressive gestures and their ability of conveying emotional states (e.g., the well-know and consolidated basic emotions) and of engaging spectators. Expressive gesture (Camurri et al. 2004a) thus assumes a central role in research that focuses on the high-level emotional, affective content gesture conveys, on how to analyse and process this content, on how to use it in the development of innovative multimodal interactive systems able to provide users with natural expressive interfaces (Camurri et al., 2004b).

5.3.1 Expressive gesture

When non-verbal communication mechanisms are mainly involved in interaction, a relevant role is played by information related to the emotional, affective sphere. In this framework, expressive gesture can be considered as a main conveyor of emotional, affective content. While the relevance of movement and gesture as a main channel of non-verbal communication becomes evident and increasing research efforts are devoted to them (see the Gesture Workshop series of conferences started in 1996 and collecting a continuously growing interest), the focus is here centered on the qualities that make a gesture expressive. An attempt of defining the concept of expressive gesture can be found in (Camurri et al., 2004a). The definition finds its basis on Kurtenbach and Hulteen's (1990) definition of gesture that states that gesture is "a movement of the body that contains information". Especially in performing arts, gesture is not only intended to denote things or to support speech as in the traditional framework of natural gesture, but the information it contains and conveys is often related to the affective, emotional domain. From this point of view, gesture can be considered "expressive" since it carries what Cowie and colleagues (2001) call "implicit messages", and what Hashimoto (1997) calls KANSEI. That is, expressive gesture is the responsible of the communication of information that we call expressive content. Expressive content is different and in most cases independent from, even if often superimposed to, possible denotative meaning. Expressive content concerns aspects related to feelings, moods, affect, intensity of emotional experience. For example, the same action can be performed in several ways, by stressing different qualities of movement: it is possible to recognize a person from the way he/she walks, but it is also possible to get information about the emotional state of a person by looking at his/her gait, e.g., if he/she is angry, sad, happy. In the case of gait analysis, we can therefore distinguish among several objectives and layers of analysis: a first one aiming at describing the physical features of the movement, for example in order to classify it, a second one aiming at extracting the expressive content gait coveys, e.g., in terms of information about the emotional state that the walker communicates through his/her way of walking. From this point of view, walking can be considered as an expressive gesture: even if no denotative meaning is associated with it, it still communicates information about the emotional state of the walker, i.e., it conveys a specific expressive content. In fact, in this perspective the walking action fully satisfies the conditions stated in the definition of gesture by Kurtenbach and Hulteen (1990): walking is "a movement of the body that contains information". Some studies can be found aiming at analyzing the expressive intentions conveyed through everyday actions: for example, Pollick (2004) investigated the expressive content of actions like knocking or drinking.

Studies on expressive gesture are grounded on several different sources coming from both science and technology and art and humanities. Such sources include humanistic theories of non-verbal communication developed for dance and choreography such as Rudolf Laban's Theory of Effort (Laban, 1947, 1963), theories from music and composition (e.g., Schaeffer, 1977), theories from psychology (e.g., Argyle, 1980; Wallbott, 1980; Pollick, 2004).

5.3.2 Experiments on expressive gesture in human full-body movement

One of the most relevant aspects of research in expressive gesture that can be highly relevant for the design and development of Enactive Interfaces fusing action and vision is the study of expressive gesture in human full-body movement.

The EU-IST Project MEGA (Camurri et al., 2004c), for example, carried out an analysis of expressive gesture in human movement with respect to two main aspects: analysis of expressive full-body movement in dance performance and analysis of expressive movement of music performers. Experiments aimed at (i) investigating specific and concrete aspects of expressive gesture in human movement, and (ii) testing and validating models and algorithms by comparing their performances with spectators' ratings of the same stimulus material. Experiments focused on the following aspects:

- *Classification of dance performances in term of basic emotions*. This experiment employed a reference archive of dances in which five dancers performed the same dance with four different emotional expressions: anger, fear, grief and joy. Psychologists (Department of Psychology, Uppsala University, Sweden) collected spectators' ratings from 32 observers. Engineers (InfoMus Lab, DIST, University of Genova, Italy) extracted motion cues from the video recordings and developed models for automatic classification of dance gestures in term of the conveyed basic emotion. Statistical analysis was carried out in order to find correlations between perceived emotions and extracted cues. This analysis allowed identifying some cues that are particularly relevant for classification in the considered dances (Camurri at al., 2003a). Machine learning techniques (e.g., decision trees) were then employed to classify dance fragment in term of basic emotions. Results from spectators' rating and automatic classification were compared (Camurri at al., 2004a).

- *Real-time prediction of emotion in dance*. Work on cue extraction was used for a performance at The Cultural Center of Stockholm, where a dancer controlled the mood of the music by changing dancing style. Three different cues extracted from the video were used to predict the emotional intention of the performer using a fuzzy logic approach. The prediction was then mapped to the expression of the musical output. A number of short dance fragments in the hip-hop/street dancing style, expressing different intentions, were recorded. The aim of the experiment was to verify how well cue extraction and emotion prediction could be performed in a real scenario and to evaluate the effectiveness of interaction strategies associating motion cues to musical output.
- *Body movements of a marimba player*. The aim of this experiment was the analysis and assessment of the communication of expressive intentions through a marimba player's movements (using the recorded archive of a marimba player's movements), including influence of played emotion and visible body parts. The experiment allowed a classification of rated emotions in terms of movement cues, such as small - large, slow - fast, uneven - even, and jerky - smooth movements, that characterized each of the performed intentions (Dahl and Friberg, 2004).
- *Body movements of a pianist and spectator's engagement*. This experiment aimed at measuring spectator's engagement in artistic communication and at correlating such measures to expressive movement (and audio) cues. Piano performances were shown to spectators. Spectators were asked to evaluate their emotional engagement while listening to the performances. Continuous measures from spectators were compared with the extracted motion and audio cues in order to find possible correlations (Camurri at al., 2004a).

5.3.3 Mapping of expressive content on visual output

Another relevant aspect for fusing Action and Vision concerns mapping of the expressive content extracted from expressive gesture (e.g., using the models and algorithms employed in the above experiments) onto real-time generation and processing of visual content (Camurri et al., 2004c, 2004d).

For example, in the EU-IST MEGA project this task was performed through the use of interaction (or mapping) strategies able to directly dealing with expressive content. These includes:

- *Expressive direct strategies* consisting of associations without any dynamics of expressive cues of analysed expressive gestures with parameters of synthesised expressive gestures (e.g., generation and/or processing of visual output). Expressive direct strategies are usually employed for implementing reactive behaviour. Several implementations are available for these strategies. One of them consists of collections of pre-defined condition-action rules, i.e., set of rules associating given configurations of parameters coming from the analysis side with given configurations of synthesis parameters. Another one employs collections of algebraic functions, computing values of synthesis parameters depending on values of analysed expressive cues. It should be noticed that while the complexity of an algebraic function can be freely increased according to any possible need, it anyway remains a static function, i.e., the mapping it induces does not change anymore once the function is defined and put at work.
- *Expressive high-level indirect strategies* including reasoning and decision-making processes and often related to rational and cognitive processes. For example, consider a software module able to make decisions based on the incoming decoded expressive content: it could select an algebraic function (an expressive direct strategy) within a collection of possible algebraic functions, thus allowing direct strategies to be adapted to the current context, i.e., implementing an adaptive and dynamic direct mapping. Indirect strategies are usually characterized by a state evolving over time (that is, they are dynamic processes) and by decisional processes. Production systems and decision-making algorithms can be employed to implement this kind of strategies.

Mapping strategies were experimented in real scenarios, namely artistic performances and museum exhibits. The following pictures show some of the results:

The first picture (above) refers to "Shells", a piece for Tarogato (a traditional Hungarian single-reed wind instrument) and interactive music system that was first written in 1993 by Robert Rowe in collaboration with Esther Lamneck. The piece was performed in Prato, Italy in July 2003 and the EyesWeb open platform (www.eyesweb.org, Camurri et al., 2000, 2004e) was employed for audio analysis and for generation of visual content in real-time. The visual content consisted of expressive manipulation (e.g., in term of colors, deformation, application of filtering techniques) of both pre-recorded material stored in avi files and live images of the player and of a dancer (Douglas Dunn) performing dance improvisation during the piece.

The second picture (below) shows an example employed in some MEGA performances: an image is deformed depending on expressive audio and motion cues. A lens metaphor is used. Different kinds of deforming lenses are available in EyesWeb. The amount of deformation is related to values of the extracted expressive cues. The dancer's body, the background or both of them can be deformed. For example, in the left image the background is deformed. The dancer's body is instead deformed in the image on the right. The background bitmap can be selected by the artist and can be dynamically changed during the performance.

5.3.4 Research on expressive gesture: potentialities and possible problems

Research on expressive gestures can open a path toward novel forms of artistic performances, in which technology is not just something added to a traditional scenario, but rather becomes a component of the artistic language. They can reveal novel perspectives leading to a redefinition of the relationship between art and technology from a condition in which art uses technology for accomplishing specific tasks that only technology can afford (or that computers can do better than humans) to a novel condition in which technology and art share the same expressive language and in which technology allows the artist to directly intervene on the artistic content and in the expressive communication process. Technology can thus increase the aesthetic value of an interactive artwork by dramatically improving the sense of presence of the active audience.

Expressive gestures are also a challenge for designers of interactive systems: as in software engineering methods for designing and implementing good software are developed and studied, the designer of interactive systems would need methods to develop and adapt his/her work with respect to the application scenarios and the requirements of the designer of a performance or of an installation.

Art is not the only application field that can benefit of research on expressive gesture. For example, another domain of interest is therapy and rehabilitation in which some pilot experiments were carried out. In the framework of the EU-IST project CARE HERE, for example, prototypes of multimodal interactive systems were developed to analyze body movements of different kinds of patients (Parkinson's patients, severely handicapped children, people with disabilities in the learning processes) and to map the analyzed parameters onto automatic real-time generation of audio and visual outputs, attempting to create aesthetic resonance (Camurri et al., 2003b).

Research on expressive gesture has of course its limitations. At the current state of the art, we cannot say that we are able to extract and interpret high-level information in every condition and context. What has been done up to now was a collection of preliminary experiments aiming at shading some light on the communication mechanisms involving expressive gesture. That is, we are moving toward the interpretation of high-level information, but this is still far to be reached. Moreover, interpreting expressive content requires the ability of capturing and measuring the subtlest nuances of human gestures since small variations within such nuances are often likely to make the difference. This is a big challenge for sensors systems for example in terms of needed temporal and spatial resolution.

Computer vision and sound analysis algorithms also are still not enough developed to handle such subtleties.

Other problems relate to the use of the expressive information. This holds in particular for artistic use in the performing arts scenario. Is really this kind of information useful for artists? How such information can be employed in a real performance (in a non-trivial way)? These questions still remain largely unanswered and at the current state of the art it seems quite difficult to get a sound and comprehensive answer in short time.

Finally, ethical concerns have to be taken into account. As an example, consider the risk related to the availability of techniques able to emotionally classify users according to their behavior and to convey them suitable emotional messages. Such techniques could allow third parties to control in some way user's behavior (e.g., as it is already happening on a certain extent in advertising, companies could use such information to control the behavior of their customers). Moreover, the emotional, affective sphere is related to the most private aspects of individuals' life and techniques able to deal with it must be carefully considered with respect to privacy safeguard.

5.3.5 References

Argyle M. (1980), "Bodily Communication", Methuen & Co Ltd, London.

Camurri A., Mazzarino B., Ricchetti M., Timmers R., Volpe G. (2004a), "Multimodal analysis of expressive gesture in music and dance performances", in A. Camurri, G. Volpe (Eds.), Gesture-based Communication in Human-Computer Interaction, LNAI 2915, Springer Verlag.

Camurri A., Mazzarino B., Volpe G. (2004b), "Expressive interfaces", Cognition, Technology & Work, 6(1): 15-22, Springer-Verlag.

Camurri A., Mazzarino B., Menocci S., Rocca E., Vallone I., Volpe G. (2004c), "Expressive gesture and multimodal interactive systems", in Proc. AISB 2004 Convention: Motion, Emotion and Cognition, Leeds, UK.

Camurri A., De Poli G., Friberg A., Leman M., Volpe G. (2004d), "The MEGA project: analysis and synthesis of multisensory expressive gesture in performing art application", submitted.

Camurri A., Coletta P., Massari A., Mazzarino B., Peri M., Ricchetti M., Ricci A., Volpe G. (2004e) "Toward real-time multimodal processing: EyesWeb 4.0", in Proc. AISB 2004 Convention: Motion, Emotion and Cognition, Leeds.

Camurri A., Lagerlöf I., Volpe G. (2003a) "Recognizing Emotion from Dance Movement: Comparison of Spectator Recognition and Automated Techniques", International Journal of Human-Computer Studies, 59(1-2): 213-225, Elsevier Science.

Camurri A., Mazzarino B., Volpe G., Morasso P., Priano F., Re C. (2003b) "Application of multimedia techniques in the physical rehabilitation of Parkinson's patients", Journal of Visualization and Computer Animation, 14(5): 269-278, Wiley.

Camurri A., Hashimoto S., Ricchetti M., Trocca R., Suzuki K., Volpe G. (2000) "EyesWeb – Toward Gesture and Affect Recognition in Interactive Dance and Music Systems" Computer Music Journal, 24(1): 57-69, MIT Press.

Cowie R., Douglas-Cowie E., Tsapatsoulis N., Votsis G., Kollias S., Fellenz W., Taylor J. (2001), Emotion Recognition in Human-Computer Interaction. IEEE Signal Processing Magazine, 1.

Dahl S., Friberg A. (2004) "Expressiveness of musician's body movements in performances on marimba" in A. Camurri, G. Volpe (Eds.), Gesture-based Communication in Human-Computer Interaction, LNAI 2915, Springer Verlag.

Hashimoto S. (1997), "KANSEI as the Third Target of Information Processing and Related Topics in Japan", in Camurri A. (Ed.) "Proceedings of the International Workshop on KANSEI: The technology of emotion", AIMI (Italian Computer Music Association) and DIST-University of Genova, 101-104.

Kurtenbach G., Hulteen E. (1990), "Gestures in Human Computer Communication", in Brenda Laurel (Ed.) The Art and Science of Interface Design, Addison-Wesley, 309-317.

Picard R. (1997), "Affective Computing", Cambridge, MA, MIT Press

Pollick F.E. (2004), "The Features People Use to Recognize Human Movement Style", in A. Camurri, G. Volpe (Eds.), Gesture-based Communication in Human-Computer Interaction, LNAI 2915, Springer Verlag.
Laban R. (1963), "Modern Educational Dance", Macdonald & Evans Ltd., London.
Rowe R. (2001), "Machine Musicianship", Cambridge MA: MIT Press.
Rowe R. (1993), "Interactive music systems: Machine listening and composition", Cambridge MA: MIT Press.
Schaeffer P. (1977), "Traité des Objets Musicaux", 2nd Edition, Paris, Editions du Seuil.
Wallbott H.G. (1980), "The measurement of Human Expressions", in Walbunga von Rallfer-Engel (Ed.) Aspects of communications, 203-228.

5.4 Input-output gesture processing

©Annie Luciani (INPG)
September 2004

Input-ouput gesture processing is related to the computer processing of gestures that are bilateral. The case on which gestural action is decoupled to gestural perception (as if we act with one hand and perceive with the other hand- will not be examine here. Thus, the paragraph is composed of two parts :
- Link between the input (gestural action) and the output (gestural returns)
- Link between the gestural input-output loop and the visual output.

5.4.1 Link between the input (gestural action) and the output (gestural returns)

As defined previously, the instrumental interaction contains necessarily ergotic interaction during with there is an exchange of energy between human and manipulated objects. The devices able to convey such coupling between sensors and actuators are necessarily retroactive mechanical transducers as defined in [EI_WP6_DLV1_UPMF_040923]. As said before, such technological situation is new in the context of computerized environments. It was the last to be implemented and its needs for this simultaneously force feedback devices and physically-based models able to produce the forces as answers to the sensed actions.

The process that have to link gestural inputs to gestural outputs is then of a nature deeply different that signal processing (filtering, extraction, reconstruction etc.) used in pure semiotic action as in motion and gesture capture, gesture recognition or expressive gesture presented just before. It is necessarily a computer simulation of a physically-based model, in its large meaning of model that correlates in a energetically consistent way, extensive variables (ex. positions) and intensive variables (ex. forces).

5.4.2 Link between the gestural input-output loop and the visual output.

In such ergotic situation, seing is the way to perceive the object' behaviors that do not exist without such manipulation (deformations, fractures, etc.). The energy communicated by the human to the manipulated object, and retracting on the human body, is the energy that produces such behaviors. What it is seen is the trace, the encoding of the bilateral gestural action-perception loop in a visual effect. We can state that the visual effect is not an effect arbitrarily or metaphorically linked with such actions, but a type of exteroceptive perception of ergotic gestural interaction. Consequently, the relation between action and correlated vision have to be taken in charge (1) by a physically-based model and (2) which have to be energetically link with the model that link the gestural input and output in a consistent way.

Such system is not totally implemented in usual VR platform. [Luciani 1991] [Florens 1998] [Uhl 1995] developed such complete processes in a generic way, that are able to link bilateral ergotic interaction and epistemic seing of the dynamics of the objects in a generic and physically consistent way.

5.4.3 References

[Florens 1998] FLORENS J.J., CADOZ C., LUCIANI A., " A real-Time Workstation for physical model of Multisensorial Gesturally controlled Instrument". Proceedings of ICMC 1998.

[Uhl 1995] UHL(C), FLORENS JL, LUCIANI(A), CADOZ (C) - «Hardware Architecture of a Real Time Simulator for he Cordis-Anima System :Physical Models, Images, Gestures and Sounds» - Proc. of Computer Graphics International '95 - Leeds (UK), 25-30 June 1995 - , Academic Press. - RA Ernshaw & JA Vince Ed. - pp 421-436

[Luciani 1991] LUCIANI (A), JIMENEZ (S), FLORENS (JL), CADOZ (C) & RAOULT (O), "Computational physics : a modeler simulator for animated physical objects", Proceedings of the European Computer Graphics Conference and Exhibition. Vienna, Austria, Septembre 91, Editeur Elsevier

6 The HCI point of view : Multimodality in Human-Computer Interaction

©Joan De Boeck, LUC-EDM.

In this chapter the current State-Of-The-Art in Human-Computer interaction within multimodal applications will be illuminated. In a first section some theoretical foundations and general terminology will be handled. The next section covers the problem of task-analysis and identification. Section three elaborates on the different interaction-metaphors currently available. As haptic interaction is central in this WP, we will clarify the different haptic definitions and elaborate on currently available haptic applications as well as available API's that can be used to develop haptic applications. As two-handed interaction is becoming more and more important in multimodal interfaces, this will be handled in section six. Finally, the last section will illustrate problems with haptic interaction and the solutions currently addressed in literature to overcome.

6.1 Theoretical Foundations of Multimodality

When discussing multimodal computer interfaces, it is necessary to have a common understanding of what is multimodality. When reading literature it becomes clear that several definitions do exist. In most psychological oriented research [Stoffregen and Bardy, 2001]
a modality is defined as a sensual observation. [Arens and Hovy, 1990] define a modality as a 'single mechanism by which to express information'. Bernsen at its turn also takes the representation into account to define a modality which comes in contrast with sensory modalities [Bernsen, 1994a]. Other important terms such as 'Channel', 'Medium' and 'Exhibit' also are defined in [Arens and Hovy, 1990] as well as in the deliverable of WP1 of the Miami project (Esprit Project 8579) [Schomaker et al., 1995].
Several attempts to develop taxonomies of input and output modalities can be found in literature; however none of them appears to be fully complete. In [Coomans and Timmermans, 1997] a first version of an input-output taxonomy has been presented. One of the more elaborated and theoretically founded taxonomies are described in [Bernsen, 1994a] and [Bernsen, 1995b] for output modalities. In essence multimodal output is seen as a combination of several uni-modal outputs, where a modality is defined as a vector of five orthogonal properties: A modality both can be linguistic or non-linguistic, analogue or non-analogue, arbitrary or non-arbitrary, static or dynamic and communicates via a certain medium such as the visual, haptic or auditory channel. Based on this definition, all output modalities can be classified in a hierarchical structure. In [Bernsen, 1995a] the extension of the taxonomy to input modalities is proposed. This results in a similar structure in which in general the haptic channel has been replaced by a kinestatic channel. In his subsequent work [Bernsen, 1994b] [Bernsen and Verjans, 1998], Bernsen et al. use this taxonomy to support multimodal interface design and make transitions from the task domain to a human computer interface and make a well-considered choices between the available possibilities.
On the other hand, P.B. Andersen adopts the concept of 'semiotics' [Andersen, 2000]
to bring the insights from older media to the task of interface design.

6.2 Tasks in HCI

6.2.1 General

As generally accepted these days, a computer interface must allow a user to execute a given task as efficient and intuitive as possible. Several rules to define good interfaces can be found in literature. As early as 1986, Norman & Draper suggest a user-centred design in order to fulfil the requirements of usability. Based on the work of [lewis&Rieman1993], [Marti, 1996] suggests a task-centred design of user interfaces. Indeed, by identifying and modelling the tasks of the user, the system can be designed and evaluated in a structured process. Several formal notations such as transition diagrams [Parnas, 1969] or ConcurTaskTrees [paterno], can be used. Although each notation has its benefits, none of them offers an ideal solution or a complete notation of the knowledge needed for the development of a user interface. Therefore, a flow of models, with model transformations, each of them adding extra data is proposed in [doct Kris]. We can distinguish the domain model, user model, task model, dialog model and the presentation model.

6.2.2 Virtual Environments

Which tasks a user can fulfil heavily relies on the application and the environment the computer-system is used. An essential difference exists between office applications, HCI in production or safety applications and HCI in design and evaluation-tools or even in virtual reality. In the latter situation, and by extension also in all other modelling and visualisation-applications, according to [Esposito] the user's task can be classified into five groups: Navigation, object query, object manipulation, object creation and modification, and application environment query and modification. However [Gabbard and Hix, 1997] reduces this set to three common tasks: Navigation and locomotion, object selection and object manipulation, modification and query. Those three groups of interaction require their own metaphors as will be handled in the next section.

6.3 Metaphors in HCI

6.3.1 General Metaphors

A common way to transfer the knowledge a user has picked up in one domain or situation to another situation, or to transfer intuitive every-day acts to a computer system is by use of metaphors. Undoubtedly the most-well known metaphor is the "desktop metaphor", which projects the knowledge of the every-day office-desktop to the computer. The same is true in other applications for other tasks that are to be executed by the user.

6.3.2 Navigation Metaphors

When navigating in a virtual world, or even when navigating in any other desktop application four questions arise: "Where am I now?", "What is my current attitude and orientation?", "Where do I want to go?", "How do I travel there". Answers to those questions, which imply the more psychological questions about navigation and way finding are answered in [Satalich, 1995]. To accomplish the task of navigation, several metaphors have been developed already. Although no metaphor exists that fits in every application, each metaphor has its specific benefits. When navigating in a 3D environment, the problem often is to find an intuitive paradigm to use a 2D input device for a 6DOF task. The best solutions however are found using 6DOF input devices. In [Ware and Osborne, 1990], three basic metaphors are presented, as there are: Flying Vehicle, Scene in Hand and Eyeball in hand. Other metaphors, for special applications, or using special hardware can be found in literature: [Tan et al., 2001] describes a metaphor that increases the camera's elevation when increasing the speed. [Koller et al., 1996] shows a method that allows orbital viewing by rotating the head in an immersive environment. [Camera-In-Hand] and [Ext Camera in Hand] test a metaphor that uses the PHANToM Device and makes usage of force feedback in order to navigate through a 3D world. Finally, WIM (world in miniature) [Mine, 1995] provides the user with a miniature representation of the world that allows him to interact on a general overview of the world.

Another category of navigation metaphors can be found with body based techniques: tracking the user's body or torso can activate a flying vehicle metaphor: by leaning or by extending ones arm, a virtual vehicle can start its movements. In this category, we also can classify all kinds of treadmills [eurohaptics04 invited].

6.3.3 Selection Metaphors

Selection of objects in 3D environment can be completed in several ways. The most intuitive metaphor is by direct manipulation, by directly touching the object with the virtual hand or the pointer, the object becomes selected [?], as we know from our everyday life. The problem with this technique is that selection is limited to the objects that are within the proximity of the user, (or better) within the bounds of the used device. To solve this problem, ray casting (or cone casting) is proposed by [Mine, 1995].

6.3.4 Manipulation Metaphors

Objects that are selected can be manipulated or queried. In this task, we also consider interaction with widgets (menus, dialogs, …). In general both selection metaphors can be used for manipulation. Although direct manipulation is limited to the user's close environment, while ray casting go without some degrees of freedom such as rotating around another axis than the axis of the ray itself [Bow-man and Hodges, 1997]. A complete taxonomy of the most frequent manipulation techniques can be found in [Pouprey et al., 1998]. In this document, a distinction has been made between exocentric and egocentric metaphors. The former defines metaphors in which the user is outside the world and is looking at it from a kind of a god-eye's view. The latter class defines metaphors in which the user is standing in the world itself: virtual pointer metaphors, such as ray casting [Bolt, 1980], flashlight [Halliday and Green, 1996], aperture [ref 22 in poupirev] or image plane [ref 25 in poupirev], or virtual hand metaphors such as classical 'virtual hand', 'gogo' [Poupyrev et al., 1996] or 'indirect gogo'.

For all metaphors (navigation, selection and manipulation) an adequate feedback is essential for acceptance by the user. Be this feedback visual, aural or haptic in nature.

6.4 Haptic Interaction

As mentioned in the previous paragraph, feedback is of crucial importance for adequate interaction between human and machine. From the human point-of-view, feedback can go over the five human senses: visual, aural, haptic sense of smell and taste, although only the first three senses are relevant in current HCI. Currently, far more research has been conducted in graphic feedback. However, concerning multimodality, haptic feedback has been gaining more and more importance for the last decade. Several researches indeed prove the benefits of force feedback (e.g. [Unger et al.,2002]), as will become clear from the remainder of this section.

6.4.1 Definitions in Haptic Interaction

According to [Oakley et al., 2000], let us define haptic as everything relating to the sense of touch. From the every-day life, it is clear that the sense of touch consists of several perceptions. We can distinguish force feedback as the result of a mechanical production of force sensed by the human kinaesthetic system. Tactile feedback on the other hand is pertaining to the cutaneous sense, specifically the sensation of pressure. Proprioceptive feedback is related to sensory information about the state and position of the body. The vestabular sense refers to the head position and acceleration [Oakley et al., 2000].

Other authors make a distinction between active and passive force feedback: according to [Lindeman et al.,1999] passive force feedback is caused by the shape of an object held or kept by the user. Active feedback is caused by a specialized actuator that actively generates forces or vibrations. [Srinivasan and Basdogan, 1997] also makes a difference between point-based and ray-based haptic interactions. In the former, only the end point of the haptic device interacts with the object. The latter the generic probe of the haptic device is modelled as a finite ray. In this work also the resolution of the human tactile sensory capabilities can be found.

6.4.2 Haptic Applications

Over the last decade an increasing number of research and commercial haptic application became present. A summarized overview of current haptic applications, going from telepresence, over training up to virtual reality can be found in [Stone, 2000].

Some work presents the haptic interaction as an improvement in a standard 2D desktop environment [Miller and Zelevnik, 1998] as the haptic feedback in those standard WIMP-interfaces mostly is restricted by the feel of the mouse of buttons. The real strength however of the haptic channel will show up in more complex interfaces. Full haptic interfaces especially appear to be efficient compared to their WIMP counterpart when dealing with 3D information [Gauldie et al., 2004]. A commercially available product that fully exploits the possibilities of the haptic feedback is FreeForm [Sensable Inc.].

Besides the already existing applications, several ready-to-use API's are available; either commercially or in the OpenSource commulinty; e.g. the Ghost API, and recently OpenHaptics Toolkit, both from Sensable [Sensable Inc.]. Other API's are developed by ReachIn [Reachin] and by Novint Technologies [Novint]

6.4.3 Two-handed interaction

Recent research proves that bi-manual interaction improves the interaction as this is a natural means of interacting by humans. In general, in a cooperative bimanual action, the non-dominant hand creates a frame of reference for the dominant hand [Hinkley et al., 1997b]. As the task becomes more difficult, the importance of the specialization between the right and the left hand increases. Even in a common bimanual task such as writing, the user's dominant hand moves relative to the position of the non-dominant hand (e.g. which brings the paper and holds it in place). In [Balakrishnan and Hinckley, 1999], the author shows how the match and mismatch between the input of the hand and the graphical output influences two-handed performance. A more theoretical consideration of this asymmetrical bimanual behaviour can be found in Guiard's kinematic chain model [Guiard,1997].

Hinkley also proves that kinestatic information of a bimanual operation provides sufficient perceptual cues to form a frame of reference, which is independent of visual feedback [Hinkley et al., 1997a]. Another technique is to introduce a grounding object to support the user's hand. The combination of a grounding object, combined with both hands provides even better possibilities. Hinckley also found that uni-manual operations are more dependent on visual feedback.

The bimanual haptic approach also can be found in other work such as the work of Ishii. Here, tangible bits [Ishii and Ullmer, 1997] are used as physical handles. Those tangible user interfaces (TUI) allow the user to intuitively work with physical objects representing the computational data.

Lindeman et al. [Lindeman et al., 1999][Lindeman et al., 2001] use passive feedback-devices such as a physical plate, held by the non-dominant hand, are used for widget interaction. The user's proprioceptive knowledge of his/her non-dominant in respect to the dominant hand also appears to be a valid approach to easily activate several functions within the world. Also [Mine and Brooks, 1997] showed that hand-held widgets, which rely on proprioceptive information, are to be preferred over floating or object-bound widgets. Asuming that that hybrid 2D/3D hatic widgets (2D menus and dialogs with haptic feedback positioned in 3D space) are a suitable way to interact with in a 3D interface [Raymaekers and Coninx, 2001], in [De Boeck et al., 2004] a proprioceptive gesture with the non dominant hand in respect to the dominant hand has been used in order to activate those widgets. In the subsequent work the same principle has been used to grab objects to bring and keep them closer to the user for easy manipulation.

6.4.4 Problems and solutions in Haptic HCI

Although haptic feedback in general allows the user to be less dependent on visual feedback, still some problems do exist when bringing haptics and vision together. Several solutions to overcome those problems have been proposed so far. Most of those solutions propose certain improvements of the visual feedback, although other solutions suggest an enrichment of the entire multimodal interface by adding other communication channels.

6.4.4.1 Improving visual Feedback

One of the solutions to result in a better depth-perception can be found in [Giess et al., 2000]. Here the author investigates if the use of shadows can improve the haptic interaction by providing the user with a better depth awareness. The same depth-perception problem has been addressed in [Bouguila et al., 2000] by introducing stereopsis. However, here we can consider several issues of the coupling between haptics and stereopsis, as there are: possible misperceptions of the stereopsis cue, difference between perception of objects in stereoscopic images and objects as used in real life and finally the fact that real objects have much stronger visual cues (such as focusing). Another clue that make users perceive depth is motion parallax. By introducing head tracking, the experience also can be improved [pet]. Other work describes the use of large projection screens with or without stereopsis in order to improve the sense of presence in the generated world. Large projection screens seem to have similar results as head-mounted displays (HMD's) but causing less simulator sickness [Patrick et al.,2000]. This result has been adopted in [De Boeck et al., 2003a] in order to bring a semi-immersive haptic setup to the desktop. Other applications or experiments in which haptic feedback in combination with visual feedback by large projection screens can be found in [Pape and Sandin, 2000] and [Burdea, 1996].

Reachin [Reachin] use another approach by placing a monitor up-side-down in such that the user sees the screen through a mirror. The haptic device is placed underneath the mirror to establish a collocation of the haptics and the vision. The Reachin display can be used either in stereo- or in monoscopic view.

6.4.4.2 *Enrich the interaction by other modalities*

Another approach to improve the haptic-vision-fusion tries to adopt other communication channels in order to enrich the interaction. As [Cohen, 1994] says: we use the strengths of one modality to overcome the weaknesses of another. Although this track has not been investigated thoroughly in combination with haptic interfaces, we believe a well-suited combination of haptic feedback and other modalities can improve the fusion between haptics and vision. Indeed, Oviat states that well-designed multimodal systems must aspire a synergetic blend between the strengths of each modality, although each random combination of modalities does not lead to a synergetic blend [Oviatt, 1999]. In this work, she also discusses several common myths about multimodal interaction. Bolt already showed in 1980 [Bolt, 1980] the benefits of gestures and speech as a natural way to communicate.

Currently the most obvious modality to improve an interface appears to be speech-input. Although speech-interfaces becomes increasingly more popular, not much work can be found that blends haptic interaction together with speech. [De Boeck et al., 2003b] shows a preliminary test that shows user's behaviour when both speech and haptic direct interaction are available. Users tend to 'over-use' the speech modality. However, [Sturm et al., 2002] shows that prolonged use of this kind of interfaces leads to a more realistic blend. One of the weaknesses of the spoken modality is the relative load on the short-time memory and the error-prone nature of current technical implementations.

In [Sturm et al., 2002] the author defines different types of multimodality. If each modality is used one after the other (e.g. first click and then speak) we speak of "sequentially multimodal". This is the most simple way of multimodality. When two modalities are used simultaneously this is "simultaneous multimodality". In the latter case, if both modalities handle a single piece of information it is called "coordinated simultaneous multimodality". The opposite obviously is described as non-coordinated simultaneous modality. It is clear that simultaneous multimodality implies several technical complications. Some solutions for the fusion of different types of data can be found in [Nigay and Coutaz,1995].

6.5 References

Andersen, P. B. (2000). What semiotics can and cannot do for HCI. In CHI 2000 Workshop on Semiotic Approaches to User Interface Design, Den Haag, NL.

Arens, Y. and Hovy, E. (1990). How to describewhat? towards a theory of modality utilization. In Proc. of the 12 Conference of the Cognitive Science Society., pages 18{26.

Balakrishnan, R. and Hinckley, K. (1999). The role of kinesthetic reference frames in two-handed input performance. In Proceedings of the 12th Annual

ACM Symposium on User Interface Software and Technology, pages 171{178, Asheville, USA.

Bernsen, N. O. (1994a). Foundations of multimodal representations: a taxonomy of representational modalities. In Interacting With Computers, volume 6 Number 4.

Bernsen, N. O. (1994b). Modality theory in support of multimodal interface design. In Proceedings of the AAAI spring symposium on intelligent multi-media multi-modal systems, pages 37{44.

Bernsen, N. O. (1995a). A taxonomy of input modalities. http://www.mrc-cbu.cam.ac.uk/amodeus/abstracts/tm/tm wp22.html.

Bernsen, N. O. (1995b). A toolbox of output modalities: Representing output information in multimodal interfaces. In CCI Working Papers in Cognitive Science and HCI, volume WPCS-95-10, Centre for Cognitive Science, Roskilde University.

Bernsen, N. O. and Verjans, S. (1998). From task domain to human-computer interface: exploring an information mapping methodology. In John Lee (Ed): Intelligence and Multimodality in Multimedia Interfaces, Menlo Park, CA.

Bolt, R. A. (1980). Put-that-there: Voice and gesture at the graphics interface. In Proceedings of Siggraph80, volume 14, pages 262{270.

Bouguila, L., Ishii, M., and Sato, M. (2000). E®ect of coupling haptics and stereopsis on depth perception in virtual environments. In Proceedings of Workshop on Haptic Human-Computer Interaction, pages 54{62, Glasgow, UK.

Bowman, D. A. and Hodges, L. F. (1997). An evaluation of techniques for grabbing and manipulating remote objects in immersive virtualenvironments. In Proceedings of the Symposium on Interactive 3D Graphics, pages 35{38, Providence, RI, USA.

Burdea, G. C. (1996). Force And Touch Feedback For Virtual Reality. Winley Inter-Science.

Cohen, P. R. (1994). The role of natural language in a multimodal interface. In Proceedings of the Fifth ACM Symposium on User Interface Software and Technology, pages 143{149, Monteray, CA, USA.

Coomans, M. and Timmermans, H. (1997). Towards a taxonomy of virtual reality user interfaces. In Proceedings of the international Conference on Information Visualisation, volume 1, London, UK.

De Boeck, J., Raymaekers, C., and Coninx, K. (2003a). Aspects of haptic feedback in a multi-modal interface for object modelling. Virtual Reality, 6(4):257{270.

De Boeck, J., Raymaekers, C., and Coninx, K. (2003b). Blending speech and touch together to facilitate modelling interactions. In Proceedings of HCI International 2003, volume 2, pages 621{625, Crete, GR.

De Boeck, J., Raymaekers, C., and Coninx, K.
(2004). Improving haptic interaction in a virtual environment by exploiting proprioception. In Proceedings of Virtual Reality Design and Evaluation Workshop, Nottingham, UK.

Gabbard, J. and Hix, D. (1997). A taxonomy of usability characteristics in virtual environments.

Gauldie, D., Wright, M., and Shillito, A. M. (2004). 3D modelling is not for wimps part II: Stylus/mouse clicks. In Proceedings of Eurohaptics 2004, pages 182{189, Munich, Germany.

Giess, C., Topfer, S., and Meinzer, H.-P. (2000). Can shadows improve haptic interaction in virtual environments. In Proceedings of the 2nd PHANToM Users Reserach Symposium 2000, volume 8 of Selected Readings in Vision and Graphics, pages 49{54, Zurich, CH.

Guiard, Y. (1997). Asymetric division of labor in human skilled bimanual action: The kinematic chain as a model. In Journal of Motor Behaviour, volume 19, pages 486{517.

Halliday, S. and Green, M. (1996). A geometric
modeling and animation system for virtual reality. In Communications of the ACM, volume 39 nr 5, pages 46{53.

Hinkley, K., Pausch, R., and Pro±tt, D. (1997a). Attention and visual feedback: The bimanual frame of reference. In Siggraph 1997: Proceedings of the 24th Annual Conference on Computer Graphics, Los Angeles, CA, USA.

Hinkley, K., Pausch, R., Pro±tt, D., Patten, J., and neal Kassell (1997b). Cooperative bimanual action. In Proceedings of CHI97: ACM Conference on Human Factors in Computer Systems, Atlanta, Georgia, USA. Ishii, H. and Ullmer, B. (1997). Tangible bits: Towards seamless interfaces between people, bits and atoms. In Proceedings of CHI 1997, pages 234{241, Atlanta, GA, USA.

Koller, D., Mine, M., and Hudson, S. (1996). Head-tracked orbital viewing: An interaction technique for immersive virtual environments. In Proceedings of the ACM Symposium on User Interface Software and Technology (UIST) 1996, Seattle, Washington, USA.

Lindeman, R. W., Sibert, J. L., and Hahn, J. K. (1999). Towards usable VR: An empirical study of user interfaces for immersive virtual environments. In Proceedings of the SIGCHI'99, pages 64{71. Lindeman, R. W., Sibert, J. L., and Templeman, J. N. (2001). The e®ects of 3D widget representation and simulated surface constraints on interaction in virtual environments. In IEEE Virtual Reality, pages 141{148.

Marti, P. (1996). Hci in italy: Task-centred design – turning task modelling into design. In SIGCHI Bulletin, volume 28 number 3.

Miller, T. and Zelevnik, R. (1998). An insidious haptic invasion: Adding force to the X desktop. In Proceedings of the 11th annual ACM symposium on User interface software and technology, pages 59{66, San Francisco, CA, USA. ACM.

Mine, M. R. (1995). Isaac: A virtual environment tool for the interactive construction of virtual worlds. Technical Report TR95-020, UNC Chapel Hill Computer Science, ftp://ftp.cs.unc.edu/pub/technical-reports/95-020.ps.Z.

Mine, M. R. and Brooks, F. P. (1997). Moving objects in space: Exploiting proprioception in virtual environment interaction. In Proceedings of the SIGGRAPH 1997 annual conference on Computer graphics, Los Angeles, CA, USA.

Nigay, L. and Coutaz, J. (1995). A generic platform for addressing the multimodal challenge. In Proceedings of ACM CHI'95 Conference on Human factors in Computing Systems, Denver, Colorado, USA.

Oakley, I., McGee, M. R., Brewster, S., and Gray, P. (2000). Putting the feel in 'look and feel'. In Proceedings of CHI 2000, pages 415{422, The Hague, NL.

Oviatt, S. (1999). Ten myths of multimodal interaction. In Communications of the ACM, volume 42, pages 74{81.

Pape, D. and Sandin, D. (2000). Quality evaluation of projection-based VR displays. In IPT2000: Immersive Projection Technology Workshop, Ames, IA, USA.

Parnas, D. (1969). On the use of transition diagrams in the design of a user interface for an interactive computer system. In Proceedings of the 24th national conference, pages 379{385.

Patrick, E., Cosgrove, D., Slavkovic, A., Rode, J. A., Verratti, T., and Chiselko, G. (2000). Using a large projection screen as an alternative to head-mounted displays for virtual environments. In Proceedings of CHI 2000, pages 478{485, The Hague, NL.

Pouprey, I., Weghorst, S., Billunghurst, M., and Ichikawa, T. (1998). Egocentric object manipulation in virtual environmnets; empirical evalutaion of interaction techniques. Computer Graphics Forum, 17(3):41{30.

Poupyrev, I., Billinghurst, M., Weghorst, S., and Ichikawa, T. (1996). The go-go interaction technique: non-linear mapping for direct manipulation in vr. In Proceedings of the ACM Symposium on User Interface Software and Technology (UIST) 1996, Seattle, Washington, USA.

Raymaekers, C. and Coninx, K. (2001). Menu interactions in a desktop haptic environment. In Proceedings of Eurohaptics 2001, pages 49{53, Birmingham, UK.

Reachin (2004). Reachin display. http://www.reachin.se/.

Satalich, G. (1995). Navigation and way¯nding in virtual reality: Finding the proper tools and cues to enhance navigational awareness. Master's thesis,

University of Washington, Seattle, USA. Schomaker, L., Nijtmans, J., Camurri, A., Lavagetto, F., Morasso, P., and Benoit, C. (1995). A taxonomy of multimodal interaction in the human information processing system. In A Report of the Esprit Project 8579 MIAMI, WP1.

Sensable Inc. Freeform. http://www.sensable.com/products/.

Srinivasan, M. and Basdogan, C. (1997). Haptics in virtual environments: Taxonomy, research status, and challenges. In Computer and Graphics, volume 12 number 4, pages 393{404.

Sto®regen, T. and Bardy, B. (2001). On speci¯cation and the senses. In Behavioral and Brain Sciences, volume 24, pages 195{265, USA.

Stone, R. J. (2000). Haptic feedback: A potted history, from telepresence to virtual reality. In Proceedings of the Workshop on Haptic Human-Computer Interaction, pages 1{8, Glasgow, UK.

Sturm, J., Bakx, I., Cranen, B., Terken, J., and Wang, F. (2002). The e®ect of prolonged use on multimodal interaction. In Proceedings of ISCA Workshop on Multimodal Interaction in Mobile Environments, Kloster Irsee, Germany.

Tan, D., Robertson, G., and Czerwinski, M. (2001). Exporing 3d navigation: Combining speed-coupled °ying with orbiting. In Proceedings of CHI 2001, Seatle, Washington, USA.

Unger, B., Nicolaidis, A., Berkelman, P., and Thompson, A. (2002). Virtual Peg-In-Hole performance using a 6-DOF mangnetic levitation

haptic device: Comparison with real forces and with

visual guidance alone. In 10th Interantional Symposium on Haptic Interfaces For Virtual Environments and Teleopreaton Systems, Orlando, Florida, USA.

Ware, C. and Osborne, S. (1990). Exploration and virtual camera control in virtual three dimentional environments. In Computer Graphics, volume 24 Number 2.

7 Perceptual and cognitive issues in Action-Vision fusion

7.1 Touch and Seing

7.1.1 Haptic modality. Some definitions and problems of classification

© Elena Pasquinelli, Institut Jean Nicod

Touch well instantiates the difficulty of providing unambiguous definitions of sensory modalities. Different terms are correlated to the notion of touch, some of them being considered as inclusive of the others. Different classifications are proposed.

Neurophysiology makes use of the term “somatic sennsory system” [] Kandel & Shwartz, 2000] comprehensive of 2 main components: a system for the detection of mechanic stimuli (light touch, vibration, pressure) and a system for the detection of pain stimuli and temperature (Purves, Augustine, Fitzpatrick, Katz, La Mambia & McNamara, 1997). This classification is based on the **physical energy** of the stimuli the captors are sensitive to. Mechanoreceptors are then sub-divided into tactile or cutaneous captors which are distributed at the surface (skin) of the body and proprioceptive captors which are located within the muscles, tendons and joints of the body (**localization** of the captors). Different perceptual qualities are then associated to the two sub-systems: in a general fashion tactile captors are described as implicated in the perception of the qualities of the objects of the external world (such as dimensions, shape, microstructure, movement relative to the skin) and the proprioceptive system as dedicated to the (more or less aware) perception of the position and movement of the body. Neurophysiology deals then with the ascription to the somesthetic system of 4 main **functions**: discriminative touch, proprioception, nociception, temperature perception. There is a difficulty in sharply separating the external and the internal mechano-captors and associating them separately with exteroceptive and proprioceptive functions respectively. Active exploration of the world’s objects implies the utilization of internal, proprioceptive mechano-captors, but it provides information about the properties of the external world. Active touch has then been considered as a separate category of touch on the basis of the role that movements (and movement captors) play into the discrimination of the properties of objects. This category is labeled “tactile-kinaesthetic perception” or “haptic perception”.

The term “haptics” was first introduced by Revesz - see [Revesz, 1958] - to incorporate cutaneous and kinaesthetic information). [Loomis & Lederman, 1986] make reference to the haptic sensory modality in terms of “kinaesthetic touch”: kinaesthetic touch is comprehensive of cutaneous and kinaesthetic receptors, provides information about objects and surfaces that are in contact with the subject and guides the manipulation of objects. The modality of touch is then composed of three sub-modalities: “*The modality of touch encompasses distinct cutaneous, kinesthetic and haptic systems that are distinguished on the basis of the underlying neural inputs. The cutaneous receptors are embedded in the skin; the kinesthetic receptors lie in muscles, tendons, and joints; and the haptic system uses combined inputs from both.*” [Lederman, Klatsky, 2002, p. 1] These classifications do not then question the divisions operated by neurophysiology and based on the energy of the stimulus and the localization of the receptors.

On the contrary, Katz’s *The world of touch*, a classic in the history of the study of touch [Katz, 1989], refused to adopt an “*atomistic approach to perception*” by individuating and separating the activity of different sensory captors (thus multiplying the number of tactile sensations) and he choose to adopt a system of classification based on the **qualities** perceived by touch. The world of touch possesses three main modifications or qualities: surface touch (the two-dimensional tactile structure that is identified when touching a continuous palpable area, localized at the surface of the object, and following the curvatures of the object), immersion touch (the tactile phenomenon without definite shape nor structure or spatial orienting, as when moving the hand in a fluid), volume touch (the perception of the shape, the spatial distribution of the object that we can have when the object is, for instance, covered by a textile or the hand is covered by a glove). The “*skin senses*” cannot then be separate since “*in the living* *organism (whose expressions, after all, are what we wish to understand), large coalitions of sensory elements always work together*.” [Katz, 1989, p. 34] The differentiation operated by the physiology of the senses is then an artifact, in that complex phenomena constitute the only real component of conscience. The physiology of the senses is then obliged to recombine the elements it created into complex phenomena, thus suggesting that complex phenomena are cognitive products of logical operations. On the opposite, Katz invites to consider tactile perception as an immediately complex phenomenon which does not require the intervention of successive cognitive operations. Katz’s suggestion does not solve the problem of differentiating touch from other sensory modalities, but is only limited to the internal classification of touch, since common qualities (as the shape of an object) can be appreciated by more than one sensory modality (as vision and touch).

A sort of middle-way position is represented by Gibson’s classification of haptic touch. In fact, Gibson maintains the distinction between physical energies and types of receptors but points more on the object properties. [Gibson, 1962, 1966] suggested that there is a great difference in the resulting percept depending on the **active or passive** role of the perceiver: when the stimulation is passive, as when being touched by an object, even if the object is moving, the subject obtains sensations of skin modification; it is only when the subject plays an active role by actively touching the object that attention is directed to the properties of the object. Active touch is then defined as an exploratory rather then a merely receptive sense, by which the variations in the skin stimulation are produced by variations in the motor activity. Thus the unitary perception of an object with more fingers doesn’t require a central integrations since the pressure of the fingers upon an object informs about the qualities (e.g. the hardness) of the object and does not give rise to separate, cutaneous sensations (on the contrary, in the case of passive touch, two separate pressures on the skin give rise to two different sensations). In the same way, in active touch, kinaesthesia is not to be separated nor simply combined with cutaneous sensations, since the patterns of change of the skin contact covary with the change in limb position giving rise to one and the same information about the object properties. Touch is exemplary of the connection of perception and movement in perception, since in its case the equipment for feeling is anatomically the same as the equipment for doing. The non-separation of the skin senses from kinaesthesia is labeled “haptic system”, and distinguished from haptic touch and dynamic touch [Gibson, 1962]: "*The sensibility of the individual to the world adjacent to his body by the use of his body will here be called the haptic system. The word haptic comes from a Greek term meaning "able to lay hold of." It operates when a man or an animal feels things with his body or its extremities. It is not just the sense of skin pressure. It is not even the sense of pressure plus the sense of kinesthesis. […] The haptic system, then, is an apparatus by which the individual gets information about both the environment and his body. He feels an object relative to his body and the body relative to an object.*" [Gibson, 1966, p. 97] The haptic system is then sub-divided into: cutaneous touch (when the skin and deep tissues are stimulated without movement of muscles and joints); haptic touch (when the skin and deep tissues are stimulated by the movement at the joints, as in catching an object, palpating, squeezing, etc. in order to extract information about its geometry and microstructure); dynamic touch (when skin and joints are stimulated in association with muscular effort, as in the discrimination of weight, which is better when the object is wielded, rigidity, viscosity, etc.); oriented touch (the combination of inputs from vestibular, joint and skin receptors); touch-temperature (the combination of skin stimuli with vasodilatation and vasoconstriction); painful touch; social touch (the affective components of touch, as in the new-born cares).

Dynamic touch is a rich domain of studies (see for instance [Turvey, 1996]. Dynamic touch is active, but it does not regard finger exploration, for instance. The perception of object properties by wielding is a prominent example of dynamic touch. The haptic properties that are thus perceived are those regarding the macro-geometry and volume of the objects, as the extension, shape, orientation and weight; in the same time properties of the limb holding the object are discriminated. [Turvey, 1996] states as follows: “*What sets kinesthetic touch apart from other forms of touch is the prominent contribution of muscular effort and its sensory consequences. As a grasped object is wielded, the receptors that interpenetrate muscular and tendinous tissues are mechanically stimulated. These mechanoreceptors, as they are called, respond to the stretching, twisting, and bending of muscles and* *tendons. Their collective response to the changing flux of mechanical energy is the primary (although not the exclusive) neural basis of dynamic touch*."

An interesting suggestion for the internal classification of touch can be extracted from the researches of Lederman and Klatsky (see for instance [Lederman, Klatsky, 1987; Klatsky, Lederman, Metzger, 1985]. The hand system is an intelligent instrument in that it makes use of its motor capacities for ameliorating its sensitive abilities. Since the movements are coupled with the properties of the objects that are extracted, it is possible to describe a set of exploratory movements or patterns that correspond to object properties as texture (slight movements on the surface), shape (contour following or wielding), presence of parts, etc. It is then possible to sub-divide the sense of touch (of active touch) with no reference to the energies, to the type of receptors or their localization, but only to observable properties of the exploratory activity such as the movement employed and the perceptual result obtained.

Recently another use of the term "haptics" has appeared in the domain of computer interfaces. Computer haptics includes the technologies and processes for the generation and proposition of force-feedback stimuli to human users in virtual reality environments. The focus in on hand exploration and manipulation: "*Haptics is concerned with information acquisition and object manipulation through touch. Haptics is used as an umbrella term covering all aspects of manual exploration and manipulation by humans and machines, as well as interactions between the two, performed in real, virtual or teleoperated environments. Haptic interfaces allow users to touch, feel and manipulate objects simulated by virtual environments (Ves) and teleoperator systems."* [Biggs, Srinivasan, 2001, p. 1]

References
Biggs, S. J., Srinivasan, M. A. (),. (2001). Haptic Interfaces.
Gibson, J. J. (1962). Observations on active touch. *Psychological Review, 69*(6).
Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin Company.
Heller, M. A., Schiff, W. (1991). *The Psychology of Touch*. Hillsdale, NJ: L. Erlbaum Associates Publishers.
Kandel, E. R., Schwartz, J. H., Hessel, T. M. (2000). *Principles of neural science*.: McGraw-Hill.
Katz, D. (1989). *The World of Touch*. Hillsdale: L. Erlbaum Associates Publishers.
Klatzky, R. L., Lederman, S. J., Metzger, V. A. (1985). Identifying objects by touch: An "expert system". *Perception & Psychophysics, 37*(4), 299-302.
Klatzky, R. L., Lederman, S. J. (2002). Touch. In A. F. H. R. W. Proctor (Ed.), *Experimental Psychology* (Vol. 4, pp. 147-176). New York: Wiley.
Lederman, S. J., Klatzky, R. L. (1987). Hand Movements: A Window into Haptic Object Recognition. *Cognitive Psychology, 19*(3), 342-368.
Loomis, J. M., Lederman, S. J. (1986). Tactual perception. In K. Boff, Kaufman, L., Thomas, J. (Ed.), *Handbook of perception and human performance*. New York: Wiley.
Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., La Mambia, A. S., McNamara, J. O. (1997). *Neurosciences*.: Sinauer Associates.
Revesz, G. (1958). *The human hand, a psychological study*. London: Routledge and Kegan Paul.
Turvey, M. T. (1996). Dynamic touch. *American Psychologist, 51*(11), 1134-1152.

7.1.2 Touch. The sense of reality.

© Elena Pasquinelli, Institut Jean Nicod

Touch modality has often been described as the "sense of reality".
In 1754, Condillac [Condillac, 1984] attributed to touch the property of distality in perception, that is to the external world. A pretended statue, which senses are opened one after the other, would not be able to distinguish between itself and the objects it is perceiving until the touch modality isn't activated; when the statue begins to tactualy explore the reality, two types of sensation arise: those regarding the

object and those regarding the body, and this allows the separation of the self from the world, and the perceptual constitution of the distal object.

In his seminal study of touch of 1925, Katz [Katz, 1989] was re-editing the tradition of the "objectifying" capacity of touch modality by describing touch as the "*sense of reality*". Katz insists on the fact that the tactual sense is bipolar: a stimulus on the dorsal part of the hand can be perceived both as a subjective, proximal, local sensation or as the sensation of the object which causes the experience (this same reflection has been made in the philosophical domain by Merleau-Ponty: touch is a reciprocal sensory modality in that it is impossible to touch without being touched; the activity of touching implies then the involvement of the body in the knowledge about the world). The objective pole dominates when touch is accompanied by movement, then in active tactual exploration. When, for instance, one hand touches the other, the static hand is perceived as touched (subjective pole of the sensation), while the hand which is moving is perceived as touching (objective pole). It is then movement, associated with touch, that produces the impression of the reality as external. Touch can be considered the "sense of reality" in that its connection with movement is particularly strong.

The objectifying role of movement is also recognized in the use of visuo-tactile substitution displays [Bach-y-Rita, 1982]. It seems in fact that the possibility of actively guiding the sensors (the camera) produces a shift from the sensation of a local, tactile stimulation to the (visual) perception of a distal object placed in the external reality.

Active movement, more than touch itself then, would constitute the proper "sense of reality".

It should be tested if active movement of the user is suitable for improving the credibility of virtual objects (otherwise called the "sense of presence"), that is for projecting them in the external space as objective realities. Some questions can be asked: Is it true that touch is more strongly connected with movement than the other senses, and that active touch contributes to the process of objectification of the stimulus more than, say, active vision? What is the role of the expectations of the user relatively to the consequences of its movements (the behavior of the object in response to his movement)?

References

Condillac, E. B. d. (1984). Traité des sensations [1754], *Traité des sensations. Traité des Animaux*. Paris: Fayard.

Katz, D. (1989). *The World of Touch*. Hillsdale: L. Erlbaum Associates Publishers.

Merleau-Ponty, M. (1943). *Phénoménologie de la perception*. Paris: Gallimard.

Bach-y-Rita, P. (1982). Sensory substitution in rehabilitation. In M. S. L. Illis, & H. Granville (Ed.), *Rehabilitation of the Neurological Patient* (pp. 361-383). Oxford, UK: Blackwell Scientific.

7.1.3 Epistemic Seeing

© Nivedita Gangopadhyay, Institut Jean Nicod

Epistemic seeing is seeing an entity in a manner that involves integrating the visual data with thought and belief and in general is the application of a concept to a currently experienced object.

F. I. Dretske in his book *Seeing and Knowing* (1969) draws a phenomenologically salient distinction between epistemic seeing and non-epistemic seeing regarding the content of perceptual experience. Epistemic seeing can be considered as the perceptual apprehension of a fact, namely, the fact that the thing before the perceiver is an object bearing a particular name and having certain functions e.g. perceiving a table *as* a table i.e. as an object bearing the name "table" and associated with certain functions. Thus epistemic seeing involves not only the grasp of the relevant concept, but also implies its application in experience. Epistemic seeing is claimed to amount to knowledge. The perceiver may *know* that an object *b* is *P* by *seeing that* it is *P*. For a visual experience to qualify as epistemic seeing certain conditions must be fulfilled. According to Dretske, a perceiver *S sees that* an object *b* is *P* only if (i) *b* is *P*, (ii) *S* simply sees that *b*, (iii) the conditions under which *S* simply sees *b* are such that it would not look to *S* as it does unless it were *P*, (iv) *S* believes that the conditions in (iii) obtain, and (v) *S* believes that *b* is *P*. When conditions (i)-(iv) are satisfied, *S* has a *conclusive reason* for believing *b* to be *P*. Hence, if condition (v) is also satisfied, *S knows that b* is *P* by *seeing that b* is *P*. Thus *b*'s looking the way it does to *S* provides *S* (in the relevant circumstances) a conclusive reason for believing that *b* is *P*. Thus, to *see that b* is *P* is to believe that it is *P because* of the way it looks and to see it in this way is epistemic seeing of the object.

References
Dretske, F.I. 1969. *Seeing and Knowing*, Chicago: The University of Chicago Press.
Dretske, F.I. 1979. "Simple Seeing," *Body, Mind, and Method*, D.F. Gustafson and B.L.Tapscott, eds., Dordrecht: Kluwer Academic publishers: 1-15. (Reprinted in F. Dretske, *Perception, Knowledge and Belief: Selected Essays*, Cambridge: Cambridge University Press, 2000: 97-112.)

7.2 Sensorimotor theories of Perception

© Nivedita Gangopadhyay, Institut Jean Nicod

The sensorimotor theories of perception maintain that perception is a skill-based activity of environmental exploration mediated by the animal's implicit knowledge and mastery of the laws governing the sensorimotor contingencies.

According to this approach seeing is a way of acting. Sensorimotor theories do not endeavor to explain vision as a process in the brain involving internal representations or as something that happens *in* individuals but rather as something that they *do*. The experience of seeing occurs when the organism masters the governing laws of sensorimotor contingencies. Sensorimotor contingencies are the structure of the rules governing the sensory changes produced by various motor actions and thus they can be said to embody the laws of interaction between the organism and the environment. The perceiver's knowledge of the sensorimotor contingencies takes the form of a practical know-how. Sensorimotor contingencies are different for different sensory modalities and thus they serve to distinguish one sensory modality from another. In case of perception, visual sensorimotor contingencies can be classified into two categories viz.- a) sensorimotor contingencies that are determined by the character of the visual apparatus itself and b) sensorimotor contingencies that are related to visual attributes. The main proponents of the doctrine are J. Kevin O'Regan and A. Noë who following Mac Kay call the laws describing the sensorimotor interactions the "sensorimotor contingencies". However, scholars like Ryle, Pessoa, Maturana, Varela, Thompson, Rosch, Järvilehto and Gibson have expressed similar ideas and in general have stressed the importance of action in perception.

References:
Mac Kay, D.M. (1962). Theoretical models of space perception. *Aspects of the theory of artificial intelligence*, ed. Muses, C.A. Plenum Press
Myin, E. and O'Regan, J. Kevin (2002). Perceptual consciousness, access to modality and skill theories. *Journal of Consciousness Studies* 9(1): 27-45
Noë, A. (2002). Experience and the active mind. *Synthese* 29: 41-60
Noë, A. (2002). On what we see. *Pacific Philosophical Quarterly* 83: 1
Noë, A. and O'Regan, J. Kevin. (2000). Perception, attention and the grand illusion. *Psyche* 6(15) http://psyche.cs.monash.edu.au/v6/psyche-6-15-noe.html
Noë, A. and O'Regan, J. Kevin. (2002). On the Brain-Basis of Visual Consciousness: A Sensorimotor Account. *Vision and Mind*, ed. Noë, A. and Thompson, E. MIT Press
Noë, A., Pessoa, L., Thompson, E. (2000). Beyond the grand illusion: what change blindness really teaches us about vision. *Visual Cognition* 7: 93-106
O'Regan, J. Kevin. and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences* 24(5) : 939-1011
http://nivea.psycho.univ-paris5.fr/OREGAN-NOE-BBS/ORegan;Noe.BBS.pdf
more at: http://nivea.psycho.univ-paris5.fr/, http://ist-socrates.berkeley.edu/~noe/
O'Regan, J. Kevin. and Noë, A. (2002). What it is like to see : A sensorimotor theory of perceptual experience. *Synthese* 29 : 79-103
O'Regan, J. Kevin., Myin, E. and Noë, A. (2004 ??). Skill, corporality and alerting capacity in an account of sensory consciousness, (published ??) http://nivea.psycho.univ-paris5.fr

7.3 Co-location of visually (stereoscopically) and haptically presented and perceived virtual objects

©Janson Gunnar, UPPSALA

In the physical world a surface may emit, transmit or reflect light, send out sound waves or make resistance to impact, but its location is the same whatever sense is considered. In virtual environments this may not be the case. Visual and haptic displays often present the same virtual object in different locations, the visual version on a screen, with or without stereo information, and the haptic version within the working area of a haptic display. This may not be considered to be very important, as perception is highly adaptable and, at least after some training, may function well also in new conditions. On the other hand, it can be expected that virtual conditions similar to the real conditions may be advantageous, as the senses in such cases work under conditions natural for them. It should be noted, however, that co-location of visual and haptic information is only one of the conditions important for perceived location. Perceived location of a surface is dependent on several other information sources.

There are efforts to increase the similarity between real and virtual worlds by rendering visual and haptic information that is calibrated to provide co-ordinate systems that are coincident. For the ReachIn device (http://www.reachin.se), for instance, co-location in this way is a basic idea. However, the expected advantage of co-location for perception and performance has been tested only in a few experimental studies.

Wall, Paynter, Shillito, Wright and Scali (2002) investigated the effects of co-location and stereoscopic information on *performance* in a targeting task. The results were significant effects on accuracy for both factors, as well as a significant interaction such that, when stereo information is available, the benefit from haptic information is significantly less. Concerning time to reach target there was an advantage to have stereoscopic information, but haptic information had no effect. There were large individual differences in both dependent variables. As the participants in this experiment were beginners in using a haptic device, Wall et al. recommend caution in generalizing the results to other levels of expertise.

Jansson & Öström (2004) studied the effects of co-locating visual and haptic information, as well as of stereoscopic information, on the precision in the *perception of object form*. The experimental problem was the following. Is there any difference in precision of judging the distortion of a spherical object, when the information is presented in the form of exploratory motion paths, if (1) the visual information is presented stereoscopically or non-stereoscopically and (2) the visual and the haptic information is co-located or not? In other words, is there any benefit of presenting the visual information stereoscopically and co-locating visual and haptic information? The result was that Co-location had a significant effect on the depth dimension under Stereo conditions. This demonstrates that Co-location has positive effects on the perception of object form in depth.

The studies mentioned were performed within relatively small environments (with a Phantom display in a Reachin co-locating device). Bouguila, Ishii and Sato (2001) used a larger SCALABLE SPIDER device. They had noted the instability of depth perception in stereoscopic presentations of virtual environments and studied the contribution of haptics to the *precision of locating virtual objects* with random-dot stereoscopic information. They found that the precision of the location of objects in depth, as well as the time of performance, was improved when haptic information was added to stereoscopic information.

A still larger virtual environment was studied at University College London (PURE-FORM, 2004). In a CAVE-like environment (ReaCToR) co-location of visual stereo information and haptic information via the Exoskeleton PURE-FORM display was arranged. Informal observations indicated that co-location was very important to enhance the experience (p. 26), but it was also found that some participants had problems to visualize a 3D stereo object and focused on the front wall instead of the 3D position of the stereo model. This led to a discrepancy between visual and haptic information and loss of the stereo effect. A planned solution was to place the haptic interaction as close as possible to the projection wall (p. 27).

A detailed analysis of the problems to get stereoscopic information to function well for depth perception in virtual environments was made by Wann, Rushton & Mon-Williams (1995), discussing.

These problems are common for all systems intending to present large spatial intervals from a dual 2D source. In contrast with natural conditions, accommodation and vergence eye movements are not coordinated when viewing these virtual displays, and the blur information available in natural contexts are missing in computer generated optical information, the conditions being more similar to 2D pictorial information. This problem increases when the range of stereoscopic depths is wider (p. 2731). Wann et al. also stated that the dissociation of accommodation and convergence are not easily solved by increased display quality and concluded that several problems remain to accurately simulate 3D space from 2D images information (p. 2735). These problems motivated the author of a recent study of distance perception in virtual environments (Messing, 2004) to refrain from using stereoscopic information, with additional reference to the lack of difference between monocular and binocular information at distances larger than two meters (Philbeck & Loomis, 1997).

In sum, it has been shown that co-location between visual (stereoscopic) and haptic information has positive effects in small environments, but that several problems remain with applications in larger environments. It seems to be recommendable, especially in virtual environments with large depth intervals, to consider in each case if the visual information should be stereoscopic or not.

References:

Bouguila, L., Ishii, M. & Sato, M. (2001). What impact does the haptic-stereo integration have on depth perception in stereographic virtual environment? A preliminary study. In Brewster, S., Murray-Smith, R. (Eds.), *Haptic Human-Computer Interaction. Lecture Notes in Computer Science,* Vol. 2058 (pp. 135-150). Berlin: Springer.

Jansson, G. & Öström, M. (2004). The effects of co-location of visual and haptic space on judgements of form. In M. Buss & M Fritschi (Eds.), *Proceedings of the 4th International Conference Eurohaptics 2004* (pp. 516-519). München, Germany: Technische Universität München.

Messing, R. (2004). Distance perception and cues to distance in virtual reality. Poster at First Symposium on Applied Perception in Graphics and Visualization, co-located with ACM SIGGRAPH, August 7-8, 2004, Loa Angeles, CA. (available at http://www.sccs.swathmore.edu/users/04/rossm/academic/thesis.html).

Philbeck, J. W. & Loomis, J. M. (1997). Comparison of two indicators of perceived egocentric distance under full-cue and reduced-cue conditions. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 72-85.

Pure-Form (2004). *The MPF system.* Deliverable 12. IST-2000-29580 PURE-FORM.

Wall, S. A., Paytner, K., Shillito, A. M., Wright, M. & Scali, S. (2002). The effect of haptic feedback and stereo graphics in a 3D target acquisition task. In Wall, S. A., Riedel, B., Crossan, A. & McGee, M. R. (Eds.): *Eurohaptics 2002. Conference Proceedings* (pp. 23-29). Edinburgh, U.K.: University of Edinburgh.

Wann, J. P., Rushton, S. & Mon-Williams, M. (1995). Natural problems for stereoscopic depth perception in virtual environments. *Vision Research, 35,* 2731-2736.

7.4 Perceptual conflict in visuo-haptic integration. the case of the pseudo-haptic feedback and its implications for haptic interfaces

©Fabien PFAENDER, Gunnar DECLERK, COSTECH

7.4.1 Intermodal integration

Our purpose here is not to make an exhaustive state of the art on sensory intermodal integration but to give a short survey of that problem in order to better characterize a particular case of visuo-haptic integration that has been recently studied and called *pseudo-haptic* phenomenon by the main searchers of the domain [LÉCUYER, 2000, 2003, 2004, CRISON, 2004, PALJIC, 2004 and others]. For a good state of the art of the problem of sensory integration, perceptual conflict and illusions, please see [PASQUINELLI, WP7, 2004].

Two main situations involving an intermodal integration are usually considered in psychology [HATWELL, 94] : First, a situation of *complementarity* in which sensorial modalities work in a complementary way by accessing to distincts properties of objects. Visual modality can for example appreciate the shape of an object, haptic modality (the hand) can perceive its weight and auditive modality can perceive the sounds it produces. Then the process of intermodal integration consists in putting in relation those different informations in order to build a single « multimodal » object. Secondly, a situation which can be characterized as situation of *redundancy* (or recurrence) : different sensory modalities can provide informations on a same property of the perceived object. For exemple the size or the texture of an object can be estimated by the visual and the haptic modality. In the same way the spatial location of an echoing object can be estimated visually or by the auditory modality, etc.

With regard to those two situations we can differentiate two kinds of properties of objects : properties that are *specific* of a particular sensory modality (for example the sound, that can only be perceived with the auditory modality) and properties that can be characterized as *amodal* [HATWELL, 94]. Acording to Hatwell, this idea has especially been introduced by Gestalt psychologists who argued that some perceptions were *amodal* because they didn't have a sensory substratum (the hidden part of a table for example) [HATWELL, 94]. This term was later used by Gibson and Bower with a more general meaning : an information is amodal because not specific to a modality.

Many cases of redundacy situations exist where the informations (also called *cues*) about the same property of an object coming from the different sensory modalites are discrepant, producing a *perceptual conflict*. Pasquinelli proposes to define the *perceptual conflict* « as the presence of two contradictory elements in one and the same perceptual unit, so that as the production of an incoherent perceptual grouping » [PASQUINELLI, WP7, 2004]. Those particular perceptive situations enable to study the way the conflicting sensory informations combine.
Different principles of integration have been identified and different models have been developped in order to explain how such partly redundant informations are integrated into an unitary final coherent percept. We won't make here an exhaustive study of those models. We just want to point three main principles of integration that are usually described : an « average » between conflictual informations ; a « total capture » with the complete recalibration in the dominant modality ; or finally a « partial capture » with a low contribution of the minor modality for the estimation of the conflictual property. The different principles of integration could depend on the situation that is considered and on the property of objects that is perceived. For example Collins and Singer (1968) showed that visual cues dominate haptic cues in the perception of the orientation of shapes. In this case the resulting percept is completely determined from the visual modality. Other principles could also hold for a large number of other situations. For example the two stage model of *weak fusion* (cf. JACOBS, 2002 for a summary) : in a first stage every single cue provides a separate estimate of a physical property and in a second stage these single estimates combine into a final estimate by weighted averaging. The weights for each estimate or cue depend on its reliability. The more reliable a cue is the larger its weight is in the linear combination.

Pasquinelli proposed to call the percepts resulting from integration of discrepant sensory informations *solved conflicts*. In that case « the experience can be suitably modified in order to give rise to a coherent final percept ». It's necessary to precise that there are other cases where conflictual situations don't give raise to those *solved conflicts* : « if two perceptual experiences are too discrepant for the system, and/or there are no good reasons to put them together in one only set, they get separated into two different final percepts » [PASQUINELLI, WP7, 2004].

The case we are here interested in concerns situations presenting conflicts between visual and haptic informations. What kind of principle is working in integration of discrepant visuo-haptic informations?
Numerous cases of visuo-haptic conflicts show a dominance of vision. For example, when visual and proprioceptive informations about the position of the arm are artificially discorrelated by the carrying of prismatic glasses there is often a visual capture or a un compromise with a visual dominance [HATWELL, 1994]. In a same way, in a experiment of I. Rock et J. Victor (1964) subjects are looking through a lens that makes them appear rectangular a square object. They have to choose a similar object among a collection of objects that they can touch but not see and a collection of objects that they can see but not touch. Most of the subjects choose an object corresponding to the shape seen and not to the shape palpated. This seems to show a dominance of vision upon kinesthesia by touching in that particular situation.
There are also experimental cases where the same weight is given to the tactile and visual information of texture of an objet [cf. Lederman et Abbott, 1981, in PASQUINELLI, WP7, 2004]. However that kind of solution seems flexible since it variates with the verbal instruction given to the subjects. It seems thus that « the cognitive knowledge implied by the verbal instruction is relevant for the aspect of the final percept » [PASQUINELLI, WP7, 2004].

7.4.2 A case of perceptual conflict in visuo-haptic integration : Lécuyer's "pseudo-haptic" feedback

We are now presenting a case of visuo-haptic solved conflict that Lécuyer, Coquillart, Kheddar, Richard and Coiffet studied and that they called the *pseudo-haptic feedback* (also called *pseudo-haptic illusion* or *pseudo-haptic phenomenon*) [LÉCUYER, 2000]. It's useful to precise that this phenomenon has been studied particularly regarding to the implications it could have upon technical developments of force-feedback devices and more globally haptic interfaces.

According to Lécuyer, Andriot and Crosnier (2003), the pseudo-haptic feedback can be understood as « the generation of haptic sensations by the use and combination of sensory feedbacks coming from other channels than the tactile channel ». Lécuyer et al. (2003) showed that it was possible to provide haptic informations to the user of an input device which passively reacts to the applied force (no force feedback) by combining its use with a visual feedback. The visual feedback is then used as a « disruptive » feedback which induces haptic sensations, « close to sensory illusion » [LÉCUYER, 2003].

The canonical experiment about pseudo-haptic feedback conducted by Lécuyer, Coquillart, Kheddar, Richard and Coiffet [LECUYER, 2000] is the following one : A *Spaceball* (isometric input device which measures the pressure whitout providing a force feedback ; it's presented in the form of a sphere) is used to measure the pressure that the user applies with a spring embedded in a piston (*real spring*). The spring displacement is visually displayed on a computer screen (*virtual spring*). The displacement of the virtual spring is a function of the pression applied on the *Spaceball* with the real spring.
For a same pressure applied on the *Spaceball*, the authors have then displaced more or less strongly the virtual spring on the screen. They observed that the more the visual displacement of the virtual spring is important, the more the real spring is perceived as soft – although the stiffness of the *Spaceball* is constant. Conversely, if the visual displacement is less important, the spring is perceived as stiffer. Lécuyer et al. conclude from this effect that « the visual deformation is so much misleading the user that the visual displacement on the screen partly acts as a substitute for the perception of the displacement of the finger pressing on the interface. This phenomenon appears to be an illusion of the proprioceptive sense which is blurred by the visual feedback. » [LECUYER, 2003].

In the same experimental way, Lécuyer, Burkhardt, and Etienne (2004), showed that it was possible to simulate friction occurring when inserting an object inside a narrow passage by artificially reducing the speed of the manipulated object during the insertion. "Assuming that the object is manipulated with an *isometric* input device, the user will have to increase his/her pressure on the device to make the object advance inside the passage" [LECUYER, 2004].

Paljic, Burkhardt and Coquillart (2004) proposed to consider the pseudo-haptic feedback as a form of *synesthesia* – also called *cross-modal transfer*. Synesthesia is described by Biocca, Kim and Choi (2001) as an extreme case of cross modal interaction "*in which sensory information of one sensory channel produces experiences in another unstimulated sensory channel that receives no apparent stimulation from the virtual environment*" [BIOCCA, 2001]. Cross modal interactions are fundamentally "*perceptual illusions in which users use sensory cues in one modality to fill in the missing components of perceptual experience*" [BIOCCA, 2001].

7.4.3 An alternative way for the conception of haptic interfaces ?

The works on the pseudo-haptic feedback, by indicating that it is possible to blur the proprioceptive sense of subjects (here the perception of the effort and « symmetrically » the resistance of an object) with an appropriate visual feedback, could open a kind of alternative way for the conception of haptic interfaces.

For example some results suggest the possibility of simulating textures in desktop applications by using such a pseudo-haptic feedback. Some experimental evaluations conducted by Lécuyer et al. [LECUYER, 2004] showed that it was possible to successfully identify macroscopic textures such as bumps and holes, by simply using the variations of the motion of the cursor. In this situation, "the coupling between the slowing down of the object on the screen and the increasing reaction force coming from the device gives the user the illusion of a force feedback as if a friction force was applied to her/him" [LECUYER, 2004].

Likewise the several levels of resistance – the inerty – of a manipulated virtual object or of the environment where this object is moving can be simulated by modifying the speed of the object and changing the « Control/Display ratio ». The Control/Display ratio can be defined as following : "The speed of hand movement (Control) to speed of cursor (or any other virtual object manipulated in the simulation) movement (Display) gives a ratio called the Control-to-Display (or C/D) ratio" [CRISON, 2004]. A strong resistance is then associated with a strong deceleration of the object on the screen.

In some situations pseudo-haptic feedback could in this manner be used as an alernative to real haptic feedback. Crison, Lécuyer, Savary, Mellet-d'Huart, Burkhardt, and Dautin (2004) tested and compared the use of haptic (with a PHANToM device) and pseudo-haptic feedback in a virtual reality system dedicated to the technical training of milling. The results of this evaluation showed that both devices were globally satisfying. However the haptic device had drawbacks – coming principally from the ergonomy of the PHANToM – that the pseudo-haptic solution could avoid. For example : "The grasping of the device was not satisfactory. The stylus was perceived as a too fine element. The rotations of the stylus disturbed the manipulation" [CRISON, 2004].

At present most of the haptic devices, used in telesurgery, for the handling of dangerous products, or more globally for all kinds of computter assisted teleoperations, present important drawbacks : they are often expansive, their utilisation present difficulties and the perceptive effects aren't always very convincing – which has repercussions on the « performances » of the user.

A « lighter » material environment that would provide comparable perceptive effects (mainly the information about resistance, rigidity, softness, texture of objects) would present great advantages. The ambition of the force feedback devices is not to provide the user information that would be exactly the same than in a « real » situation but rather to provide a « feeling of reality » and to produce reactions comparable to reactions occurring in a natural environment. The aim is also to understand which components are the basis for providing that « feeling of reality ». And we can imagine to go trough « short cuts » regarding to the natural functioning of the human perceptive system to reach comparable perceptive results.

However the mastering of the production of pseudo-haptic « illusions » needs to deeply understand the perceptive mecanisms at work. At present the pseudo-haptic effects still seem hard to canalize and Lécuyer et al. notice that subjects of their experiment are differently sensitive to the visuo-haptic illusion brought to the fore. « In front of that conflictual perceptive situation, some subjects behaved in a way "rather visual" and others reacted in a way "rather haptic" » [LECUYER, 2003].

We want finally to precise that the technical solution afforded by pseudo-haptic perceptive phenomenon doesn't seem to us to throw doubt on the relevance and interest of force feedback devices. But it offers a kind of alternative solution that could be used in particular cases – for example for a matter of low costs and if the use of such a perceptive phenomenon is sufficient to perform the expected tasks, for example in virtual reality systems for learning [CRISON, 2004].

7.4.4 References

[BIOCCA, 2001] F. Biocca, J. Kim, and Y. Choi (2001). Visual touch in virtual environments : An exploratory study of presence, multimodal interfaces and cross-modal sensory illusions. In *Presence*, volume 10, pages 247–265. MIT.

[COLLINS & SINGER, 1968] Collins, J.K. & Singer, G. (1968). Interaction between sensory spatial aftereffects and persistence of response following behavioral compensation. *Journal of Experimental Psychology, 77,* 301-307.

[CRISON, 2004] F. Crison, A. Lécuyer, A. Savary, D. Mellet-d'Huart, J.M. Burkhardt, and J.L. Dautin (2004). The Use of Haptic and Pseudo-Haptic Feedback for the Technical Training of Milling. Poster in EuroHaptics Conference (Eurohaptics), June 5-7, Munich, Germany.

[HATWELL, 1994] Hatwell, Y. (1994). Transferts intermodaux et intégration intermodale. In M. Richelle, J. Requin, & M. Robert (Eds.), *Traité de Psychologie Expérimentale, Volume 1* (pp. 543-584). Paris: Presses Universitaires de France.

[JACOBS, 2002] Jacobs, R.A. (2002). What determines visual cue reliability? *Trends in Cognitive Science, 6*(8)*,* 345-350.

[LÉCUYER, 2000] A. Lécuyer, S. Coquillart, A. Kheddar, P. Richard and P. Coiffet (2000). Pseudo-Haptic Feedback : Can Isometric Input Devices Simulate Force Feedback ? IEEE Int. Conf. on Virtual Reality, pages 83-90, New Brunswick, US.

[LÉCUYER, 2003] A. Lécuyer, C. Andriot, A. Crosnier (2003). Interfaces Haptiques et Pseudo-Haptiques. Proceedings of JNRR'03 (4ème Journées Nationales de la Recherche en Robotique) (*in french*)

[LÉCUYER, 2004] A. Lécuyer, J.M. Burkhardt, and L. Etienne (2004). Feeling Bumps and Holes without a Haptic Interface: the Perception of Pseudo-Haptic Textures. ACM Conference in Human Factors in Computing Systems (ACM SIGCHI), April 24-29, Vienna, Austria.

[PALJIC, 2004] A. Paljic, J. Burkhardt, S. Coquillart (2004). Evaluation of pseudo-Haptic feedback for simulating torque: a comparison between isometric and elastic input devices. Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2004. HAPTICS '04. Proceedings. 12th International Symposium on

[PASQUINELLI, WP7, 2004] E. Pasquinelli (2004). The awareness of perceptual illusions. Problems of coherence and dynamic knowledge. In documents of the WP7. *Article submitted to the journal Philosophical Psychology.*

[ROCK & VICTOR, 1964] Rock, I., & Victor, J. (1964). Vision and touch: an experimentally induced conflict between the two senses. *Science, 143*, 594-596

[VARELA, 1993] F. Varela, E. Thompson et E. Rosch (1993). L'inscription corporelle de l'esprit. Editions du Seuil.

7.5 Crossmodal attentional capture

© Nicolas J. Bullot, Institut Jean Nicod

Acording to recent findings, attentional capture seems to entail crossmodal upshots, which contribute to the continuous perceptual tracking of distal elements across distinct sensory fields.

It has long been described by phenomenological analyses that attention can undergo involuntary shifts (see Hatfield, 1998: 10, for an historical overview). This has led to the distinction between (i) automatic or reflex and (ii) voluntary attention within various lexical idioms. For instance, James (1890: 416-17) distinguishes between 'passive' and 'reflex' attention on one hand and 'active' and 'voluntary' attention on the other. At the end of the nineteenth century, Wundt (1897: 217-18) or Titchener (1899) mention also a related distinction. Experimental cognitive sciences brings also this distinction into play, but use the phrases 'exogenous attention' and 'endogenous attention' respectively (e.g., Driver & Spence, 1998; Jones, 2001; Spence, 2001).

The notion of exogenous capture refers to an apparent truncation of a voluntary search due to the presence of a distinctive attractor element which "pulls" attention to a specified location, event or object. Attractor elements may be for instance events with abrupt onsets. Capture relates thus to an involuntary access to (unexpected) events or objects that deserve attention due to salient features. According to recent empirical findings, exogenous capture is a crossmodal phenomenon, since attentional capture in one perceptual modality facilitates overt and covert access in other modalities access (to the properties of the attractor element – object, cue or event).

First, consider the case of publicly observable bodily movements, the overt case. Think of a distal event constituted by the fall from a table of a heavy object, such as a book or a glass, and its collision with a parquet floor. The collision will cause an abrupt impact sound that operates as an alerting signal for surrounding perceivers. Typically, the acoustic and auditory event will elicit (overt) saccadic eye movements and bodily orientation toward the source of the sound (Driver & Spence, 1998, 2004; Gibson, 1966: 75). Hence, this event has usually typical overt crossmodal consequences for each perceiver in its vicinity. In particular, the auditory event triggers an audiovisual perceptual tracking of the sound source which may be followed by visuo-haptic tracking of the same (e.g., if the perceiver searches, reaches and grasps the object).

Second, this description seems underpinned by experimental findings about the covert outcomes of attention capture. For instance, experimental works with the 'orthogonal cueing' paradigm studied by Driver & Spence (1998; 2004) and colleagues indicate that multimodal priming and crossmodal links in exogenous spatial attention seem to occur even before any overt bodily motion, in a covert manner. Spatially non-predictive cue in one modality can attract covert attention towards its location in other sensory fields, not solely within the cued modality (Driver & Spence, 1998: 255). For instance, abrupt sounds seem to 'attract' visual and tactile attention, not merely auditory attention. If this be true then it suggests that the capture of auditory attention in a given region of space facilitates the detection within the same particular region of space of events/objects by other sensory modalities. Hence, attentional capture seems to obtain (involuntary) cognitive spatial access to salient objects or cues within one single multi-modally tracked location. In addition, since the properties of each physical object are usually co-localized, this kind of crossmodal capacity is likely to facilitate the continuous tracking of the target object across distinct sensory fields.

It is appealing to conceive of this overt and covert crossmodal anchoring with the function of perceptual tracking of distal things. The crossmodal links should contribute to the keeping track of the uniqueness of the distal target by means of providing a continuous tracking across modalities.

References

Calvert, G. A., Brammer, M. J., & Iversen, S. D. (1998). Crossmodal identification. Trends in Cognitive Sciences, 2(7), 247-253.

Driver, J. (1996). Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. Nature, 381, 66-68.

Driver, J., & Spence, C. (1994). Spatial synergies between auditory and visual attention. In C. Ulmità & M. Moscovitch (Eds.), Attention and Performance XV: Conscious and Nonconscious Processing (pp. 311-331). Cambridge, MA: MIT Press.
Driver, J., & Spence, C. (1998). Attention and the crossmodal construction of space. Trends in Cognitive Sciences, 2(7), 254-262.
Driver, J., & Spence, C. (2004). Crossmodal spatial attention: Evidence from human performance. In C. Spence & J. Driver (Eds.), Crossmodal Space and Crossmodal Attention (pp. 180-220). Oxford: Oxford University Press.
Gibson, J. J. (1966). The Senses Considered as Perceptual Systems. London: George Allen and Unwin.
Hatfield, G. (1998). Attention in early scientific psychology. In R. D. Wright (Ed.), Visual attention (pp. 3-25). Oxford: Oxford University Press.
James, W. (1890). The Principles of Psychology. New York: Dover Publications.
Jones, M. R. (2001). Temporal expectancies, capture, and timing in auditory sequences. In C. L. Folk & B. S. Gibson (Eds.), Attraction, Distraction and Action: Multiples Perspectives on Attentional Capture (pp. 191-229). Amsterdam: Elsevier.
Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. R. Davies (Eds.), Varieties of Attention (pp. 29-62). Orlando: Academic Press.
Logan, G. D., & Compton, B. J. (1998). Attention and automaticity. In R. D. Wright (Ed.), Visual attention (pp. 108-131). Oxford: Oxford University Press.
Reisberg, D. (1978). Looking where you listen: visual cues and auditory attention. Acta Psychologica, 42(331-341).
Shiffrin, R. M. (1997). Attention, automatism, and consciousness. In J. D. Cohen & J. W. Schooler (Eds.), Scientific Approaches to Consciousness (pp. 49-64). Hillsdale, NJ: Erlbaum.
Spence, C. (2001). Crossmodal attentional capture: A controversy resolved? In C. L. Folk & B. S. Gibson (Eds.), Attraction, Distraction and Action: Multiples Perspectives on Attentional Capture (pp. 231-262). Amsterdam: Elsevier.
Titchener, E. B. (1899). An Outline of Psychology. New York: The Macmillan Company.
Wundt, W. M. (1897). Outlines of Psychology (C. H. Judd, Trans.). Leipzig: Wilhelm Engelmann.

7.6 Interceptive actions

©Le Runigo Cyrille, Benguigui Nicolas, Bardy Benoit, UPS XI

7.6.1 Introduction

The ecological approach to perception and action (Gibson, 1950, 1966, 1979) offers a theoretical support to study visuo-motor coordination in general, and sport skills in particular. He described very simples and parsimonious control mechanisms grounded on circular links between information and movement We will first recall some of the principles of the ecological approach which seem to give a better understanding of the control mechanisms underlying sport skills. We will then present some of the most representative studies of the interceptive actions which rest on this suitable framework.

7.6.2 Ecological Approach

According to Gibson (1950,1979), information is not obtained from a treatment of visual cues, but it is directly available in the environment and more precisely in the optical flow, i.e. in the changes of optical configuration, in the form of invariants. It is why Gibson called this approach the “direct perception“ theory. More precisely, this invariants specify affordances, that are the acts or behaviors permitted by objects, places, and events. “The affordances of the environment are what it *offers* animals, what it *provides* or *furnishes*, either for good or ill” (Gibson, 1979). An affordance is what the environment means to a perceiver, in terms of opportunities for action. To say that affordances are perceived means that information specifying these affordances is available in the stimulation of the optical flow and can be detected by a properly attuned perceptual system. Animals and environments join together to form systems. This joining is both characterized and permitted by compatibilities between properties of the animals and properties of the environment. Affordances belong to animal-environment systems and nothing less.

The detection of this information, allowing to specify the property of the actor-environment system, makes it possible for the actor to proceed with the regulation of its movement by modifying the produced internal forces. This produces in turn a change in the optical structure that reaches the eye. Movement control is thus possible step by step on the basis of a reciprocal coupling between information and movement. Warren (1988) proposed a formalization of the information-movement cycle, through the concept of laws of control, which connects an information and a force in order to ensure the success of the required action: _ Fint = G (_ Information) (an internal change in force is a function of a change in information). Gibson (1958) gave several convincing illustrations of this concept of invariance, as for example the "focus of expansion" (the center of the flow field), which corresponds to the direction of displacements. By using this invariant, the actor can change the current direction of his/her displacement in order to place the focus of expansion on the required direction. The use of invariants gives rise to a prospective type of control based on a circular link between information and movement.

7.6.3 Ecological Approach and Interceptive Actions

The interest of the ecological approach to perception and action is largely based on this concept of perception-action coupling and the study of interceptive actions is a paradigmatic example of this approach. Indeed, interceptive skills require great ability in the coordination of actions with the transformations of the environment. In such actions, accuracy can be measured in a temporal window of interception that can be less than 10 ms (e.g., Mc Leod, Mc Laughlin & Nimmo-Smith, 1986). This accuracy suggests a very close relationship between the perceptive and motor processes (e.g., Bootsma and Van Wieringen, 1990). The interceptive actions are directed towards a space-time goal. The action must be accurate in both time and space, in order to allow a participant to take along the effector to the right place at the right time (e.g., Pepper, Bootsma, Mestre, and Bakker, 1994). To coordinate his/her action with the displacement of a moving object, the actor has to determine the time remaining before the occurrence of the collision. This time remaining before the contact (time-to-contact, TC, (Lee, 1976)), requires the perception and the use of information relating to the displacement of the object.

According to the proponents of the cognitive approach, the total execution of the action and its details of implementation would be envisaged in advance. In this prospect, Tydesley and Whiting (1975) proposed the "operational timing" hypothesis corresponding to the selection of a motor plan, specifying the parameters of the movement before its execution, in a motor program adapted to the situation. These twenty last years, this model of a predictive planning of the movement was the subject of many criticisms. Hofsten (1987) and Lee (1980) supposed that such a complex system would have doubtless great difficulties to adapt to the temporal constraints imposed by certain problems and would undoubtedly be excessively sensitive to the informational noise. Indeed, this approach seems too complex to explain the way in which individuals control their action in the case of unexpected events like a velocity change of an object to be intercepted.

There is no prediction but an on-line adaptation of the movement on the basis of a law of control allowing to continuously control the displacement of the effector. This type of control strategy is often called prospective strategy. When an actor has to intercept a moving object at constant velocity, the optical invariant tau (_ ; *the inverse of the relative rate of dilatation of the object's optical contours on the retina* ; Lee, 1976), directly specifies the time-to-contact (TC1 ; *first order temporal relation between the individual and the mobile which approaches*). The adaptation of the movement is carried out on the basis of a law of control, making it possible to continuously control the displacement of the effector according to the optical invariant tau.
The study of Savelsbergh, Whiting and Bootsma (1991) provides evidences on the use of the optical variable tau to adjust a gesture in progress. Indeed, by modifying the optical invariant tau (by changing the diameter of a ball), they show a continuous adaptation of the gesture compared to the information provided by the dilation of the ball, and thus the proof of a perception-action coupling to control a

movement already initiated. Moreover, it is possible to consider that the specificity of this coupling determines the level of expertise precisely reached by the actor.

In this context, Bootsma and van Wieringen (1990) set up a cinematographic study of the attacking forehand of players of high level table tennis in which the participants had to strike a ball in approach. The authors showed that the variability of the racket orientation accross trial was greater at the beginning of the movement than at the moment of ball-racket contact. The decrease of the movement variability with the approach of the ball and the extreme accuracy of contact in expert players were interpreted by the authors as being the result of a continuous regulation of the movement with the approach of the ball or, in other words, as the result of a strong coupling between perception and action on the part of the experts (see also Bootsma, Houbiers, Whiting, & van Wieringen, 1991). This hypothesis was also provided by Tresilian (1995) who suggested that the nature of perception-action coupling could become more specific with practice.

It appears that the constraints of the task provide the principles of organization for action and perception at the same time. The fact that the action is more or less fast or starts more or less late does not seem to have consequences as long as a coupling between the perceptive and motor variables is maintained. This greater variability at the time of the initiation of the propelling phase of the gesture rather than at the time of the contact ball-racket is clearly opposed to the assumption of Tydesley and Whiting (1975). Indeed, if the movement were controlled exclusively on a prescriptive mode, the variability should be more important at the end of the action than at the beginning, because of the noise necessarily produced by the effector system during the movement. It thus appears that the problem of the control of interceptive actions cannot be simplified with the only control of the moment of initiation of the movement, but that it rests on a prospective strategy of control.

Recent studies provided experimental evidences supporting this type of strategy. Peper, Bootsma, Mestre & Bakker (1994) proposed the “velocity required model”, capturing a control mechanism that depends on the use of an informational support specifying the difference between the current behavior and the required behavior. In accordance with the model, the actor couples action and perception without having recourse to any prediction or anticipation. The authors showed that the actor control the amount of hand acceleration produced (Ah) on the basis of the difference between the required hand velocity (Vh req) and the current hand velocity (Vh) (Eq.1). At every moment, the required hand velocity is equal to the ratio of the current lateral distance hand-ball (Xh - Xb) to the current TC1 between the ball and the axis of the hand movement (Eq.2). The TC1, at every moment, is equal to the ratio of the current distance between the ball and the axis of the hand movement (Yb) to the current velocity of the ball(Vb) (Eq.3).

Ah = _Vh req – _Vh (Eq.1)
Vh req = (Xh – Xb) / TC1 (Eq.2)
TC1 = Yb / Vb (Eq.3)
with _ and _ constants.

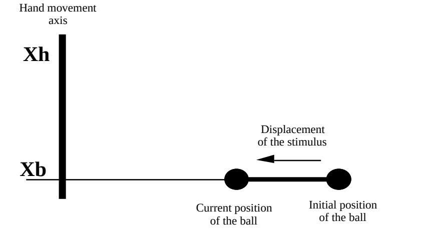

Hand movement axis
Xh
Displacement of the stimulus
Xb
Current position of the ball
Initial position of the ball

Figure 2 : Schema of the task of Peper et al. (1994).

The required velocity model was tested by Montagne, Laurent, Durey and Bootsma, (1999), in a catching task in which the gesture is mechanically constrained according to an axis, by modifying the current lateral distance (from the angle of approach of the mobile) whereas the space-time characteristics of the point of contact remained the same ones. The results show on the one hand that for a starting position of the hand, the kinematic of the produced movements depends on the angles of approach (returning or outgoing). In addition, even if the hand position, at the beginning, coincides with the place of capture, movement reversals, on the axis of the movement, appear when the angle of approach creates a lateral distance, because of the use of a law of control in order to cancel continuously the lateral distance. Moreover, the results obtained indicate that the differential between the current behavior and the required behavior is stabilized around zero, on average 300 ms before the contact like Peper et al. (1994) had already noted it. Thus, it appears that the regulations produced by the subject consist in making equivalent the current velocity towards the required velocity.

From the same point of view, a study of Montagne, Fraisse, Ripoll & Laurent (2000) tested the validity of the required velocity model by modifying now the (TC1) between the object and the axis of the hand movement. A hitting task was proposed to the subjects in which the gesture was again mechanically constrained according to an axis, perpendicularly with the displacement of the moving object. According to these authors, the distance covered by the object and its velocity would not have consequences taken separately, but the relation between the two, i.e. the TC1, would have an interest for the control of the movement. Various combinations of distance covered and velocity which lead to the same TC1, should involve the same behavior, and thus a modification of the TC1 should cause specific adaptations. The results showed a significant influence of the TC1 on various variables. A reduction in the time of presentation of the mobile involves a reduction of the latency time, an increase of the required velocity at the time of initiation of the gesture, an increase of the maximum hand velocity and an earlier appearance of the moment of maximum velocity. Conversely, the distance covered by the mobile and its velocity do not have effects on these variables.

This results provide strong evidence in favour of a prospective control strategy to a better understanding of the regulation in interceptive actions. Moreover, Tresilian (1995) suggested that the nature of perception-action coupling could become more specific with practice. A more specific coupling could mean that experts produce more continuous regulations in their movements. Experts could also produce more adapted regulations with accelerations of the hand which are more precisely tuned to the approach of the ball (Peper et al., 1994). On the temporal level, it can be assumed that the latency of the perception-action coupling or the visuo-motor delay (VMD) could be shorter in expert players. The VMD is generally defined as the time period between visually registering some information to be used to produce an adjustment and the resulting observable movements (e.g., Brenner, Smeets, & Lussanet, 1998; Carlton & Carlton, 1987; Tresilian, 1993).

This VMD has been estimated in interceptive action to lie somewhere between 100 ms and 200 ms depending on various factors (e.g., Michaels, Zeinstra, & Oudejans, 2001). For example, Lee, Young, Reddish, Lough, and Clayton (1983) calculated theoretical VMD on the order of 50 ms to 135 ms in their ball striking task. In a ball-catching experiment in which the final part of the trajectory was occluded, Whiting, Gill and Stephenson (1970) showed that performance in catching could be affected for occlusion as short as 100 ms, suggesting that information was used until 100 ms before contact. Bootsma and van Wieringen (1990), in a table-tennis ball-striking task, as well as Savelsbergh, Whiting & Bootsma (1991), in a catching task, observed that the variability was minimal at about 100 ms before contact. These authors concluded that this phase of minimal variability corresponds to the end of the control of the action and that the time interval might correspond to a VMD.

In contrast with these results, some studies have showed longer VMD in interceptive actions. Lee et al. (1983) suggested that VMD could be longer when information is used to initiate the action than when it is used to control an action already in progress. This hypothesis was confirmed by Michaels et al. (2001) as well as Benguigui et al. (2003) who showed that the VMD between the occurrence of a critical information to start action and the beginning of the action was indeed around 200 ms.

In sum, the length of VMD is directly dependant of the use of information and could be either near 100 ms when the use is on a continuous mode or near 200 ms when this use is on a discrete mode (beginning of action or important correction of the action). It can be also be supposed that the VMD is also dependant of the expertise level of the performer.

McLeod (1987) was the first to propose and investigate the idea that experts in ball sports could have a shorter VMD. He analyzed the movement of high level cricket batters and hypothesized that they would be able to adapt their action to an unexpected deviation of the ball on the ground in a time shorter than an experimental reaction time (RT) measured on a press-button response (e.g., Keele &

Posner, 1968). The VMD for correction were located between 190 ms and 240 ms. These VMD partially confirmed the hypothesis. They were not very different from experimental simple RT (e.g., Hick, 1952). However, these VMD could be regarded as very short if one takes into account (1) the weight of the bat and the inertia that must be overcome in order to start a correction, (2) that the change of the ball trajectory was unpredictable and that consequently the VMD should be compared to multiple choice RT which are a function of the number of possible responses and are largely superior to 200 ms (e.g., Hyman, 1953).

This result was confirmed by Carlton (1992) in an experiment in which expert tennis players had to carry out forehands, hitting balls that could accelerate or decelerate after the rebound. For that, the surface of the court was manipulated using either a slippery or a rough texture ribbons to cause respectively a fast or slow unexpected rebound. Carlton showed that the participants were able to adapt their responses in a time from 150 ms to 190 ms.

However, it should be noted that the data obtained from these two studies do not make it possible to determine if the delay in adapting the action can truly be regarded as one of the determinants of expertise in ball sports because the comparison was not carried out according to the level of expertise of the players. Could this be proven, it would not only make it possible to validate the idea that experts are quicker when faced with unexpected events, but also to speculate that the necessary delays due to the perception-action coupling in all catch or hit actions are shorter. Shorter delays would, in the end, explain why experts are more precise when carrying out such actions.

7.6.4 Conclusion

In summary, the ecological approach to perception and action propose the assumption that action is organised from an adaptative and on-line regulation of the movement on the basis of a continuous coupling between the perceptual and motor systems. The ecological psychology supports nicely the idea of Enaction and we believe that interceptive actions can provide a interesting paradigm to highlight the concept of Enaction.

7.6.5 References

Benguigui, N., Ripoll, H., & Broderick, M. P. (2003). Time-to-contact estimation of accelerated stimuli is based on first-order information. *Journal of Experimental Psychology: Human perception and Performance., 29, 6,* 1083–1101.

Bootsma, R. J., Houbiers, M. H. J., Whiting, H. T. A., & van Wieringen, P. C. W. (1991). Acquiring an attacking forehand drive: the effects of static and dynamic environmental conditions. *Research Quarterly for Exercice and Sport, 62*, 276-284.

Bootsma, R. J., & van Wieringen, P. C. W. (1990). Timing an attacking forehand drive in table tennis. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 21-29.

Brenner, E., Smeets, J. B. J., & Lussanet, M. H. E. (1998). Hitting moving objects: Continuous control of the acceleration of the hand on the basis of the target's velocity. *Experimental brain research, 122,* 467-474.

Carlton, L. G. (1992). Visual processing time and the control of movement. *Vision and control* (pp. 3-29).

Carlton, L. G., & Carlton, M. J. (1987). Response amendment latencies during discrete arm movement. *Journal of Motor Behavior, 19,* 333-354.

Gibson, J. J. (1950). *The perception of the visual world.* Boston: Houghton Mifflin.

Gibson, J. J. (1958). Visually controlled locomotion and visual orientation in animals. *British Journal of Psychology, 49,* 182-194.

Gibson, J. J. (1966). *The senses considered as perceptual system.* Boston: Houghton Mifflin.

Gibson, J. J. (1979). *An ecological appoach to visual perception.* Boston: Houghton Mifflin.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology, 4*, 11-26.

Hofsten, C., von (1987). Catching. In H. Heuer & A. P. Sanders (Eds.), *Perception and action.* Berlin: Sprecherverlag, 33-46.
Hyman, R. (1953). Stimulus information as a determinant of reaction time, *Journal of Experimental Psychology, 45*, 423-432.
Keele, S. W., & Posner, M. I. (1968). Processing in visual feedback in rapid movement. *Journal of Experimental Psychology, 77,*155-158.
Lee, D. N. (1976). A theory of visual control of braking based on information about time-to-collision. *Perception, 5,* 437-459.
Lee, D. N. (1980). Visuo-motor co-ordination in space-time. In G. E. Stelmach et J. Requin (Eds.), *Tutorials in motor behavior* (pp. 281-293). Amsterdam, North-Holland.
Lee, D. N., Young, D. S., Reddish, P., Lough, S., & Clayton, T. (1983). Visual timing in hitting an accelerating ball. *Quarterly Journal of Experimental Psychology, 35*, 333-346.
McLeod, P. (1987). Visual reaction time and high-speed ball games. *Perception, 16,* 49-59.
McLeod, P., McLaughlin, C., & Nimmo-Smith, I. (1986). Information encapsulation and automaticity: Evidence from the visual control of finely timed actions. In M. I. Posner & O. S. M. Marin (Eds.), *Attention and performance XI* (pp. 391-406) Hillsdale, NJ: Erlbaum.
Michaels, C. F., Zeinstra, E. B. & Oudejans, R. R. D. (2001). Information and action in 9 punching a falling ball. *Quarterly Journal of Experimental Psychology, 54A*, 69-93.
Montagne, G., Fraisse, F., Ripoll, H., & Laurent, M. (2000). Perception-action coupling in an interceptive task : First-order time-to-contact as an input variable. *Human Movement Science, 19,* 59-72.
Montagne, G., Laurent, M., Durey, A., & Bootsma, R. J. (1999). Movement reversals in ball catching. *Experimental brain research, 129,* 87-92.
Peper, C. E., Bootsma, R. J., Mestre, D. R., & Bakker, F. C. (1994). Catching balls: How to get the hand to the right place at the right time. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 591-612.
Savelsbergh, G. J. P., Whiting, H. T. A., & Bootsma, R. J. (1991). 'Grasping' tau. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 315-322.
Tresilian, J. R. (1993). Four questions of time to contact: a critical examination of research on interceptive timing. *Perception, 22,* 653-680.
Tresilian, J. R. (1995). Perceptual and cognitive processes in time-to-contact estimation : Analysis of prediction motion and relative judgment tasks. *Perception and psychophysics, 57,* 231-245.
Tydesley, D. A., & Whiting, H. T. A. (1975). Operational timing. *Journal of Human Movement Studies, 1,* 172-177.
Warren, W. H. (1988). Actions mode and laws of control for the visual guidance of action. In O. G. Meijer and K. Roth (Eds.), *Complex movement behavior: 'the' motor-action controversy* (pp. 339-380). Amsterdam: Noth-Holland.
Whiting H. T. A., Gill, E. B., & Stephenson, J. M. (1970). Critical time intervals for taking in flight information in a ball-catching task. *Ergonomics, 13*, 265-272.

7.7 Temporal delays in action-vision loop

©Damien Couroussé
September 9th, 2004

When one speaks about latency or time delay in VE or teleoperation, it usually refers to the period of time from input action to visual display. This is one of the main technological bottlenecks since it deals with incompressible time delays due to the processing of data coming from tracking systems, the processing of data that have to be delivered back, and the displaying of these new data.

The study of the human perception of time delays falls within many areas. It is of first interest for the psychophysics, but also for fields where the technological design has a direct impact on human behavior, such as teleoperation, immersive and non-immersive VE, training simulators, etc. Moreover, the study of human perception of time delays is of prime interest for VE designers, since the cost of time required for data processing is one of the main bottlenecks.

The following article review mainly deals with VE systems, where the user's movements of both head and hands are visually displayed in real-time. Studies of time delay perception in VE commonly proceed as follow: the subject is equipped with a visual head mounted display, which provides him or her a stereoscopic view of the virtual scene. A tracking system gives head and hands position and orientation, and allows integrating the user in the virtual scene. Thus, display of the virtual scene depends on the position and the orientation of the user's head.

Usual tracking systems provide refresh rates of 60 or 120Hz, and the minimum overall full latency of the full system is about 30ms – Adelstein & al (2003) report a minimum overall latency in their system 33±5ms, Ellis & al (1999) report 27±5ms; Ware and Balakrishnan (1994) present a head lag of 114ms, and a hand lag of 87ms. Ellis & al (2004) system ensures a base latency of 10,4ms at a tracking refresh rate of 60Hz.

The literature on manual control has long established that latency in displays or controls has a major negative impact on performance (Sheridan & Ferrel, 1963, Sheridan, 1992). Overall latency in visual display is a barrier to perceived image stability, drastically weakens the quality of performance, causes diseases such as disorientation and cybersickness –or as known as simulator sickness (Frank & al 1988, Allison & al 2001)– and impedes the subjective sense of presence. Lags in hand movements can degrade performance in grasping or pointing tasks, and lags in head movements can generate apparent motion of objects or space where they are still. Adelstein & al (2003) report this phenomenon as "image slip", which they define as "the virtual scene's artifactual concomitant motion with the observer's head resulting from time lag".

Richard & al (1996 – quoting Wloka, 1995) report that the "high ceil" of time delays threshold for control is about 300ms, above which the user loses the immersive feeling in the simulation and modifies his control strategy from continuous regulation to a "move and wait" strategy. Under this "high ceil" for time delay, continuous control remains possible and delay do not interfere so much with the user, but it still has an effect on the quality of performance and might be perceived by the user.

It has long been believed that the human perception of lag would match up with Weber's law, i.e. that the amount of perceived variation of delay would be a ratio of the base latency of the system[3]. First studies report results that go along with Weber's law: Watson & al (1997) report that (symmetrical) deviations of lag up to 40% do not affect task performance in the range of frame time commonly accepted in VE (50ms or 20fps). However, as the frame rate falls (frame time of 100ms or more, 10fps), the amount of fluctuation interacts significantly with the performance (placement time).

7.7.1 Human processing time:

One part of the studies in this domain tries to give an evaluation of the human processing time (from action to vision). The process by which an evaluation of the human processing time is provided is based on the Fitts' law (Ware and Balakrishnan, 1994).

The Fitts' paradigm describes the time taken to acquire a visual target with a hand or arm movement. The most commonly used formula is applied to movements performed along one dimension or to movements that can be reduced along an axis:

$MeanTime = C_1 + C_2 \log_2(\frac{D}{W} + 1.0)$, where:

[3] Ernst Weber was the first to describe the difference threshold mathematically. Weber's law can be stated as follows: for any particular sensory system, the ratio of the difference in stimulation divided by the original stimulation is a constant. Different sensory systems have different constants.

$$\frac{\Delta I}{I} = K$$

Weber's Law, more simply stated, says that the size of the *just noticeable difference* (i.e., *delta I*) is a constant proportion of the original stimulus value.

- D is the distance from the end of the movement to the center of the target
- W is the target width
- C_1 and C_2 are experimentally determined constants. $1/C_2$ is also defined as the index of performance [bit.s^{-1}], and ID is defined as: $ID = \log_2(\frac{D}{W} + 1.0)$

The process modeled by Fitts is a series of movements each of which gets the hand-guided probe closer to the target, until the probe falls within the target area. However, as the hand does not come to a complete stop in reaching movements, a series of corrections are applied in a dynamic feedback loop. Thus, the model of Fitts provides a good accuracy for a mean time evaluation.

In order to introduce machine-processing lags, Ware and Balakrishnan propose the following formula, still for movements performed along one direction:
$MeanTime = C_1 + C_2(C_3 + MachineLag)ID$, where:

- C_1 represents the sum of the initial response time and the time required to confirm the acquisition of the target.
- C_2ID represents the average number of iterations of the control loop
- C_3 is the human processing time to make a corrective movement
- *MachineLag* is the machine processing time (i.e. the delay introduced by the machine)

7.7.2 Human sensibility to overall latency:

In three experiments, by comparing head and hand movements in base conditions (i.e. with the base latency provided by the system) and in four other conditions where supplementary lags are added independently to head or hand movements, Ware and Balakrishnan provide an estimation of the human processing time of 166ms (C_3). Therefore, they show that in the specific proposed task there is no significant effect of increased lags applied to head movements, and that frame rates below 15Hz drastically increase the mean response time. These results are consistent with previous estimates provided by the literature (Carleton 1981, Keele and Posner 1968, quoted by Ware and Balakrishnan), which are given between 100 and 200ms.

In an other study based on the Fitts' paradigm, Ware and MacKenzie (1993), still on the study of a target acquisition task, show that lag has effects on movement time, error rate, and bandwidth (bits/s) as the value becomes greater or equal to 75ms. Lags of 8.3 and 25 ms induce quite the same results on the overall performance, and therefore do not seem to perturb hand movements.

7.7.3 Discrimination of changes in latency

Another way to approach the human sensibility to time delay is to determine how much is the human perceptive to *variations* of latency. Ellis & al (1999) had subjects comparing a reference situation, of which latency was equal to the minimal full system latency of 27±5ms or was fixed to 97 or 196ms, with a test situation, in which several latency increments of 16.7ms were added. The subjects were asked to discriminate between the reference situation and a test situation, by observing a virtual scene composed of a ball of 10cm diameter, while moving the head back and forth.
The results show that the discrimination of latency does not follow the Weber's law: JND[4] seems indeed to be roughly independent of the base latency introduced.
Therefore, the authors recommend designers of VE environments to avoid important variations of latencies, even if the base latency is important: users of long latency VE systems will be as sensitive to changes in latency as those who use prompter systems.

In another study, still based on the comparison of a base reference situation to a test situation, in which supplementary latency is added, Adelstein & al (2003) show that the average JND is about 17ms, and that the mean PSE is about 50ms, whatever the duration of the base latency (33, 100 or 200ms) is. The

[4] The latency difference which is detected 50% of time when actually present in two alternative forced choice situation is the point of subjective equality (PSE).
The change in latency required to increase or decrease detection 25% from the PSE is the just noticeable difference (JND) (in Ellis & al, 2004).

explanation proposed is that the perception of delays in VE might be due in part to "image slip", which is the apparent movement of the image as a consequence of head tracking latency, or as defined by Adelstein & al "the virtual scene's artifactual concomitant motion with the observer's head resulting from time lag". Therefore, as the subject moves his head in a sinusoidal movement, for certain sinusoidal motion frequencies, the amount of added incremental delay produces the same proportions of image slip in displacement, regardless of the base latency. This could explain in part why the results are not according to Weber's law.

Ellis & al (2004) confirmed these results. They found JND about 10-15ms, and PSE about 30ms, for a system where the base latency was only of 10.4ms. The aim of the study was to determine if the perception of latency was depending on the complexity of the visual scene or not. It was shown that three different visual scenes, one with a single object displayed in foreground with a black background (i.e. no background), one with of a single background (interleaved staircases) without foreground, and the last with both a foreground object and a background, provided no difference on the results. In a further experiment (to be published), the authors shown that latency discrimination is not depending on the complexity of the visual scene.

7.7.4 Adaptation to temporal latency

The problems of spatial and temporal misalignment between sensorialities cause diseases to the users of VE and impair overall performance. However, it has been long shown that a short learning stage of a few minutes permits human to adapt to spatial misalignment between sensorialities (this is the case study of *prism adaptation*). After this recalibration, the spatial misalignment does not impair performance anymore, nor is even felt as disturbing. Cunningham and al. (2001) studied the case of temporal misalignment in a driving task without force feedback, and shown that the subjects managed to adapt to a temporal delay of 250ms, after a training period had been performed. However, after training to temporal delay, a negative *aftereffect*[5] was observed when the temporal delay removed or set to its minimum value (35ms – this minimum value was said not to be noticed by the subjects). Therefore, the error rates drastically increased and gave similarly results to a performance of the delayed task without training. In another study, Cunnigham et al. (2000) report that, if the subjects complained at first about the delay, they did not noticed it after some training, and even improved their performance compared to the pre-test, which was measured with a 35ms delay. Several subjects "spontaneously reported that, toward the end of training, the visual and proprioceptive feedback seemed simultaneous". The authors furthermore add that the subject's perception of latency was modified such that "when the delay was removed, the plane seemed to move before the mouse did".

Bürki-Cohen and al. (1996) established a review of comparative studies between full flight simulators and fixed-based simulators. Fisex-based simulators only differ from full flight simulators in all respects except for absence of platform motion. Quoting Levison & Junker (1977), they reported that large reductions of mean-squared error were observed in a tracking task, in all conditions of delay (no delay, 80, 200 and 300ms). The group submitted to the 200 and 300 ms of time delay was however exposed to some aftereffects. For that reason, the authors to be avoid motion platform in flight simulators if the time delay could not be minimized, since they stated that badly synchronized motion is in fact worse than no motion at all.

7.7.5 References

Adelstein, Bernard D., Thomas, G. L., Ellis, Stephen R. *(2003). Head tracking latency in virtual environments: psychophysics and a model,* Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting.

Crowley, James L., Coutaz, Joëlle, Bérard, François (2000). Things That See, *Communications of the ACM 43, 3, pp 54-64, ACM Press New York, NY, USA.*

[5] The aftereffect is the fact that the adaptation to an intersensory discrepancy reduces a subject's ability to accurately perform the task once the time delay is removed: the benefits of training are gone or reversed.

Bürki-Cohen, J., Soja, N., Longridge, T. (1996). Simulator Platform Motion – The Need Revisited. Draft Submitted to the *International Journal of Aviation Psychology*. Washington Dulles Airport Hilton, June 1996.

Cunningham, Douglas W., Chatziastros, Astros, von der Heyde, Markus, Bülthoff, Heinrich H. (2000). Temporal Adaptation and the role of temporal contiguity in spatial behavior, *Technical Report No. 85*, MPI, December 2000

Cunningham, Douglas W., Biillock, Vincent A., Tsou, Brian H. (2001). Sensorimotor Adaptation to Violations of Temporal contiguity, *Psychological Science,* Volume 12: Issue 6

Ellis, Stephen R., Mania, Katerina, Adelstein, Bernard D. and Hill, Micheal, (2004). Generalizeability of latency detection in a variety of virtual environments, *Human Factors and Ergonomics Society, 48th Annual Meeting, New Orleans, USA (to appear)*

Ellis, Stephen R., Young, Mark J. and Adelstein, Bernard D. (1999). Discrimination of changes in latency during head movement, *Proceedings of Computer Human Interfaces, pp. 1129-1133*

Frank, L. H., Casali, J. G., & Wirewill, W. (1988). Effects of visual display and motion system delays on operator performance and uneasiness in a driving simulator. *Human Factors, 30,* 201-217

MacKenzie, I.S., Ware, C. (1993). Lag as determinant of human performance in interactive systems. *Proceedings of the ACM Conference on Human Factors in Computing Systems – INTERCHI'93,* 488-493. New-York:ACM.

Richard, P., Birebent, G., Coiffet, P., (1996). Effect of frame rate and force feedback on virtual object manipulation, *Presence, Vol. 5, No. 1, 95-108*

Richard, Paul, Birebent, Georges, Coiffet, Philippe, Burdea, Grigore, Gomez, Daniel, Langrana, Noshir (1996). Effect of Frame Rate and Force feedback on Virtual Object Manipulation *Presence, Vol 5, No. 1, 95-108.*

Stanney, Kay M., Mourant, Ronald R., Kennedy, Robert S., (1998). Human Factors Issues in Virtual Environments: A Review of the Literature, *Presence, Vol 4, No. 4, pp. 327-351*

Ware, Colin, and Balakrishnan, Ravin, (1994). Reaching for Objects in VR Displays: Lag and Frame Rate, *ACM Transactins on Computer-Human Interaction, Vol 1, No 4, pp. 331-356.*

L. James Smart, Jr., Miami University, Oxford, Ohio, Thomas A. Stoffregen, University of Minnesota, Minneapolis, Minnesota, and Benoît G. Bardy, University of Paris Sud XI, Orsay, France: Visually Induced Motion Sickness Predicted by Postural Instability, *HUMAN FACTORS*, Vol. 44, No. 3, Fall 2002.

Watson, B., Spaulding, V., Walker, N., Ribarsky, W. (1997) Evaluation of the Effects of Frame Time Variation on VR Task Performance, *VRAIS '97, IEEE Virtual Reality Annual Symposium, 38-44*

8 STAR in Technologies for action-vision fusion

©Jarlier Sophie, George Papagiannakis, HyungSeok Kim, UNIGE

8.1 Gestural devices. Terminology

©Annie Luciani, INPG

The terms “haptic devices” or “tactile devices”, currently used, lead to the confusion between the human perception and the characteristics of the device. Haptic and Tactile refer to human sensory-motor apparatus. In addition, they are related to various approachs as shown by E. Pasquinelli [Pasquinelli 2004]. Thus, it is unclear to qualify transducers by the human sensory or motor modality : on one hand the classification of sensori channels, as shown by E. Pasquinelli are various according to basic assumptions, not yet unified. In other hand, there is no bijective relation between the signals produced and received by the transducers and the sensori-motor modalities. For example, force feedback devices stimulate the internal muscular mechano receptors as well as the cutaneous or pain ones, and conversely, tactile devices have to be resistant to the penetration and thus they stimulate also the deep tissues and muscular mechanoreceptors.

The stimuli received by the human who manipulate the device (and through it, the distant object) are complex and at least – but not only - composed of muscular kinaesthetic and tactile perceptions.

That is the first care to take into account in these new type of transducers and the haptic human skills, as well know in others fields (as optic flow, transducers and perception or acoustical flow, transducers and perception): the well known non bijective relation between the external universe space and the perceptual and cognitive space.

To avoid these misunderstanding, we propose to generalyse the terms “haptic”, “force feedback”, “tactile” in the global term “gestural”. Gesture means all what it could be done by the human body using all the mechano sensors-actuators apparatus during a body performance. Thus:

- gestural action is the motor part of the gestural channel involved in the gestural performance. It involves all the physical components (articulated skeleton and muscles) of body.
- gestural perception is the sensory part of the gestural channel. All the terms declined from “gesture” or “gestural”, (gestural channel, gestural perception, gestural action) allow to avoid the detailed description of each sub-means (subset of sensors, subset of motor capabilities) as well as the human perceptual and/or cognitive results of the use of these means. In addition, it allows to avoid the discrepancies of meanings of the term ‘haptic” (See “Haptic modality. Some definitions and problems of classification” by E. Pasquinelli, WP4b State of the Art).
- In addition, Cadoz [Cadoz1994] and Luvciani [Luciani 2004] suggest to use the term “gestural action” instead of “action”: action, in a general use, might refer not only to the performance itself but also the result of this performance.

In such terminology, Gestural devices states all the electromechanical devices that are inserted between the human body and a electrified machines to convey gestural action and/or perception.

8.1.1 Retroactive Gestural Transducers

8.1.1.1 Definition

Among all the gestural devices, that could be any electromechanical sensors or actuators, some of them are devoted to convey ergotic bilateral interaction during which there is an energu exchange between th ehuman body and the machine through the gestural device. That are the device usually called either “haptic device” or “force feedback device”. The two terms are restrictive. The amin properties of these devices being to convey mechanical gestural data and to be retroactive, it is suggested to use the term of “mechanical retroactiove transducers (or devices”) or “gestural retroactive transducers”, ‘or devices”. The term « Mechanical », or “getsural” address the nature of these physical data ; « Retroactive » addresses the link between what it is sensed (i.e. the signals sensed from the human mechanical action via the sensors 1) and it is returned to the human mechanical perceptions (i.e. the signals provided by the pair 2 of sensors-actuators). The term « Transducers » addresses the property of fidelity in the signal coding of the physical data, that is the main issue of the field of instrumentation.

8.1.1.2 Systems

©Jarlier Sophie, George Papagiannakis, HyungSeok Kim, UNIGE

During the past fifteen years, we attended to several successful attempts that integrate visual and acoustic perception channels in VR systems. Nevertheless, one major component was missing in order to enhance the realism of virtual worlds: the sense of touch or haptics (force/tactile). Haptics allow a higher rate of interaction and the simulation of virtual object physical characteristics and provide physical constraints in a virtual world. Gestural sensori-motor channel is more complex than other purely sensory channels (sight and audition) seeing that it allows humans to emit and in the same time to receive information to and from the external environment. First development of gestural retroactive trabsducers devices appeared in 1978 [Batter and al. 1971][Atkinson 1977][Florens 1978][, but incorporating gestural perception in VR systems comes with many unresolved problems and is still at the beginning.

Nowadays haptic devices are commercially available in certain companies such as SensAble Technologies, Immersion Corp., Force Dimension, etc. Most of them use three degrees of freedom (DOF); also six DOF devices have then been introduced.

We don't present here all the gestural retroactive transducers developed in laboratories and that are not already commercialized. The state of the art on existing systems and technologie swill be done in WP3.

We will summarize here, the best known commercial systems to illustrated each possible classes of devices:

- **Arm-like devices**: The user grasps a robotic arm with several degrees of freedom that can apply forces. The **PHANTOM** (see figure3), from *SensAble technologies* is probably the most widely used haptic device of all. It is an example of manipulandum (mechanical interface between the user and the virtual environment). It has 3 (positional) or 6 DOF (positional + yaw, pitch and roll) depending on the version. It can be used for any application that requires manipulating objects with a probe in a virtual environment. The **Dextrous Arm** from *Sarcos* is a big robotic arm that is used most of the time for remote manipulation of actual robotic arms. It isn't used at all in virtual reality probably due to its oversized dimensions and extra features that aren't useful in VR systems. **Freedom 6S** from *MPB Technologies* is a device that does almost the same as the PHANTOM: it has 6 DOF and can be used with most standards PCs. However, its design is slightly different and it is by far less used than the PHANTOM is.

- **Exoskeletons**: These devices are less disseminated than the previous class, but they offer a much higher level of realism. They allow the simulation of mechanical properties such as stiffness and stretch. The **CyberGrasp** (see figure4), from *Immersion Corp.* is a hand exoskeleton that can separately apply up to 12 N to each finger, but it must be used jointly with the **CyberGlove** that will track the configuration of the hand and an external motion tracking system (Polemus, Vicon …) that will provide the position of the hand in space. It can be used to manipulate virtual objects by directly grasping them: the feeling of grasping is induced by the exoskeleton that will constrain the finger to match the shape of the virtual object that is being manipulated. The **CyberForce** also from Immersion is an extension of the CyberGrasp: it associates this device with a Phantom-like one, thus enabling one to constrain not only the fingers, but also the position and orientation of the palm itself. The **Utah/MIT Dextrous Hand Master** is yet another hand exoskeleton which is very similar to the CyberGrasp (however, it doesn't require to use a DataGlove for evaluating the configuration of the hand and the measure is much more accurate). The **Rutgers Master II** is a hand exoskeleton like the two previous ones, but it is attached inside the hand. It exerts its force only on 4 fingers (the pinky is excluded) and it uses the device itself for evaluating the configuration of the hand.

- **Tactile displays**: These devices are not force feedback but rather tactile feedback devices i.e. they give the sense of touch: shape, texture, temperature. The **CyberTouch** from *Immersion Corp.* (see figure5) has six vibro-tactile stimulators (one for each finger, one on the palm) and it can generate simple pulses or sustained vibration or more complex tactile feedback patterns. The **Fingertip Stimulator Array** from the *University of Exeter* is also a vibro-tactile stimulator. However, it is not designed to fit on a glove and it has a much higher spatial (up to 400Hz) and temporal resolution than the CyberTouch (1 pin per square millimeter). The **Elastic Force Sensor** from *Iwata Lab* is a deformation based stimulator: it is cylinder shaped and can deform itself to closer the shape of the virtual object. Also, it measures the pressure that is applied to it (while grasped for example) so that the simulation can combine the deformations due to the user and the ones that are induced by the simulation itself.

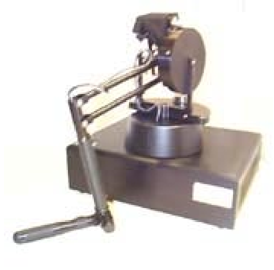


Figure3. Phantom

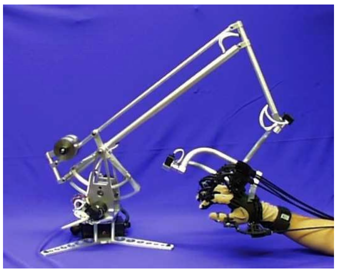


Figure4. CyberGrasp

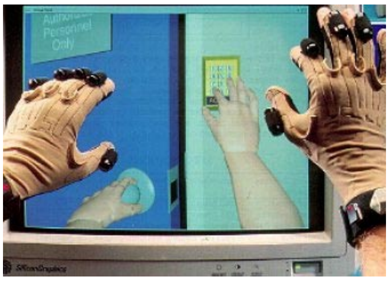


Figure5. CyberTouch

Haptic technology is difficult to achieve convincingly. If the interface were sensitive enough, the degree of specificity of manipulation could be very high. But today, the technologies are not mature and their effectiveness is not conclusive. Some problems of manipulation arise with the use of haptic devices such lack of precision, inexactness of user movement, force feedback engine not realistic and difficult maintenance of the arm by the user at the same position. CyberGrasp and Exoskeleton cause tiredness (muscular, blood pressure and circulation). Grasping, manipulation with force feedback with the goal to have the same feeling in a real world or to perceive a semblance or reality are not convincing. The combination of haptic display devices and visual display requires synchronizing the tactile/haptic and visual interfaces in order to achieve a consistent and physically possible experience.

References

[Atkinson 1977] ATKINSON, W.D., BOND, K.E.,TRIBBLE, G.L., WILSON, K.R. - Computing with feeling - Comput. and Graphics, Vol 2 – 1977

[Florens 1978] FLORENS, J.L. - Coupleur gestuel rétroactif pour la commande et le contrôle de sons synthétisés - Thèse INPG, Grenoble, France – 1978

[Batter and al. 1971] Batter, J.J. and Brooks, F.P., Jr. GROPE-I : A computer display to the sense of feel. *Information Processing, Proc. IFIP Congress 71, 759-763.*

8.1.2 Non-Retroactive Gestural Transducers: Sensors for motion capture

©Jarlier Sophie, George Papagiannakis, HyungSeok Kim, UNIGE

Nowadays, two main trends made it up to the wide spread commercial level in motion capture: optical and magnetic systems. Several companies fight each other on the market of optical systems and, among others, one can cite Vicon Motion Systems Ltd [VICON] and Motion Analysis [MOTIONANALYSIS] as key actors of the market. All these companies have more or less the same technical level and thus offer almost the same performances in terms of hardware features: 4 million pixels cameras seem to be the norm now and 1Khz capture rate isn't exceptional any more (however, one must be aware that the highest the frame rate is, the smallest the capture volume becomes). All these systems use infrared light for isolating reflecting markers (figure 1) on the scene: infrared emitters (figure 2) are placed on the cameras and infrared filters are placed in front of the aperture so that only the light re-emitted by the markers (which are basically just small spheres with reflecting tape on them) make its way to the camera sensors. Once this happens, and being given that the cameras are calibrated (i.e. that the internal geometry and its location in the 3D space are known), simple triangulation calculation enable the system to reconstruct the location of the markers in 3D. Of course, one can suspect that each company that manufactures such system do have secret algorithms for making the reconstruction more robust, but none of these were released to the public so far. Because markers are often occluded by other elements of the scene or by the body itself, a high number of cameras are required for acquiring the data efficiently: if one has many different angles of view, then the chances for a marker to be occluded drop down. Most systems offer up to 24 cameras, but the techniques that are used for reconstructing the markers should in theory support as many cameras as you wish.

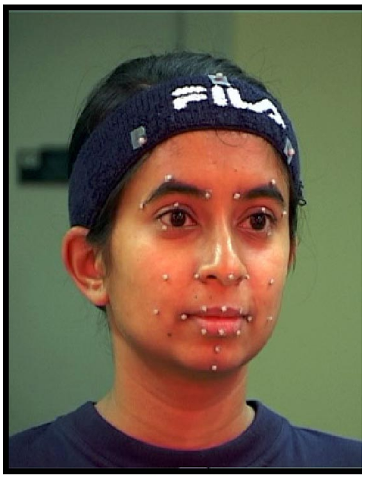


Figure 1: a setup of reflective markers for a facial animation capture session

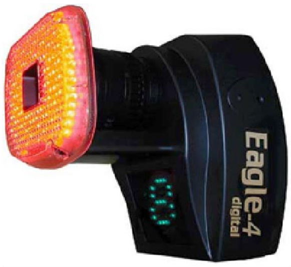


Figure 2: A motion capture camera with an array of infrared emitters on its front

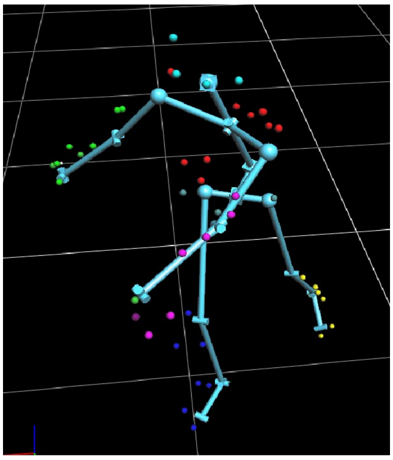


Figure 3: A skeleton estimati (light blue colored) from t markers reconstruction (t small rainbow colored balls).

Even though this class of system is generally used in combination with an offline processing of the data, for high quality production or motion analysis, there exist some software tools that enable one to perform the capture and visualize it in real time. For instance, Vicon IQ [VICONIQ] has an embedded tool kit that allows real time optical capture with skeleton fitting and real time post processing (up to a limited number of markers to post process though). Because this process uses as much computational resources as it can for improving the skeletal fitting onto the reconstructed markers (figure 3), it is generally used in a networking configuration: one mocap server capture and post process the data, and then sends the markers location plus the rigid body estimation over the network for future use by another machine (like interacting with a virtual environment for instance)

Figure 4: a magnetic receiver

The second wide spread way of capturing motion is the use of magnetic systems [ASCENSION] [POLEHMUS]: Markers are small magnetic receivers (figure 4) that analyze the magnetic field produced by an emitter, thus computing their position and orientation in space relatively to this emitter. This class of systems has several advantages such as no calibration of the system and ease of setup. However, the drawbacks are at least as numerous as the advantages: high sensitivity to metallic pieces in the surrounding and low frame rate (no more than 150 Hz).

Probably because of its ease of use, this kind of trackers is often found as a mean of interaction with virtual worlds, either by itself (for tracking the general position and orientation of a user inside a VE for instance) either in combination with another device (e.g. the data glove from immersion corp. which track the motion of the fingers needs another tracker for an actual 3D interaction as it cannot locate itself in the 3D space.)

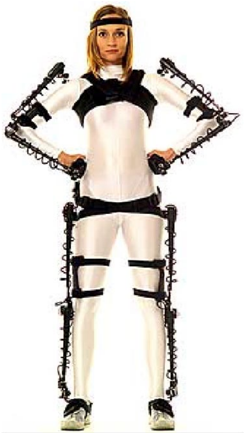


Figure 5: an exoskeleton for mocap

There exist a third class of system which, even though it is used quite a lot, is still limited to a very specific class of motion to capture: the mechanical systems (figure 5) [IMMERSION] [METAMOTION]. Basically, the user either wears an exoskeleton on the part of his body to be captured or manipulate a kind of puppet, and the motion is recorded through this device. The limitations implied by such technologies prevented them from really breaking through the motion capture field so far (indeed, it is much more difficult to modify the setup of an exoskeleton than to move small markers around).

Besides these commercial trends, a huge amount of alternative systems, some of them being experimental, some others limited to very specific applications are used case by case, depending on the goal that is aimed: estimating the pose of a character, the motion of a rigid object, etc. However, the biggest achievement that many are currently investigating would be to accurately motion capture a full body without any kind of markers. Even if some attempts have been made in this direction, the way is still long before one obtains the same kind of performances as commercial systems offer. Amongst the attempts to get rid of any kind of markers, Luck and Al. [LUCK] used 4 cameras and silhouette extraction techniques for tracking a full body in real time, Chu and Al.[CHU] used generic volumes fitting for capturing body motion, Alexander and Al.[ALEX] used non rigid body estimation for improving the result of their markerless captures.

Because the virtual character that is animated rarely fits exactly to the subject that is captured, quite a lot of work has been done on motion retargeting and editing. Shin and al. [SHIN] recomputed the physical parameters out of a motion capture sequence and then reinforce their physical correctness by modifying the motion. Choi and Al. [CHOI] developed a system that enables one to edit and to retarget a motion in real time. Popovic and Witkin [POPO] develop an algorithm based on space time constraints that enables one to edit a captured motion while preserving its essential features.

References

[VICON] The vicon web site: http://www.vicon.com
[MOTIONANALYSIS] The Motion Analysis web site: http://www.motionanalysis.com
[VICONIQ] The IQ technical sheet: http://www.vicon.com/proddetail.jsp?id=164
[ASCENSION] The Ascension technologies corp. web site: http://www.ascension-tech.com
[POLHEMUS] The Polhemus web site: http://www.polhemus.com
[IMMERSION] The immersion corp. web site: http://www.immersion.com
[METAMOTION] The Meta Motion corp. web site: http://www.metamotion.com
[SHIN] Hyun Joon Shin, Lucas Kovar, Michael Gleicher, *Physical Touch-up of human motions*, 11th Pacific conference on computer graphics and applications, 2003
[CHOI] Kwang-Jin Choi; Hyeong-Seok Ko. *Online motion retargeting*, The Journal of Visualization and Computer Animation 11(5):223-235, 2000
[POPO] Zoran Popovic and Andrew Witkin. *Physically Based motion transformation*, proceedings of SIGGRAPH 99, 1999.

[ALEX]: Alexander, E.J., Bregler C., Andriacchi, T.P.: *Nonrigid Modeling of Body Segments for Improved Skeletal Motion Estimation*, Computer Modeling in Engineering and Science, Vol 4, Number 3 & 4, pp. 351-364, 2003.
[CHU]: Chi-Wei Chu, Odest Chadwicke Jenkins, Maja J Matari´c, *Markerless Kinematic Model and Motion Capture from Volume Sequences*, Proceedings of IEEE computer vision and pattern recognition, Madison, Wisconsin, USA, June 16-22, 2003
[LUCK]: Jason Luck, Dan Small, Charles Q. Little, *Real-Time Tracking of Articulated Human Models Using a 3D Shape-from-Silhouette Method*, Proceedings of the International Workshop

8.2 Stereo hardware devices. Stereoscopic visualisation devices

©Jarlier Sophie, George Papagiannakis, HyungSeok Kim, UNIGE

8.2.1 Introduction

Human and computer stereo vision are based on the same principle: Two eyes separated by a few inches see the objects in a scene slightly differently. The objects seen by the left eye appear shifted horizontally compared to the same objects seen by the right eye. The size of the shift varies with an object’s distance from the eyes; nearby objects are shifted more than far ones. A stereo camera has two imagers placed like eyes. It produces two images simultaneously; the left and the right, which are together called a stereo pair or binocular perspective (see **Erreur! Argument de commutateur inconnu.**). Each pixel in the right image has a corresponding pixel in the left image, which represents the same element of the scene as viewed by the left imager. If matching pixels have the same coordinates in the left and right images, they represent an element of the scene that is infinitely far from the camera. If matching pixels have different coordinates (that is, one is shifted with respect to the other), the pixels represent a scene element that is closer than infinity. The size of the shift, called the disparity, varies according to the range of the object the pixels represent; the closer the object is to the camera, the larger the disparity.

8.2.2 Stereoscopic displays

A stereoscopic display is one that differs from a planar display in only one respect: It is able to display parallax values of the image points. Parallax produces disparity in the eyes, thus providing the stereoscopic cue. The stereoscope is a means for presenting disparity information to the eyes. In the CrystalEyes or SimulEyes VR display (**Erreur! Argument de commutateur inconnu.**), the left and right images are alternated rapidly on the monitor screen. When perceiving Stereoscopic Images through shuttering eyewear, each shutter is synchronized to occlude the unwanted image and transmit the wanted image. Thus each eye sees only its appropriate perspective view. The left eye sees only the left view, and the right eye only the right view. If the images (the term “fields” is often used for video and computer graphics) are refreshed (changed or written) fast enough (often at twice the rate of the planar display), the result is a flickerless stereoscopic image. This kind of a display is called a field-sequential stereoscopic display.

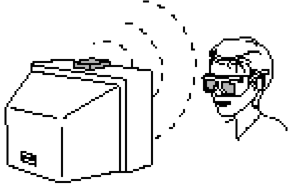
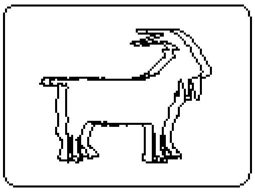
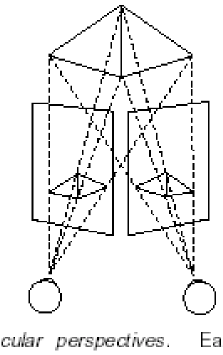
 *CrystalEyes.* An emitter, sitting on the monitor, broadcasts an infrared signal (dotted lines). The eyewear receives the signal and synchronizes its shutters to the video field rate.	 *Retinal disparity.* The image that would be seen on overlayed left and right retinae (provided the eyes were not converged on the goat).	 *Binocular perspectives.* Each eye sees a slightly different perspective view, shown by the projection of the pyramid onto the plane surfaces.

Figure Erreur! Argument de commutateur inconnu. **Stereo images**

When you observe an electro-stereoscopic monitor image without our eyewear, it looks like there are two images overlayed and superimposed. The refresh rate is so high that you can't see any flicker, and it looks like the images are double-exposed (**Erreur! Argument de commutateur inconnu.**). The distance between left and right corresponding image points (sometimes also called "homologous" or "conjugate" points) is parallax, and may be measured in inches or millimeters. Parallax and disparity are similar entities. Parallax is measured at the display screen, and disparity is measured at the retinae. When wearing our eyewear, parallax becomes retinal disparity. It is parallax which produces retinal disparity, and disparity in turn produces stereopsis. Parallax may also be given in terms of angular measure, which relates it to disparity by taking into account the viewer's distance from the display screen. CrystalEyes. An emitter, sitting on the monitor, broadcasts an infrared signal (dotted lines). The eyewear receives the signal and synchronizes its shutters to the video field rate.

Active stereo Eyewear

The active eyewear approach as it was first introduced above, uses wireless battery-powered eyewear with liquid crystal shutters that are run in synchrony with the video field rate. Synchronization information is communicated to the eyewear by means of an infrared (IR) emitter. The emitter looks at the computer's video signal and seeing the vertical blanking synchronization pulse broadcasts coded IR pulses to signify when the left eye and the right eye images are being displayed. The eyewear incorporates an IR detection diode that sees the emitter's signal and tells the eyewear shutters when to occlude and transmit. An active eyewear product, like CrystalEyes by StereoGraphics, has shutters with a dynamic range better than 1500:1,

Passive stereo Eyewear

An alternative to active eyewear is the ZScreen, which is a special kind of LC polarization modulator. It is placed in front of the projection (**Erreur! Source du renvoi introuvable.**) lens(es) like a sheet-polarizing filter. The device changes the characteristic of polarized light and switches between left and right-handed circularly polarized light at field rate. Audience members wear circular polarizing analyzing eyewear. Although the great majority of theatres showing stereo movies use linearly polarized light for image selection, circular polarized light has the advantage of allowing a great deal of head tipping before the stereoscopic effect is lost.

Stereo VR-Helmets/Head Mounted Displays

These are basically wearable monitors (**Erreur! Argument de commutateur inconnu.**). To allow stereoscopic vision there is a little LCD or CRT monitor for each eye. Consumer products like VFX-1, i-glasses or Cybermaxx have two LCD monitors with an effective resolution of 640x480 pixels or even higher. In addition some of these helmets have a head tracker (which replaces or complements keyboard, mouse or joystick input by head-movement) and some stereo-headphones.

Figure Erreur! Argument de commutateur inconnu. **HMD stereo**

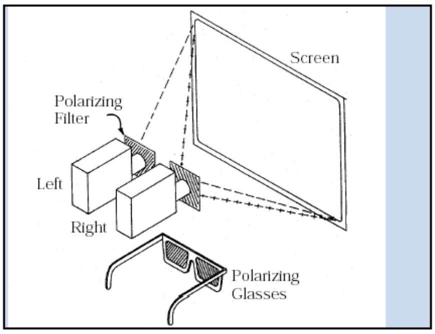


Figure **Erreur! Argument de commutateur inconnu.** Passive stereo

Figure **Erreur! Argument de commutateur inconnu.** HMD stereo

8.2.3 Evaluation

Monocular depth cues are part of electronic displays, just as they are part of the visual world. While the usual electronic display doesn't supply the stereoscopic depth cue, it can do a good job of producing a seemingly three-dimensional view of the world with monocular cues. These depth cues, especially perspective and motion parallax, can help to heighten the stereoscopic depth cue. We have seen that a stereoscopic display differs from a planar display by incorporating parallax. Parallax is classified into positive parallax, which produces an image distant from the viewer (in CRT space), and negative parallax, which produces off-screen effects or images in viewer space. The perceptual artifacts introduced by the limitations of display tubes and the human visual system may be held to tolerable limits by properly adjusting the parallax. A 3-D database may be displayed stereoscopically and many graphics packages or games may be upgraded to produce a true binocular stereoscopic image. In most cases, this is best left to the software developer who has access to the source code. Moreover, the developer is best able to support his or her product. Experience has shown that adding stereoscopic capability to an existing package is straightforward — a word not used lightly in a world where software development often requires heroic efforts. All computer-generated images produced from a three-dimensional database require the computation of an image from a single perspective point of view or location in space. A stereoscopic image differs only in that the image must be generated from two locations.

As an alternative, the image may be captured from the real world by means of a camera. In other words, it may be photographed or videographed. Whether the image is computed or photographed, the generation or capture of images from two perspectives — a left and a right — is required. If a single perspective viewpoint can be produced, it is conceptually straightforward to create the additional perspective. The creation of such images is the subject matter of this handbook. People in the field of interactive computer graphics have also been playing with language by using the terms "3-D" or "three-dimensional" to mean a "realistic"-looking image which may be appreciated with one eye. Unfortunately, this is confusing nomenclature because most people who speak English outside of the field of computer graphics use the term "3-D," and on occasion "three-dimensional," when they mean a stereoscopic image. Such an image requires the beholder to use both eyes. The usual so-called 3-D computer image loses nothing when you close one eye, but a true stereoscopic image loses its raison d'être when viewed with one rather than two eyes. In this section we have covered a number of solutions that are provided today and according to the authors, best results are provided with Passive Z-Screen stereo equipment albeit with a high cost projection device.

8.3 Real time computer Graphics

©Jarlier Sophie, George Papagiannakis, HyungSeok Kim, UNIGE

In the following section, the current State of the Art (STAR) of commercial middleware and scenegraph APIs acting as 3D virtual frameworks for Real-time Graphics simulations is presented.

As needs of the real-time graphics grow, many different types of middleware has been developed and used. From the low level graphics system such as GL (Graphics Library), OpenGL, Direct3D and Glide, the needs of the managing scene graph for interaction requires higher level of software abstraction. Inventor and Performer[6] [Inventor][Performer] are the early middleware APIs. Early APIs has characteristics of;

- Optimizing rendering performances by re-ordering entities of scene graph.

Caching and compression methods are implemented as a base layer of these systems, thus generic high-level scene graph is compiled to low-level primitives which can be rendered efficiently on the hardware.

- Primitive interaction

These APIs only supports only primitive interactions which are based on the transform of the position.

[6] Initially created by SGI based on GL and further extended to Open Inventor and OpenGL Performer which are based on OpenGL

Next generation of the middleware like WorldToolKit [WTK] supports more higher-levels of interactions, such as navigation and manipulation. Also, these middleware begin to support some levels of simulations including basic level of autonomous behaviors and simple collision detection.
On the other aspect, a set of systems has been proposed to support multi-user interactive virtual worlds [DIVE][NPSNET]. These systems are based on the simple middleware and supports networked interaction which is mostly directly manipulation of an object.

Recent development on the middleware can be categorized as;

• Enhanced simulation

Current systems especially those for the 3D games, have a facility of simulating physical environment in real-time [HAVOK]. These physical parameters include dynamic (force and acceleration) parameters of environment, the object and the control of the participants. These parameters are not only used for the dynamic simulation of movement but also used for simulating non-rigid objects like fluids, clothes, etc. The most of systems also include real-time accurate collision detection mechanism.

• Animation support

Character animation using articulated figure became basic feature of high-level scene APIs. Furthermore, many systems support skinning based animation and cloth animation which gives more natural animation results.
Some systems support more complex animation mechanism using motion blending and retargeting techniques [LithTech][Phoenix3D][Genesis3D]. It enables the application designer to implement more complex animation systems to make an animation according to different environments (e.g. different terrain conditions, slopes, etc.)

• Sophisticated scene optimization for the rendering

In addition to the caching and compression based optimizations, various high-level optimization mechanisms are implemented in the middleware. Efficient scene culling mechanisms using space partitioning (e.g. BSP based partitioning) are portal based scene management adopted widely since its introduction by a few systems for games.
Also, more aggressive scene optimization method is applied to some systems by adopting multiresolution meshes [ALCHEMY][NETIMMERSE][UNREAL]. In addition with the development of hardware, these optimizations give higher rendering speeds in acceptable rendering quality.

• Advanced rendering primitives

Polygon mesh has been the most popular primitives for the 3D rendering for decades. Although it is a powerful yet efficient way to represent and render 3D shapes some effects are not suitable to be represented by polygon mesh such as flames and fluids. Some systems supports particle system and/or other special effects [ALCHEMY][NETIMMERSE][RENDERWARE].
In addition to the special effects, illumination models are evolved through out the generation. Currently, environment mapping, shadow generation are light mapping become one of the widely adopted illumination primitives. In addition, more sophisticated illumination models such as subsurface scattering is begin to be adopted in the middleware following advance of the hardware in these area.

• Autonomous behavior

Starting from the primitive path following behavior, some advanced 'artificial intelligence' is adopted especially for entertainment applications [LITHTECH][UNREAL]. These features are implemented through scripts or codes that conduct activities of sensing the environment and reacting to the environments.

References

[ALCHEMY] Alchemy, http://www.intrinsic.com
[BLAXXUN] Blaxxun, http://www.blaxxun.com
[DIVE] Dive: Distributed Interactive Virtual Environment, http://www.sics.se/dive/
[FILMBOX] FilmBox, http://www.kaydara.com
[GENESIS3D] Genesis3D, http://www. Genesis3D.com/
[Inventor] IRIS Inventor Programming Guide, Silicon Graphics, 1992
[JAVA3D] Java3D, http://java.sun.com/products/java-media/3D/
[LITHTECH] LithTech, http://www.lithtech.com/

[MAYART] Maya Real-Time SDK, http://www.aliaswavefront.com/en/WhatWeDo/maya/see/solutions/realtime_sdk.shtml)
[NETIMMERSE] NetImmerse, http://www.ndl.com/
[NPSNET] NPSNET, http://www.npsnet.org/~npsnet/v/
[OPENPERFORMER] OpenGL Performer, http://www.sgi.com/software/performer/
[Performer] John Rohlf and James Helman, IRIS performer: a high performance multiprocessing toolkit for real-time 3D graphics, ACM SIGGRAPH 1994, pp. 381-394, 1994
[PHOENIX3D] Phoenix3D, http://www.4xtechnologies.com
[QUAKEIII] Quake III, http://www.quake.com/
[RENDERWARE] RenderWare, http://www. renderware.com/
[UNREAL] Unreal, http://unreal.epicgames.com/
[WTK] World Tool Kit, Sense8, http://sense8.sierraweb.net

8.4 Spatial information through 3D sounds (Not in this Deliverable, For the next deliverable, Could you summarize for this Deliv.)

©Jarlier Sophie, George Papagiannakis, HyungSeok Kim UNIGE

3D spatial audio in virtual environments is a relatively new and wide research topic, although spatial audio in general has been under investigation since the beginning of the last century. Rendering audible a modeled acoustical space in such a way that the three-dimensional sound illusion is preserved is called auralization according to Kleiner [Kleiner1993]. Devices to render 3D sound are described below.

8.4.1 3D Auditory Display

The final stage of the auralization pipeline consists in reproducing a three-dimensional sound field for the ears of the listener. We can highlight two main kinds of techniques:

- **Binaural and Transaural Techniques:** Head Related Transfer Functions (HRTFs) provide filters that model the overall effects of head, ear and torso on sound propagation.

The goal is to recreate the wave field at both ears of the listener, using headphones (binaural techniques). The MIT media lab made available HRTFs measurements of a KEMAR dummy head [Gardner1994].

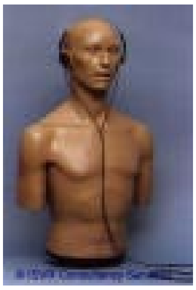


Figure2. KEMAR dummy head

- **Multi-channel Auditory Displays (wave field synthesis)**

An array of loudspeakers is placed around the listening area to reproduce directional sound waves. This method can reproduce correct localization cues without the expense of HRTF filtering. Such techniques are usually used for large audiences but tend to suffer from sweet spot problems.

8.4.2 3D spatial, binaural and surround sound systems

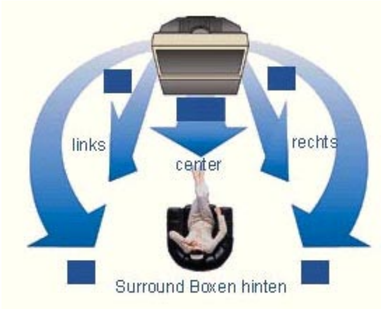
After the use of mono with the sound emanating from a single point, and stereo with directional information spread along a line in front of the listener, surround sound arose necessary to improve the realism of perceived sound by providing information from all directions around the listener. Surround sound became an essential cue in media production, film, television, radio and computer games. Sound systems are numerous and are a wide field of research for companies such as Dolby laboratories, Sensaura, aBSOLUTE sOUND SySTEMs, Surround Assiciates, Linn Products, Lexicon, Kelly industries, etc. In the movies, there are three main formats:

8.4.2.1 Dolby Stereo Digital (DSD)

Dolby Surround:
It is the consumer version of the original analog Dolby multichannel film sound format. Four channels of audio (Left, Centre, Right and mono Surround, LCRS) are encoded into a single stereo pair (Left total Lt – Right total Rt). Dolby Stereo moved on to Dolby Digital and then to Dolby Surround used in the television industry and home video.

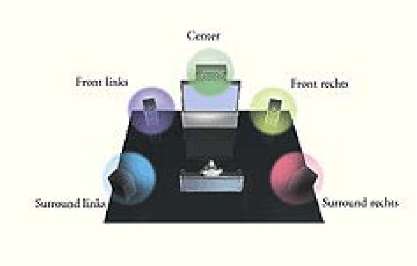
Dolby Surround Pro Logic:

It decodes program material encoded in Dolby Surround. It is used into virtually every home theater audio system. It reconstructs the original four channels – left, center, right and surround – that were encoded onto the program material's stereo soundtracks.

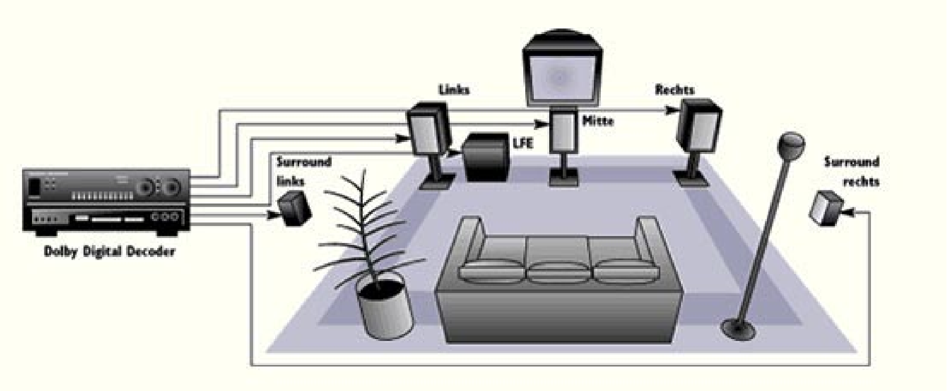


Dolby Surround Pro Logic II:

It improved matrix decoding technology that provides better spatiality and directionality on Dolby Surround program material.



Dolby Digital:



The digital surround format relies upon Dolby's AC3 data reduction process by eliminating not distinguishable audio information. It is essentially used in film industry and adopted for consumer products. This format comprises left, centre and right front channels, left and right surround channels and sub-bass/effects channel and is called 5.1.**:** identifies the use of Dolby

Digital (AC-3) audio coding for such consumer formats as DVD and DTV. As with film sound, Dolby Digital can provide up to five full-range channels for left, center and right screen channels, independent left and right surround channels and a sixth ("..1") channel for low-frequency effects.

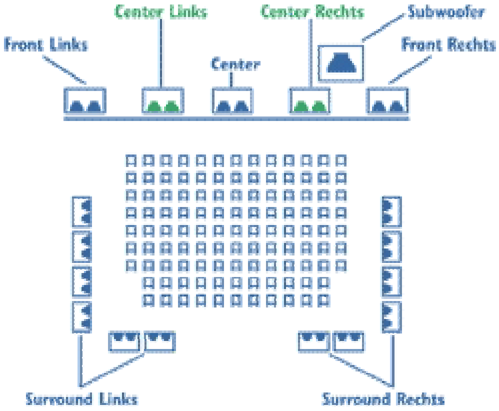
- **DTS (Digital Theater Systems) Systems**

DTS Digital Surround:

It is the standard for providing 5.1 channels of discrete digital audio in consumer electronics products and software content. DTS 5.1 encoding is made for CD. You can use a DTS encoder to encode your 5.1 mixes and put them onto a CD-R

- **Sony Dynamic Digital Sound (SDDS)**

SDDS is the newest of these three formats. It provides an easy ability to record and print SDDS films. It is for the moment only available for movie theaters.



It is also important to quote the ***Ambisonics sound system*** which has the potential to localize a sound anywhere. It is used to create a better sound reproduction system by providing more information than with stereo or Dolby. Is not limited to creating a reverberant rear sound field, and requires a different arrangement of speakers. All speakers cooperate to localize sounds, so the front-rear time delay is unnecessary.

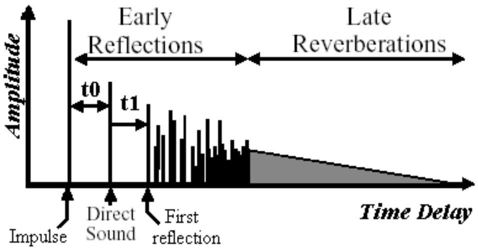
8.4.2.2 Issues of sound in immersive reality systems

The audio is as or even more important than the video. If the surround sound is believable to your ear and brain, you will be there but not here, you will be transported. The surrounding sound defines the environment all around you. Nevertheless, the problem encountered to give this sense of realism and believability comes from speakers which produce noises and distortions. Plus, higher the number of sources and the complexity of the virtual environment is, higher the rendering of an accurate sound in real-time is complex.

8.4.3 Fundamental Elements for Sound Rendering

8.4.3.1 Impulse Response

The impulse responses due to sound creations construct the acoustic environment in three parts: direct sound (earliest strongest waves), early reflections (waves arrived within the first t_e ms period) and late reverberations (their density is low enough that the human ear is able to distinguish individual paths, less than 2000reflections/s. e.g.) as show below in figure 1. Early reflections stand between 20, 30 to 80ms with a speed of the sound of 345m/s in the air.



8.4.3.2 3D rendering

For basic knowledge of simulation and rendering of sound in virtual environments, we can refer to Funkhouser [Funkhouser2002], Savioja [Savioja1996], Huopaniemi [Huopaniemi1999], and the book written by Begault [Begault1994]. Some fundamental elements already existent are necessary for a complete spatial audio system:

- **Transmission** – this takes into account the absorption of the sound by the air, energy distribution, and also how sound is affected by surfaces between the source and the listener either by direct path, or via any of the reflections.
- **Reflections** – this takes account of the sound propagation around an environment by multiple early reflections possible due to reflective surfaces.
- **Reverberation** – reverberations are later reflections that are not uniquely distinguishable; they are generally only possible in closed environments such as a room.
- **Head Related Transfer Function (HRTF)** – this computation calculates the sound propagation / interference changes due to the listener's head and the ear's pinnae.
- **Diffraction** – This deal with how sound is diffracted when striking either a very irregular surface, or edges of surfaces.
- **Refraction** – refraction is the bending of waves when they enter a medium where their speed is different.

As can be observed, some of these elements are affected by the position of the sound source relative to the listener (or receiver) and others are affected by the environments itself. This is absolutely necessary to take this fact in consideration for an immersive environment, where the user makes the correspondence between the 3D visual effect and the 3D audio effect. Indeed if the user is in a normal room, but its 3D audio response is for metal or concrete, it will sound very bizarre.

Methods

Different methods have been studied to propagate sound in an immersive virtual reality environment, such as:

- **Finite Element/Boundary Method (FEM/FBM)**

  This numerical solution solves wave equation by subdividing space into elements. Wave equations are expressed as a discrete set of linear equation for these elements. This method is especially used in vibration mechanics in room acoustics and gains interest for low frequencies and simple environments but require high compute time and storage space which increase dramatically with frequencies.

- **Propagation methods**

3D spatial acoustics can be rendered using several different models, most having their own particular advantages and drawbacks. Different methods have been developed for the rendering of sound. They are described here:

Image Source [Allen1979]: Image Source methods compute specular reflection paths by generating virtual sources by mirroring the location of the audio source S over each polygonal surface of the environment.

Ray Tracing [Kulowski1984]: Ray Tracing methods find reverberation paths between a source and receiver by generating rays emanating from the source position through the environment until an appropriate set of rays has been found that reach the receiver position.

Beam Tracing [Funkhouser2004]: Beam Tracing methods classify propagation paths from a source by recursively tracing pyramidal beams (i.e. sets of rays) through the environment.

8.4.3.3 Validation of Acoustical Models

When implementing a virtual acoustics rendering pipeline, we should ask ourselves “Is the acoustics modeling method accurate enough?” and “Is the signal processing pipeline delivering appropriately spatialized sound to the listener?”. Unfortunately, there is no definitive guideline for answering these questions. Based on subjective observations (exchange of opinions between acousticians and musicians in the case of concert halls), a relative set of objective criteria have been designed that describe how energy is distributed in the impulse response:

- Energy decay such as reverberation time, early decay time, etc…
- Clarity that measures a ration of early to late energy in the impulse response
- Binaural linked to our stereophonic perception, which measure the “sensation of space” or “envelopment” perceived by the listener.

Results of simulations can be compared against measurements in real spaces or using scale models by using graph which depends on time and frequencies.

8.4.4 References

[Allen1979] J.B. Allen and D.A. Berkeley: Image method for efficiently simulating small-room acoustics. *Journal of Acoustics Society of America,* vol.65, pp.943-950, 1979

[Begault1994] D.R. Begault: *3D sound for virtual reality and multimedia*. Academic Press Professional, 1994

[Casier2003] G. Casier, P. Plenacoste, C. Chaillou, B. Semail: The DigiHaptic, a new three degrees of freedom multifinger haptic device. *In Virtual Reality International Conference*, pp. 35, 2003

[Fisch2003] A. Fisch, C. Mavroidis, J. Melli-Huber, Y. Bar-Cohen: Chapter 4: Haptic Devices for Virtual Reality, Telepresence, and Human-Assistive Robotics. *In Biologically-Inspired Intelligent Robots. SPIE Press,* 2003

[Fong2003] T. Fong, C. Thorpe, B. Glass: PDADriver: A handheld system for remote driving. *In Proceedings of IEEE International Conference on Advanced Robotics, Coimbra,* 2003

[Frisoli2002] A. Frisoli, F. Simoncini, M. Bergamasco : Mechanical Design of a Haptic Interface for the Hand, *ASME International DETC- 27th Biennial Mechanisms and Robotics Conference,* 2002

[Funkhouser2002] T. Funkhouser, JM. Jot and N. Tsingos: "Sounds good to me" computational sound for graphics, virtual reality and interactive systems. *In SIGGRAPH 2002 Conference Proceedings,* 2002

[Funkhouser2004] T. Funkhouser, N. Tsingos, I. Carlbom, G. Elko, M. Sondhi, J.E. West, G. Pingali, P. Min and A. Ngan: A beam tracing method for interactive architectural acoustics. *Journal of the Acoustical Society of America*, 2004

[Gardner1994] B. Gardner and K. Martin: HRTF measurements of a KEMAR dummy-head microphone. *MIT Media Lab Perceptual Computing, technical report 280*, 1994

[Gomez1995] D. Gomez, G. Burdea, N. Lagrana: Integration of the Rutgers Master II in a virtual reality simulation. *Proceedings of Virtual Reality Annual International Symposium ’95, IEEE Computer Society Press,* pp. 198-202, 1995

[Gregory2000] A. Gregory, S. Ehmann, M.C. Lin: In-Touch:interactive multiresolution modeling and 3d painting with haptic interface. *In Proceedings of IEEE International Conference on Virtual Reality 2000,* 2000

[Huopaniemi1999] J. Huopaniemi. Virtual acoustics and 3D sound in multimedia signal processing. *Thesis*, 1999.

[Kawasaki2003] H. Kawasaki, J. Takai, Y. Tanaka, C. Mrad, T. Mouri: Control of multi-fingered haptic interface opposite to human hand. *In Proceedings of the International Conference on Intelligent Robots and Systems,* 2003

[Kleiner1993] M. Kleiner, B.-I. Dalenbäck and P. Svensson: Auralization – an overview. *Journal of the Audio Engineering Society vol.41, pp.861-875*, 1993

[Kulowski1984] A. Kulowski: Algorithmic representation of the ray tracing technique. *Applied. Acoustics, vol.18, pp.449-469,* 1984

[Mendoza2001] C. Mendoza, C. Laugier: Realistic haptic rendering for highly deformable virtual objects. *In Proceedings of IEEE International Conference on Virtual Reality 2001,* pp. 264–269, 2001

[Nguyen2001] L.A. Nguyen, M. Bualat, L.J. Edwards, L. Flueckiger, C. Neveu, K. Schwehr, M.D Wagner, E. Zbinden: Virtual reality interfaces for visualization and control of remote vehicles. *In Autonomous Robots,* Vol. 11, 2001

[Nitzsche2001] N. Nitzsche, U. Hanebeck, G. Schmidt: Mobile haptic interaction with extended real or virtual Environments. *In Proceedings of 10th IEEE International Workshop on Robot and Human Interactive Communication*, 2001

[Savioja1996] L. Savioja, J. Huopaniemi, L. Lokki and R. Vnnen: Virtual environment simulation – Advances in the DIVA project. *Proceedings of ICAD'96*, 1996

[Snibbe2001] S.S Snibbe, K.E. MacLean, R. Shaw, J. Roderick, W.L. Verplank, M. Schee®: Haptic techniques for media control. *In Proceedings of the 14th annual ACM symposium on User interface software and technology, ACM Press*, 2001

[Springer1997] S.L. Springer, R. Gadh: Haptic feedback for virtual reality computer aided design. *Concurrent Product Design and Evironmentally Conscious Manufacturing American Society of Mechanical Engineers,Design v94 1997, ASME,* pp1-8, 1997

[Springer1999] S. Springer, N. Ferrier: Design of a multifinger haptic interface for teleoperational grasping. *In ASME Int'l Mech. Eng. Congress and Expo,* 1999

[Takala1992] T. Takala, J. Hahn: Sound rendering. *In SIGGRAPH 1992*, 26(2):211-220, 1992.

[Tarrin2003] N. Tarrin, S. Coquillart, S. Hasegawa, L. Bouguila, M. Sato: The stringed haptic workbench: a new haptic workbench solution. *In Proceedings of Eurographics,* 2003

[Williams1998] H.R.L. Williams, D. Noorth, M. Murphy, J. Berlin, M. Krier: Kinesthetic Force/Moment Feedback via Active Exoskeleton. *Proceedings of the Image Society Conference,* 1998

II. ANNEXE 1. List of scientific fields addressed by the action-vision relationship

	Field	Short definition
C1	Human-Machine Interfaces	Including : Human-Computer Interfaces and Interaction Multimodal interfaces Tangible Interfaces Multimodal interactive systems
C2	**Virtual Reality**	Including : Immersive and non immersive VR Mixed Reality Augmented Reality
C3	**Computer Graphics**	Including : Computer animation Geometric modeling Interactivity in Synthetic Images
C4	Teleoperators and teleoperation	
C5	Simulators	Including : driving simulators Surgical simulators Interactive simulations
C6	**Computer games**	Including: Interactive dance and music systems Music theatre applications Interactive art installations Systems for distributed performances
C7	Performance Arts	Music, Dance, Visual Arts, Multimedia
C8	Artificial Intelligence	
C9	Artificial Life	
C10	Psychophysics of action-vision	Including : Co-location of visual and haptic spaces Interceptive actions
C11	Philosophy of action-vision	
C12	Entertainment	Including: Interactive games Applications for home entertainment
C13	Training and edutainment	Including: Interactive systems for training Interactive systems for learning
C14	Cultural heritage	Including: Museum applications Interfaces for museum visitors Interactive exhibits
C15	Information visualisation	
C15	Scientific visualisation	

III. ANNEXE 2. List of Keywords

keywords	Author of the suggestion	**short definition (if necessary)**
Motion capture devices	A. Luciani (INPG)	Including : Technologies for motion and gesture capture
Gesture tracking	S. Jarlier (UNIGE)	Including: Gesture recognition, motion recognition
Auralization	S. Jarlier (UNIGE)	Including: 3D audio spatialization, sound rendering, virtual acoustics, spatial hearing, sound visualisation
3D modeling	S. Jarlier (UNIGE)	Including: Image warping, sketching, 3D reconstruction
Real time 3D synthesis and rendering	S. Jarlier (UNIGE)	Including: Animation systems, deformation, morphing, computer games, physically-based modeling, virtual reality
Haptic interface	S. Jarlier (UNIGE)	Tactile feeling, haptic device, haptic force feedback
Crossmodal attention	N. Bullot (NICOD)	
Deictic sensorimotor primitive	N. Bullot (NICOD)	
Epistemic seeing (versus simple seeing)	N. Gangopadhyay (NICOD)	
Haptic modality	E. Pasquinelli (NICOD)	
Sensorimotor theories of Perception	N. Gangopadhyay (NICOD)	
Touch as sense of reality	E. Pasquinelli (NICOD)	
Co-location (of visually and haptically presented and perceived objects)	Gunnar Jansson (UPPSALA)	There is a physical and a perceptual definition. The physical one says that co-location means that the visual and the haptic displays are calibrated to have coincident co-ordinate systems. The perceptual one says that co-location means that a visually and haptically perceived object is experienced to be at the same location in a common space.
Sign Languages	F. Pfaender (COSTECH)	Semiology
Perception-Action Coupling	C. Le Runigo (UPS)	
Invariants	C. Le Runigo (UPS)	Specify the state of the actor-environment system.
Affordances	C. Le Runigo (UPS)	Opportunities for action, property of the actor-environment system.
Perception-Action Coupling	C. Le Runigo (UPS)	Circular and reciprocal link between perception and action.
Law of control	C. Le Runigo (UPS)	Link between perceptual invariant and control parameter (force).
Prospective control	C. Le Runigo (UPS)	“On-line” adaptation of the movement.

Prescriptive control	C. Le Runigo (UPS)	Predictive planning of the movement.
Interceptive Actions	C. Le Runigo (UPS)	Catching, hitting, contact with objects...
Visuo-Motor Delay	C. Le Runigo (UPS)	Time period between the occurence of the critical information and the resulting observable adaptation.
Expressive Gesture	G. Volpe (DIST)	A gesture conveying an expressive content
Expressive Content	G. Volpe (DIST)	Information related to the emotional sphere.
Expressive Cues	G. Volpe (DIST)	Features responsible of conveying expressive content in expressive gestures.
Expressive non-verbal communication	G. Volpe (DIST)	Communication of expressive content using non-verbal mechanisms, e.g., with expressive gestures.
Expressive interface	G. Volpe (DIST).	An interface for human-computer interaction taking explicitly into account communication of expressive content.
Motion Analysis	G. Volpe (DIST).	Automatic analysis of human movement in order to extract and process motion features.
Motion Features	G. Volpe (DIST).	Descriptors of human motion (or of specific aspects of human motion).
PSE – Point of Subjective Equality	Damien Couroussé, INPG	Only available for the two alternative forced choice technique The latency difference that is detected 50% of time when actually present in two alternative forced choice situation is the point of subjective equality (PSE).
JND: Just Noticeable Difference	Damien Couroussé, INPG	Only available for the two alternative forced choice technique: The change in latency required to increase or decrease detection 25% from the PSE is the just noticeable difference (JND) (in Ellis & al, 2004).

IV. ANNEXE 3. List of commented references

1 Commented references in geometrical, light and visualization modeling and computation

Real-Time Rendering, Tomas Akenine-Möller and Eric Haines, A.K. Peters Ltd., 2nd edition, ISBN 1568811829
Rendering is the intensive process of creating a picture or sequence of frames based on geometry. The duration of this process is dependent on the complexity of the scene (a forest with many trees and thousands of leaves will take much longer to render than a scene consisting of a white box over a gray background) and the speed of the hardware doing the calculations. The authors answer the question, not only asserting that it can be done, but since this book is a programmer's guide, they list snippets of programming algorithms that help outline *how* it can be done.

Hierarchical geometric models for visible surface algorithms, James H. Clark, Communications of the ACM, 19(10): 547-554, October 1976
The geometric structure inherent in the definition of the shapes of three-dimensional objects and environments is used not just to define their relative motion and placement, but also to assist in solving many other problems of systems for producing pictures by computer. By using an extension of traditional structure information, or a geometric hierarchy, five significant improvements to current techniques are possible. All of them are presented inside the paper.

Fast backface culling using Normal Masks, H.Zhang e K.Hoff, , ACM Interactive 3D Graphics Conference, 1997
This paper presents a method for fast and efficient backface culling. The backface test is reduced to one logical operation per polygon while requiring only two bytes extra storage per polygon. The normal mask is introduced, where each bit is associated with a cluster of normals in a normal-space partitioning. A polygon's normal is approximated by the cluster of normals in which it falls; the cluster's normal mask is stored with the polygon in a pre-processing step.

Occlusion (HP and NV Extensions), Ashu Rege, NVIDIA website, www.nvidia.com/developer
This article illustrates how to use OpenGl extensions, provided by new-generation graphical boards, to perform hardware-accelerated occlusion culling. Basically, they provide a mechanism for determining "visibility" of a set of geometry. After rendering geometry, query if any of the geometry could have or did modify the depth buffer. If occlusion test returns false, geometry could not have affected depth buffer. If it returns true, it could have or did modify depth buffer. This means that the object is not visible if the test fails (returns false). It is visible if the test passes (returns true) in "usual" circumstances.

Visibility preprocessing for interactive walkthroughs, Seth J.Teller, Carlo H.Sèquin, Proceedings of the 18th annual conference on Computer graphics and interactive techniques 1991
The number of polygons comprising interesting architectural models is many more than can be rendered at interactive frame rates. However, due to occlusion by opaque surfaces (e.g., walls), only a small fraction of a typical model is visible from most viewpoints. The authors describe a method of visibility preprocessing that is efficient and effective for axis-aligned or *axial* architectural models. A model is subdivided into rectangular *cells* whose boundaries coincide with major opaque surfaces. Non-opaque *portals* are identified on cell boundaries, and used to form an *adjacency graph* connecting the cells of the subdivision. Next, the *cell-to-cell* visibility is computed for each cell of the subdivision, by linking pairs of cells between which unobstructed *sightlines* exist. At each frame, the cell containing the observer is identified, and the contents of potentially visible cells are retrieved from storage.

Progressive meshes, H. Hoppe, proceedings of SIGGRAPH '96, pp.109-118 (1996)
Highly detailed geometric models are rapidly becoming commonplace in computer graphics.

These models, often represented as complex triangle meshes, challenge rendering performance, transmission bandwidth, and storage capacities. This paper introduces the *progressive mesh* (PM) representation, a new scheme for storing and transmitting arbitrary triangle meshes. This efficient, lossless, continuous-resolution representation addresses several practical problems in graphics: smooth geo-morphing of level-of-detail approximations, progressive transmission, mesh compression, and selective refinement.

Simulation of Wrinkled Surfaces, James F. Blinn, Computer Graphics, Vol. 12 (3), pp. 286- 292 SIGGRAPH-ACM (August 1978)

Computer generated shaded images have reached an impressive degree of realism with the current state of the art. They are not so realistic; however, that they would fool many people into believing they are real. One problem is that the surfaces tend to look artificial due to their extreme smoothness. What is needed is a means of simulating the surface irregularities that are on real surfaces. This paper presents a method of using a texturing function to perform a small perturbation on the direction of the surface normal before using it in the intensity calculations. This process yields images with realistic looking surface wrinkles without the need to model each wrinkle as a separate surface element.

Dynamically generated Impostors, Gernot Schaufler,MVD '95 Workshop "Modeling – Virtual Worlds – Distributed Graphics" (Nov.95) D.W.Fellner (ed.) Infix pp. 129-136

Interactive graphics systems need to generate more than 20 frames per second which usually bear a close resemblance to each other but each are rendered from scratch without using these similarities. Although costly graphics hardware is usually built into virtual reality systems to rapidly render the whole geometry database for each frame, low system performance in complex virtual environments is very disturbing to the user. This paper presents a new rendering algorithm which exploits the coherence in computer-generated frame sequences by avoiding the actual re-rendering of the major part of the geometry database. It reuses most of the image data generated during previous frames to decrease the number of polygons actually rendered by an order of magnitude. Complex distant objects are replaced by textured polygons with an image of the objects mapped onto them.

Texturing & Modeling: A Procedural Approach, 3rd Edition, David S. Ebert, F. Kenton Musgrave, Darwyn Peachey, Ken Perlin, and Steven Worley

This book provides a tutorial and a reference on procedural texturing and modeling, thoroughly updated to meet the needs of today's 3D graphics professionals and students. There are chapters devoted to real-time issues, cellular texturing, geometric instancing, hardware acceleration, futuristic environments, and virtual universes. In addition, other authoritative chapters on which readers have come to rely contain all-new material covering L-systems, particle systems, scene graphs, spot geometry, bump mapping, cloud modeling, and noise improvements.

Point-Based Modeling, Markku Reunanen, Helsinki University of Technology

This book provides a tutorial and a reference on procedural texturing and modeling, thoroughly updated to meet the needs of today's 3D graphics professionals and students. There are chapters devoted to real-time issues, cellular texturing, geometric instancing, hardware acceleration, futuristic environments, and virtual universes. In addition, other authoritative chapters on which readers have come to rely contain all-new material covering L-systems, particle systems, scene graphs, spot geometry, bump mapping, cloud modeling, and noise improvements.

Artificial Evolution for Computer Graphics, Karl Sims, July 1991, Computer Graphics, Vol. 25, No. 4, pp. 319 - 328 1996

This paper describes how evolutionary techniques of variation and selection can be used to create complex simulated structures, textures, and motions for use in computer graphics and animation. Interactive selection, based on visual perception of procedurally generated results, allows the user to direct simulated evolutions in preferred directions. Several examples using these methods have been implemented and are described. 3D plant structures are grown using fixed sets of genetic parameters. Images, solid textures, and animations are created using mutating symbolic lisp expressions. Genotypes consisting of symbolic expressions are presented as an attempt to surpass the limitations of fixed-length genotypes with predefined expression rules. It is proposed that artificial evolution has potential as a powerful tool for achieving flexible complexity with a minimum of user input and knowledge of details.

GPU Gems: Programming Techniques, Tips, and Tricks for Real-Time Graphics, Randima Fernando, Addison-Wesley Professional;ISBN: 0321228324
This book describes real-time graphics techniques arising from the research and practice of cutting-edge developers. The book focuses on the programmable graphics pipeline available in today's graphics processing units (GPUs) and highlights tricks used by leading developers as well as fundamental, performance-conscious techniques for creating advanced visual effects.

Texture and Reflection in Computer Generated Images, J.Blinn and M. Newell, Communications of the ACM, volume 19, 1976, pp. 542-546
In 1974 Ed Catmull developed a new algorithm for rendering images of bivariate surface patches. This paper describes extensions of this algorithm in the areas of texture simulation and lighting models. The parameterization of a patch defines a coordinate system which is used as a key for mapping patterns onto the surface. The parametric values within each picture element are input to a pattern definition function. A weighted average of the values of this function over the picture element scales the intensity of that picture element. By suitably defining the pattern function, various surfaces textures can be simulated. The shape and size of this weighting function is chosen using digital signal processing theory.The other problem addressed here concerns lighting models. The patch rendering algorithm allows accurate computation of the surface.

Illumination and Reflection Maps: Simulated Objects in Simulated and Real Environments, G. Miller and R. Hoffman, SIGGRAPH ‘84 Course Notes Advanced Computer Graphics Animation, 1984
Blinn and Newell introduced reflection maps for computer simulated mirror highlights. This paper extends their method to cover a wider class of reflectance models. Panoramic images of real, painted and simulated environments are used as illumination maps that are convolved (blurred) and transformed to create reflection maps. These tables of reflected light values are used to efficiently shade objects in an animation sequence. Shaders based on point illumination may be improved in a straightforward manner to use reflection maps. Shading is by table-lookup, and the number of calculations per pixel is constant regardless of the complexity of the reflected scene. Anti-aliased mapping further improves image quality. The resulting pictures have many of the reality cues associated with ray-tracing but at greatly reduced computational cost. The geometry of highlights is less exact than in ray-tracing, and multiple surface reflections are not explicitly handled. The color of diffuse reflections can be rendered more accurately than in ray-tracing

Recovering high dynamic range radiance maps from photographs, P. E. Debevec and J. Malik, Proceedings of SIGGRAPH'97, pp. 369 – 378
The authors present a method of recovering high dynamic range radiance maps from photographs taken with conventional imaging equipment. In their method, multiple photographs of the scene are taken with different amounts of exposure. Their algorithm uses these differently exposed photographs to recover the response function of the imaging process, up to factor of scale, using the assumption of reciprocity. With the known response function, the algorithm can fuse the multiple photographs into a single, high dynamic range radiance map whose pixel values are proportional to the true radiance values in the scene.

The CAVE--Audio Visual Experience Automatic Virtual Environment, Cruz-Neira, C., Sandin, D., DeFanti, T., Kenyon, R., & Hart, J. (1992). Communications of the ACM, 35, 6, 65-72.
The CAVE is a new virtual reality interface. In its abstract design, it consists of a room whose walls, ceiling and floor surround a viewer with projected images. Its design overcomes many of the problems encountered by other virtual reality systems and can be constructed from currently available technology. Suspension of disbelief and viewer-centered perspective, are often used to describe such systems.

Towards Image Realism with Interactive Update Rates in Complex Virtual Building Environments, Airey, John M., John H.Rohlf, and Frederik P. Brooks Jr., Computer Graphics (1990 Symposium on Interactive 3D Graphics), vol. 24, nº 2, pp. 41-50, March 1990.
This paper presents the concept of rendering a potentially visible subset rather than the entire virtual environment. Moreover, in indoors scenes, the authors proposed to divide the rooms into cells. One of the architectural databases characteristic is that there are not considerable viewing changes inside the cell except when looking at doors or windows (portals). This fact makes desirable to spent time in a pre-processing step in order to get an interactive frame rate in run-time.

Hierarchical Z-Buffer Visibility, Greene, Ned, Michael Kass, and Gavin Miller, Computer Graphics (SIGGRAPH'93 Proceedings), pp. 231-238, August 1993.
The authors proposed to change the GPU architecture in order to achieve an effective occlusion by hardware. This method uses a pyramid of z-buffer in order to determine the visibility of objects.

Visibility Culling using Hierarchical Occlusion Maps, Zhang, Hansong, D. Manocha, T. Hudson, and K.E Hoff III, Computer Graphics (SIGGRAPH'97 Proceedings), pp. 77-88, August 1997.
This method is similar to the Hierarchical Z-Buffer but designed to work in current graphic hardware. The visibility is verified through two tests: An overlap test that analyses if an occluder object hides others elements. And a depth test to resolve the proximity of the occluder.

Virtual reality for aircraft engines maintainability, A.Amundarain, D.Borro, A.Garcia-Alonso, JJ.Gil, L.Matey and J.Savall, Mécanique & Industries 5, pp.121-127 (2004)
This method is applied in the visualization of aeronautical engines digital mock-ups and customizes the HOM algorithm. Such environments are non-densely occluded where less of 50% of the scene can be occluded. Furthermore, the elements of the model have small dimensions compared to the complete model. Contrary to architectonical walkthrough, the navigation within the scene never penetrates inside the model.

3D Games: Real-Time Rendering and Software Technology Volume 1, Alan Watt, Fabio Policarpo, Pearson Education, ISBN 0201619210
With this book, authors Watt and Policarpo introduce the theory behind the design of computer games and detail advanced techniques used in the industry, such as: physically based animation; advanced scene management; pre-calculation techniques/image-based rendering; advanced motion capture; and artificial intelligence. Though these topics are usually related with computer games, they are at the same time deeply connected with the virtual reality field, as the technology requirements and the problems to be faced are often the same

Project FEELEX: Adding Haptic Surface to Graphics, Hiroo Iwata Hiroaki Yano Fumitaka Nakaizumi Ryo Kawamura, Proceeding of SIGGRAPH 2001
This paper presents work carried out for a project to develop a new interactive technique that combines haptic sensation with computer graphics. The project has two goals. The first is to provide users with a spatially continuous surface on which they can effectively touch an image using any part of their bare hand, including the palm. The second goal is to present visual and haptic sensation simultaneously by using a single device that doesn't oblige the user to wear any extra equipment. In order to achieve these goals, it was designed a new interface device comprising of a flexible screen, an actuator array and a projector. The actuator deforms the flexible screen onto which the image is projected. The user can then touch the image directly and feel its shape and rigidity.

The Effect of Haptic Feedback and Stereo Graphics in a 3D Target AcquisitionTask
Steven A. Wall, Karin Paynter, Ann Marie Shillito, Mark Wright, Silvia Scali, Prooceedings of EuroHaptics 2002
Interaction in three dimensional virtual environments is difficult, often resulting in physical or mental fatigue. Haptic interfaces have previously been employed with 2D and 2.5D desktop metaphors in order to improve targeting performance. This paper extends the principle to a 3D environment targeting

task. Subjects completed a targeting task with and without haptic feedback, in the form of "virtual magnets" that physically attract the user towards targets in the environment, and with and without the provision of stereo depth cues via a stereo emitter and shutter glasses. It was found that the virtual magnets improved subjects accuracy, but did not improve the time taken to reach the target. Stereo cues improved both the subjects' spatial accuracy, and significantly improved the temporal measure of performance.
Category: C1-C10, pdf

RELEVANT REFERENCES BY PERCRO:

Simulators:

Washout Filter Design for a Motorcycle Simulator, Federico Barbagli, Diego Ferrazzin, Carlo Alberto Avizzano, Massimo Bergamasco, Virtual Reality 2001 Conference Proceedings, March 13 - 17, 2001, Yokohama, Japan
Many motion base simulators have been developed in the last thirty years for many different types of vehicles. In order to make a simulation more realistic, linear accelerations and angular rates are exerted on the pilot by moving the platform on which the mock-up vehicle is located. This has to be accomplished without driving the simulator out of its workspace.
Washout filters have been widely investigated in the past, mainly in the field of flight simulators. In this article it is presented a washout filter designed for a motorcycle simulator.
The solution is preliminary and follows, as a reference point, techniques previously adopted for large aircraft simulators. Differences between motorcycle and aircraft simulation
are analyzed and a preliminary customized solution is proposed.
Category: C5, pdf

La prototipazione virtuale per veicoli a due ruote (italian language), Fabio Salsedo, Franco Tecchia, Marcello Carrozzino et al., Convegno Nazionali XIV ADM, Bari 2004-08-02
In motorcyclist field, the final product is made up of subsystems and groups frequently designed and realized by third party on purchaser's specifications. A Virtual Prototyping System can be very useful for interacting between components suppliers and production firm, reducing costs and time-to-market.
PERCRO (PERCeptual RObotics) and Piaggio S.p.A. have developed a VR system available in two distinct configurations, each one of them exploiting an integrated virtual environment:
• Interactive 3D Visualization System (IVS), designed for increasing assemblies quality.
• Advanced Ergonomic Simulator (AES), for increasing driving seat ergonomicity.

Dynamic modeling of primary commands for a car simulator, Frisoli A, Avizzano CA, Bergamasco M, Data S, Santi C, IEEE 2001 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM '01) 8-12 July 2001 Como, Italy.
This paper concerns the dynamic modeling of the pri mary commands of a car, i.e. the steering wheel and the gear shift. Models have been set up to replicate the forces felt by the driver during a real drive in a car simulator, aimed at evaluating the ergonomics of internals of car in Virtual Environments. The mechanical response (i.e. the exerted force related to the displacement imposed by the driver) of these primary controls are controlled via software.

Complex Virtual Environment Management, Image Based Rendering:

Image caching algorithms and strategies for real time rendering of complex virtual environments, M. Carrozzino, F.Tecchia, C.Falcioni, M.Bergamasco, Proceedings of Afrigraph 2001
This paper presents a study about the comparison of several methods of image cache management; image caching is a recent approach of rendering, based on the concepts of impostors and hierarchical scene subdivision, which exploits the coherence of consecutive frames in a graphic sequence. Anyway

existent methods don't exploit the aspects in common with cache memory management; therefore they organize the image cache without using an optimal strategy. An implementation with variable factors and strategies is presented along with the results from the comparison of their performances.

Image-Based Crowd Rendering, F.Tecchia, C.Loscos, Y.Chrysanthou, IEEE Computer Graphics and Applications, IEEE Computer Graphics and Applications Journal, March/April 2002 (Vol. 22, No. 2)
Populated urban environments are important in many applications such as urban planning and entertainment. However, rendering in real time many people in a complex environment is still challenging. In this article, methods for rendering real-time animated crowds in virtual cities are proposed. Taking advantage of the properties of the urban environment, and the way a viewer and the avatars move within it, a fast rendering is produced, based on positional and directional discretization. To allow the display of a large number of different individual people at interactive frame rates, texture compression with multi-pass rendering is combined.

Visualizing Crowds in Real-Time, F.Tecchia, C.Loscos, Y.Chrysanthou, Computer Graphics Forum Volume 21, Issue 4 (November 2002)
Real-time crowd visualization has recently attracted quite an interest from the graphics community and, as interactive applications become even more complex, there is a natural demand for new and unexplored application scenarios. In this paper methods are presented to deal with various aspects of crowd visualization, ranging from collision detection and behaviour modeling to fast rendering with shadows and quality shading. These methods make extensive use of current graphics hardware capabilities with the aim of providing scalability without compromising run-time speed. Results from a system employing these techniques seem to suggest that simulations of reasonably complex environments populated with thousands of animated characters are possible in real-time.

2 Commented references in expressive gestures

Argyle M. (1980), "Bodily Communication", Methuen & Co Ltd, London.
This book concerns general aspects on non-verbal communications. It analyses the role of several features (e.g., movement, postures, gaze, etc.) and their importance in non-verbal communication.

Camurri A., Mazzarino B., Ricchetti M., Timmers R., Volpe G. (2004a), "Multimodal analysis of expressive gesture in music and dance performances", in A. Camurri, G. Volpe (Eds.), Gesture-based Communication in Human-Computer Interaction, LNAI 2915, Springer Verlag.
This paper gives a preliminary definition of expressive gesture. It also discusses two experiments on specific aspects of expressive gesture processing: an experiment on the role of expressive gesture in human full-body movement in dance performances and an experiment on the role of expressive gesture for engaging listeners of a musical piece.

Camurri A., Mazzarino B., Volpe G. (2004b), "Expressive interfaces", Cognition, Technology & Work, 6(1): 15-22, Springer-Verlag.
This paper discusses the role of expressive gesture in designing interfaces for human-computer interaction. It particularly focuses on performing arts applications. Approaches for analysis of expressive gesture (e.g., the microdance approach and the subtractive approach) are discussed and some examples of applications are given.

Camurri A., Mazzarino B., Menocci S., Rocca E., Vallone I., Volpe G. (2004c), "Expressive gesture and multimodal interactive systems", in Proc. AISB 2004 Convention: Motion, Emotion and Cognition, Leeds, UK.
This paper gives a quite broad overview of research on expressive gesture with a particular reference on human movement and interaction strategies for mapping extracted expressive features onto real-time generation of visual and audio content.

Camurri A., De Poli G., Friberg A., Leman M., Volpe G. (2004d), "The MEGA project: analysis and synthesis of multisensory expressive gesture in performing art application", submitted.
This paper provides a survey of the research activities carried out in the framework of the EU-IST MEGA project. It includes experiments on expressive gesture in movement (dance performance) and sound (music performances). It also discusses the methodological approach and the outcomes of the project.

Camurri A., Coletta P., Massari A., Mazzarino B., Peri M., Ricchetti M., Ricci A., Volpe G. (2004e) "Toward real-time multimodal processing: EyesWeb 4.0", in Proc. AISB 2004 Convention: Motion, Emotion and Cognition, Leeds.
This paper presents the upcoming new EyesWeb platform (version 4.0). It discusses the deep updating of requirements and the research issues that led to such updates. It also illustrates the new features included in the new platform.

Camurri A., Lagerlöf I., Volpe G. (2003a) "Recognizing Emotion from Dance Movement: Comparison of Spectator Recognition and Automated Techniques", International Journal of Human-Computer Studies, 59(1-2): 213-225, Elsevier Science.
This paper describes an experiment on the role of expressive gesture in communicating basic emotion in dance performances. In particular, it describes some of the expressive cues extracted from dance fragments and how they were used to perform a first classification of dance fragments in terms of basic emotions.

Camurri A., Mazzarino B., Volpe G., Morasso P., Priano F., Re C. (2003b) "Application of multimedia techniques in the physical rehabilitation of Parkinson's patients", Journal of Visualization and Computer Animation, 14(5): 269-278, Wiley.
This paper illustrates the application of research on expressive gesture to therapy and rehabilitation. Some multimodal interactive systems were developed to analyze gestures of Parkinson's patients and to provide them an aesthetically resonant feedback.

Camurri A., Hashimoto S., Ricchetti M., Trocca R., Suzuki K., Volpe G. (2000) "EyesWeb – Toward Gesture and Affect Recognition in Interactive Dance and Music Systems" Computer Music Journal, 24(1): 57-69, MIT Press.
This paper presents the first publicly available version of the EyesWeb platform and the research problem that led to its development. It also illustrates some applications (mainly in the performing arts scenario) in which the EyesWeb platform was employed.

Cowie R., Douglas-Cowie E., Tsapatsoulis N., Votsis G., Kollias S., Fellenz W., Taylor J. (2001), Emotion Recognition in Human-Computer Interaction. IEEE Signal Processing Magazine, 1.
This very comprehensive paper covers many aspects on current and past research on emotions and expressiveness. Models, algorithms, and examples are provided.

Dahl S., Friberg A. (2004) "Expressiveness of musician's body movements in performances on marimba" in A. Camurri, G. Volpe (Eds.), Gesture-based Communication in Human-Computer Interaction, LNAI 2915, Springer Verlag.
This paper describes an experiment on the role of expressive gesture in communicating basic emotions through the movement of a music performer, in particular a marimba player. The expressive cues extracted from players' movement are described. The experiment and its results are discussed.

Hashimoto S. (1997), "KANSEI as the Third Target of Information Processing and Related Topics in Japan", in Camurri A. (Ed.) "Proceedings of the International Workshop on KANSEI: The technology of emotion", AIMI (Italian Computer Music Association) and DIST-University of Genova, 101-104.
This paper describes the Japanese approach to KANSEI Information Processing. KANSEI is a Japanese word that hardly finds a correspondence in Western languages. It refers to feeling, emotions, affect,

mood, sense. KANSEI Information Processing is characterized by a holistic approach considering expressive communication as global complex coding-decoding system.

Kurtenbach G., Hulteen E. (1990), "Gestures in Human Computer Communication", in Brenda Laurel (Ed.) The Art and Science of Interface Design, Addison-Wesley, 309-317.
This paper gives a definition of gesture in the context of human-computer interaction. Gesture is defined as a movement of the body conveying information. The definition of expressive gesture is partially derived from that definition.

Picard R. (1997), "Affective Computing", Cambridge, MA, MIT Press
This book describes the approach of the MIT Media Lab Affective Computing group to research on emotions and affect in computer systems. The book includes both theoretical aspects of affective computing and examples of applications in several fields.

Pollick F.E. (2004), "The Features People Use to Recognize Human Movement Style", in A. Camurri, G. Volpe (Eds.), Gesture-based Communication in Human-Computer Interaction, LNAI 2915, Springer Verlag.
This paper presents a survey of psychological studies involving human movement and gestures. Attention is focused on the features that seem to be more relevant for human perception of movement. Experiments are also discussed and examples (and further references) are given.

Laban R., Lawrence F.C. (1947), "Effort", Macdonald&Evans Ltd., London.
Laban R. (1963), "Modern Educational Dance", Macdonald & Evans Ltd., London.
These two books present Rudolf Laban's Theory of Effort, a theory describing human motion in terms of basic efforts and their qualities. This theory has a high relevance in dance and choreography (further studies have been carried out after Laban's death), and it can be a useful starting point for the development of models and algorithms for the analysis of high-level expressive qualities of movement.

Rowe R. (2001), "Machine Musicianship", Cambridge MA: MIT Press.
Rowe R. (1993), "Interactive music systems: Machine listening and composition", Cambridge MA: MIT Press.
These two books deal with design and development of interactive multimedia systems with a particular focus on music applications. They encompass several aspects: real-time analysis of music performance, strategies for mapping analysis data into generation and/or processing of audio content in different contexts, from variations on predefined scores to improvisation.

Schaeffer P. (1977), "Traité des Objets Musicaux", 2nd Edition, Paris, Editions du Seuil.
This book discusses Shaeffer's Morphology of sounding objects (Objets Musicaux). It can be useful for investigating analogies among movement and music in the communication process.

Wallbott H.G. (1980), "The measurement of Human Expressions", in Walbunga von Rallfer-Engel (Ed.) Aspects of communications, 203-228.
This paper presents a collection of motion features that are deemed important for communication of emotional, affective content in human movement. After reviewing a collection of works concerning movement features related with expressiveness and techniques to extract them (either manually or automatically), the paper classifies these features by considering six different aspects: spatial aspects, temporal aspects, spatio-temporal aspects, aspects related to "force" of a movement, "gestalt.

3 Interceptive actions Commented References

Bootsma, R. J., Houbiers, M. H. J., Whiting, H. T. A., & van Wieringen, P. C. W. (1991). Acquiring an attacking forehand drive: the effects of static and dynamic environmental conditions. *Research Quarterly for Exercice and Sport, 62,* 276-284.

This experiment was designed for two purposes: (a) to obtain a description of the observed changes in the movement patterns during the acquisition of the attacking forehand drive in table tennis. This changes support the notion of a funnellike type of control. (b) to evaluate the importance of the availability of time-to-contact information in the form of a dilating image generated by an approaching ball.

Bootsma, R. J., & van Wieringen, P. C. W. (1990). Timing an attacking forehand drive in table tennis. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 21-29.
This experiment showed that the players did not fully rely on a consistent movement production strategy because of a higher temporal accuracy at the moment of ball/bat contact than at initiation. It is argued that task constraints provide the organizing principles for perception and action at the same time, thereby establishing a mutual dependency between the two.

Lee, D. N. (1976). A theory of visual control of braking based on information about time-to-collision. *Perception, 5,* 437-459.
A theory of how a driver might visually control his braking is presented in this article. A mathematical analysis of the changing optic array at the driver's eye indicates that the simplest type of visual information is information about time-to-collision, rather than information about distance, speed or acceleration/deceleration. This information would be sufficient for controlling braking and would also be likely to be easily picked up by the driver.

Montagne, G., Fraisse, F., Ripoll, H., & Laurent, M. (2000). Perception-action coupling in an interceptive task : First-order time-to-contact as an input variable. *Human Movement Science, 19,* 59-72.
The required velocity model (Peper & al., 1994) is tested in this study by manipulating the first-order time-to-contact (TC1). The results show a significant influence of TC1 on the kinematic adaptation of the participant's movements, whereas changes in the distance travelled or in the travel speed has no effects on their own. It is the relationship between the two, namely the TC1, that is relevant for movement control.

Montagne, G., Laurent, M., Durey, A., & Bootsma, R. J. (1999). Movement reversals in ball catching. *Experimental brain research, 129,* 87-92.
The aim of this study is to test the required velocity model (Peper & al., 1994). This study shows kinematic adaptation of the movement when the current lateral distance is manipulated. The authors show that the current velocity is equal to the required velocity, 300 ms before contact, and remains so until contact. They show the existence of reversal points (back and forth movements of the hand along the axis) under certain experimental conditions, because of the 'creation' of a current lateral distance.

Peper, C. E., Bootsma, R. J., Mestre, D. R., & Bakker, F. C. (1994). Catching balls: How to get the hand to the right place at the right time. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 591-612.
The results demonstrate that in catching a ball, humans do not first predict where it can be caught and then move the hand to this position (no prediction), but their hand's velocity is continuously geared to information specifying the currently required velocity. It guarantees that the hand will be at the right place in the right time. Instead of spatiotemporal estimates, continuous action-related information is required in order to control one's actions.

Savelsbergh, G. J. P., Whiting, H. T. A., & Bootsma, R. J. (1991). 'Grasping' tau. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 315-322.
This study deals with the timing of the grasp movements involved in catching a ball by manipulating the optical expansion pattern. Adjustments to the aperture of the hand to the different ball sizes, especially to the deflating ball, point to a finely attuned perception-action coupling. Moreover, this study confirm that information occuring in the last 200 ms of ball flight before contact is indeed used to tune the catching action.

Tresilian, J. R. (1995). Perceptual and cognitive processes in time-to-contact estimation : Analysis of prediction motion and relative judgment tasks. *Perception and psychophysics, 57,* 231-245.
The authors examine important differences between three classes of tasks (that appear to involve time-to-contact (TTC) information: Coincidence anticipation (CA) tasks, relative-judgement (RJ) tasks and interceptive actions (Ias)), and inquire wether results from PM (prediction-motion) and RJ tasks can be generalized to Ias. They propose a revised version of the tau hypothesis as an account of the perceptual information processing involved in the control of fast Ias.

Brouwer, A. M., Brenner, E., & Smeets, J. B. J. (2001). Perception of acceleration with short presentation times: Can acceleration be used in interception? *Experimental brain research, 133,* 242-248.
To investigate whether acceleration can be used in interception tasks, the authors determined how well subjects could detect acceleration at short presentation times in two different tasks. Most subjects could do these tasks even when the presentation time was only 300 ms. About 25% change of the average velocity was needed to detect acceleration with reasonable confidence. The authors conclude that acceleration is not used to initiate locomotion in catching balls.

Brenner, E., Smeets, J. B. J., & Lussanet, M. H. E. (1998). Hitting moving objects: Continuous control of the acceleration of the hand on the basis of the target's velocity. *Experimental brain research, 122,* 467-474.
After having showed that the direction in wich the actor move his hand is continuously adjusted on the basis of the target's perceived position, with a delay of about 110 ms, the authors conclude that the acceleration of the hand is continuously adjusted on the basis of the speed of the target, with a delay of about 200 ms, in a task in which the actor had to hit moving targets as quickly as possible with a rod.

Carlton, L. G. (1992). Visual processing time and the control of movement. *Vision and control* (pp. 3-29).
This article is concerned with visual processing delays associated with the control of ongoing movements, like aimed or pointing movements of the hand and arm. Firstly, it provides a review of some of the classic studies examining manual aiming and visual processing time. Secondly, it provides work suggesting that visual feedback information may be used in aimed movements with latencies much shorter than previously supposed. Recent evidence is then presented supporting rapid visual processing when visual stimuli in the environment are changed. The last section provides a synthesis of research on visual processing time for the control of movement.

McLeod, P. (1987). Visual reaction time and high-speed ball games. *Perception, 16,* 49-59.
This study shows that, when batting, highly skilled professional cricketers show reaction times of around 200 ms, times similar to those found in traditional laboratory studies. It is suggested tha the contrast between the ability of skilled and unskilled sportsmen to act on the basis of visual information does not lie in differences in the speed of operation of the perceptual system, but in the organisation of the motor system.

Michaels, C., & Beek, P. (1995). The state of ecological psychology. *Ecological psychology, 7(4),* 259-278.
In this article the authors discuss the state of ecological psychology, dealing with the coordination of activity with respect to perceptual information. They identify and evaluate three approaches: direct perception, kinetic theory and pattern dynamics. After a brief summary of these approaches, they elaborated the distinctions among these approaches by examining their merits and limitations with regard to four problems areas.

Montagne, G., & Laurent, M. (1996). Le support informationnel des tâches d'interception. *Science et motricité, n° 28, pp. 3-11.*
The interception of a mobile presupposes the perception of the physical caracteristics of its trajectory. The perceptive activity requires the use of sources of visual information. This article highlights the

diversity of the sources of available visual information and the mechanisms underlying the process of visual perception.

Montagne, G. (2002). Le contrôle des actions finalisées. Une approche comportementale. *Habilitation à diriger des recherches. Spécialité staps.*
The first part of this document is intended to briefly present the conceptual framework in which various work was completed. The synthesis of work is then presented there, leading to a conception of perceptive organisation and a conception of the integration modes of information for the control of the movement.

McLeod, P., & Dienes, Z. (1996). Do fielders know where to go to catch the ball or only how to get there? *Journal of experimental psychology: Human perception and performance, vol 22. n° 3. 531-543.*
Skilled fielders were filmed as they ran backward or forward to catch balls projected toward them. They ran at a speed that kept the acceleration of the tangent of the angle of elevation of gaze to the ball. This algorithm does not tell them where or when the ball will land, but it ensures that they run though the place where the ball drops to catch height at the precise moment that the ball arrives there.

Benguigui, N., Ripoll, H., & Broderick, M. P. (2003). Time-to-contact estimation of accelerated stimuli is based on first-order information. *Journal of Experimental Psychology: Human perception and Performance., 29,* 6, 1083–1101.
The goal of this study was to show that information about acceleration is not used (1) to estimate the time-to-contact (TC) of an accelerated stimulus after the occlusion of a final part of its trajectory and (2) to indirectly intercept an accelerated stimulus with a thrown projectile. In the course of this study, it was possible to show that the participants produced their temporal estimates –in a judgment arrival task– and produced their actions –in an indirect interceptive task– on the basis of TC1 information.

Tresilian, J. R. (1999b). Visually timed action: time-out for 'tau'? *Trends in Cognitive Sciences, 3,* 301-310.
This article reviews recent work that shows conclusively that the hypothesis which proposes the variable tau as the informational basis for TTC estimation is false. It describes an alternative approach showing that the information used in judging time-to-collision is task- and situation-dependant, is of many different origins (of which tau is just one) and is influenced by the information-processing constraints of the nervous system.

Day, B. L., & Lyon, I. N. (2000). Voluntary modification of automatic arm movements evoked by motion of visual target. *Experimental Brain Research, 130:* 1596168.
The authors have investigated whether the processes underlying the visually evoked, automatic adjustments to a reach are: (1) modifiable by the subject's intention, and (2) available to initiate movement of a stationary arm. They proposed two possible mechanisms: the first is a dual-pathway model resting on separate visuo-motor processes with different properties; the second model resting on a single visuo-motor mechanism that is under the control of a higher, attentional process.

Port, N. L., Lee, D., Dassonville, P., & Georgopoulos, A. P. (1997). Manual interception of moving targets. I. Performance and movement initiation. *Experimental Brain Research, 116,* 406-420.
The authors investigated the capacities of human subjects to intercept moving targets in a 2D space. Three models of movement initiation were investigated: the threshold-distance model; the threshold-tau model; a dual-strategy model was developed which allowed for the adoption of either of the two strategies for movement initiation: namely, a strategy based on the threshold-distance model ("reactive" strategie) and another based on the threshold-tau model ("predictive" strategy). In fact, individual subjects preferred to use one or the auther strategy.

Lee, D., Port, N. L., & Georgopoulos, A. P. (1997). Manual interception of moving targets. II. On-line control of overlapping submovements. *Experimental Brain Research, 116,* 421-433.

The authors studied the kinematic characteristics of arm movements and their relation to a stimulus moving with a wide range of velocity and acceleration. They suggest a control mechanism that produces a series of discrete submovements. Their results were consistent with the hypothesis that the end-point of each submovement is linearly related to the target location estimated from the position and velocity of the target at the submovement onset.

4 Commented references in temporal delay in action-vision loop

Adelstein, Bernard D., Thomas, G. L., Ellis, Stephen R. ***(2003). Head tracking latency in virtual environments: psychophysics and a model,*** **Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting.**

The authors of this paper presents some experiments that tend to quantify the human perception of latency, and to describe the mechanism by which the latency is perceived. The field of investigation is the immersive VE, in which the user is equipped with a head-mounted display for stereo rendering of scenes, and a head tracking system in order to display the scene depending of the position of the user's head.

One of the main ideas presented here, is that the perception of delays might be due in part by "image slip", which is the apparent movement of the image as a consequence of head tracking latency, or as defined by Adelstein & al "the virtual scene's artifactual concomitant motion with the observer's head resulting from time lag".

Subjects are asked to move their heads "smoothly and sinusoidally" in a yaw movement from side-to-side. Latency conditions are presented in sequential pairs composed of a reference and a probe level, which is made of the reference level latency plus an added latency. Reference levels of latency are either 33, 100 or 200ms. Subjects have to decide if the intervals of the reference and of the probe level differ or not.

The methods used are a combination of the staircase method (Method of Limits) with randomized series. The results are discussed for *added latency*, and show that the average Just Noticeable Difference is about 17ms, and the mean Point of Subjective Equality is about 50ms. At last, JND and PSE results are not significantly varying depending on the reference level of latency.

Crowley, James L., Coutaz, Joëlle, Bérard, François (2000). Things That See, *Communications of the ACM 43, 3, pp 54-64, ACM Press New York, NY, USA.*

This paper focuses on the techniques used in HCI that bring together human gesture and machine vision, which is defined as the observation of an environment using cameras. The main advantage of machine vision for HCI is that the user doesn't need any tool; this means that human actions can be tracked without being constrained.
In that kind of interaction environment, this article particularly specifies that "the latency of machine perception must be less than 50ms for direct manipulation using finger tracking".

Cunningham, Douglas W., Chatziastros, Astros, von der Heyde, Markus, Bülthoff, Heinrich H. (2000). Temporal Adaptation and the role of temporal contiguity in spatial behavior, *Technical Report No. 85,* MPI, December 2000

It has now long been established that humans are sensitive to spatial misalignment relationships between sensorialities (study of *prism adaptation*), but that after a few minutes of recalibration, this misalignment does not impair performance anymore, nor is even felt as disturbing. This kind of adjustment is known as Spatial Adaptation.
This paper addresses the similar problem of alignment between sensorialities in the temporal dimension. Although no compensation effects were noticed for temporal delays between the different sensorialities, a previous experiment has shown an adaptation to intersensory temporal discrepancy after long learning (Cunningham et al. 2001), and the experiments presented some adaptation to temporal delays. As in the case of spatial misalignment, an aftereffect is noticed in the two experiments

presented here: learning to perform the task with temporal delay greatly reduces the performance with no time delay afterwards.
At last, the second experiment is processed in order to reduce the simulator sickness felt by some subjects. One subject still experienced simulator sickness, but the decrease of the proportion of affected subjects might show that temporal delay is not an increasing factor of simulator sickness, which could only be due to a conflict between the visual and vestibular perception of acceleration.

Cunningham, Douglas W., Biillock, Vincent A., Tsou, Brian H. (2001). Sensorimotor Adaptation to Violations of Temporal contiguity, *Psychological Science,* Volume 12: Issue 6
The perception of one event that is addressed to different sensorialities is unique, despite the fact that the different sensorialities are processed at different speeds, and that neurons may respond with different latencies to an identical stimuli. This suggests that the brain is able to process the different modalities by compensating the time variations among the stimuli.

The experiment proposed here shows that in a guidance task, where a displayed airplane is controlled by mouse movements (no force feedback), the subjects perform adaptation to an added time delay (235ms) and performed the task as well in presence of such a delay compared to the task with no delay. The time delay was at first said to be disturbing and subjects complained about it. However, it is shown that training to time delays has strong negative aftereffects: after training, the subjects are no more able to perform the task with an unnoticeable delay (35ms).

Ellis, Stephen R., Mania, Katerina, Adelstein, Bernard D. and Hill, Micheal, (2004). Generalizeability of latency detection in a variety of virtual environments, *Human Factors and Ergonomics Society, 48th Annual Meeting, New Orleans, USA (to appear)*
This study deals with the discrimination of latency changes in VE during head movement.
Before the experiment, subjects are carefully instructed about the effects of latency on the display of the visual scene. The equipment is composed of a head-mounted display and a tracking system for head and hand movements. In the three experiments presented, subjects are asked to detect latencies introduced in the display of a virtual scene. The virtual scene is composed either of a single object displayed in foreground with a black background (i.e. no background), either of a single background without foreground, or both a foreground object and a background. Subjects are then asked to compare a reference situation, which present a base latency of 10.4ms, to a test situation, in which supplementary steps of 8.5ms delay are added to the reference latency.
The results confirm previous studies conducted with somewhat different psychophysical techniques and very different viewing techniques: the JND is about 10-15ms, and PSE is about 30ms. These thresholds are a bit smaller than those presented in similar experiments, and authors expect this to be due to the fact that subjects were instructed about effects of delay. Therefore, the authors recommend designers of VE environments to avoid important variations of latencies (below 16ms), even if the base latency is important.
Furthermore, latency discrimination is shown not to be depending on the complexity of the environment (i.e. the combination or existence of foreground and background). In a further experiment (to be published), the authors show that latency discrimination is not depending on the complexity of the displayed scene.

Ellis, Stephen R., Young, Mark J. and Adelstein, Bernard D. (1999). Discrimination of changes in latency during head movement, Proceedings of Computer Human Interfaces, pp. 1129-1133
When the latency exceeds 300ms in VE, the user is said to adopt a “move and wait” strategy, but the thresholds for first effects of short latencies remain imprecisely known. This paper focuses on the discrimination of latency changes in VE during head movement.
In the experiment presented here, subjects are asked to compare a reference situation, which presents base latencies of are 27, 97 or 196 ms, to a test situation, in which a supplementary delay is added to the reference latency.
The results show that the discrimination of latency does not follow the Weber’s law: JND seems indeed to be roughly independent of the base latency introduced. Therefore, the authors recommend designers of VE environments to avoid important variations of latencies, even if the base latency is important.

Richard, Paul, Birebent, Georges, Coiffet, Philippe, Burdea, Grigore, Gomez, Daniel, Langrana, Noshir (1996). Effect of Frame Rate and Force feedback on Virtual Object Manipulation *Presence, Vol 5, No. 1, 95-108.*

This paper deals with the effects of frame rate (graphics refresh rate) and viewing mode on the user performance in Immersive Virtual Environment.

Two experimental studies are done. In the first one, users are asked to grasp as quick as possible a virtual ball, which is moving relatively fast in random directions. In the second one, the subjects are asked to reach and pick up a virtual ball along a determined path (which is materialized by the vertical projection of the ball between two lines on the ground), while applying a low but stable amount of deformation on the ball, and achieving the task in less than 15 sec.

Results show that:

- High frame rates and stereoscopic view require less adaptation from the user, since these conditions are said to be closer to the natural conditions in which a human interacts with its environment.
- Completion times and efficiency are worse as the frame rate falls below 14 fps.
- Redundant information (for instance information about force feedback applied to the user on visual or auditory modalities) improve the completion time and reduce the task error, especially in the case of redundant information by auditory cues.

Stanney, Kay M., Mourant, Ronald R., Kennedy, Robert S., (1998). Human Factors Issues in Virtual Environments: A Review of the Literature, *Presence, Vol 4, No. 4, pp. 327-351*

While the number of studies and applications using techniques coming from the Virtual Realities is still growing up, the knowledge about human factors related to these new uses of technology remains quite poor. This paper thus provides an overview of many of the human-factor issues related to VR. Starting with the question of efficiency and performance of applications using VR techniques, the authors present the current issues in increasing the degree of presence and effectiveness of VEs. These factors are then related to tasks characteristics and user characteristics (including a short review of human sensory capabilities in the visual, auditory and haptic fields). The two last sections of this paper deal with health and safety issues in VE (including review of knowledge about cybersickness effects and diseases or injuries by displays) and social implications.

Ware, Colin, and Balakrishnan, Ravin, (1994). Reaching for Objects in VR Displays: Lag and Frame Rate, *ACM Transactins on Computer-Human Interaction, Vol 1, No 4, pp. 331-356.*

This paper presents three experiments verifying the previous results about human perception of time delays. According to the law of Fitt, the mean time necessary for a reaching task is depending on the size of the target, and psychophysical constants depending on the motor and cognitive capacities of the subject. The authors propose slight modifications of this law in order to take into account the delays introduced by the data processing and display time in VE.

L. James Smart, Jr., Miami University, Oxford, Ohio, Thomas A. Stoffregen, University of Minnesota, Minneapolis, Minnesota, and Benoît G. Bardy, University of Paris Sud XI, Orsay, France: Visually Induced Motion Sickness Predicted by Postural Instability, *HUMAN FACTORS*, Vol. 44, No. 3, Fall 2002.

Studies performed in the 10 past years report that motion sickness may be correlated with the perception of vection, which is defined by the authors as the *experience of self-motion relative to the inertial environment*.

The present study tries to provide qualitative and quantitative measures in order to predict motion sickness. In the experiment, standing up subjects are placed in a moving room, which exposes them to a visual stimulus, and which one is supposed to induce the feeling of vection to the participants. The results show that the subjects who more experienced motion sickness more often reported the feeling of motion. Furthermore, it is shown that postural instability precedes motion sickness.

Watson, B., Spaulding, V., Walker, N., Ribarsky, W. (1997) Evaluation of the Effects of Frame Time Variation on VR Task Performance, *VRAIS '97, IEEE Virtual Reality Annual Symposium, 38-44*
The presented study tries to show the effects of lag fluctuation over time in VE.
The proposed task is the combination of two ones: (1) the grab of a moving target (open loop task, i.e. movements that do not allow feedback or correction), and (2) the placement of the grabbed target on a pedestal (closed loop task, i.e. movements that are continuously corrected with the help of feedback information in order to reach a goal). In the proposed task, the frame delay continuously changes during the movement according to a sinusoidal law of variation.

MacKenzie, I.S., Ware, C. (1993). Lag as determinant of human performance in interactive systems. *Proceedings of the ACM Conference on Human Factors in Computing Systems – INTERCHI'93*, 488-493. New-York:ACM.
The sources of lag (the delay between input action and output response) and its effects on human performance are discussed. The authors measured the effects of time delay in a study of target acquisition using the Fitts' law paradigm with the addition of four lag conditions. The results show that lag is xxx on movement as its value becomes greater or equal to 75ms.

5 Commented references in motion capture devices

[VICON] The vicon web site: http://www.vicon.com
Vicon Ltd is a company that provides optical mocap solutions both in hardware and software. They not only sell all the equipments for doing mocap (cameras, data stations, etc) but also software for optimizing the use of their systems (the software they sell ranges from simple workstation for capturing the data to complex kinematics estimators for efficiently post-processing the data or plugin-like products for gait estimation)
[MOTIONANALYSIS] The Motion Analysis web site: http://www.motionanalysis.com
Motion Analysis corp. is a company that claims to be the leader in optical mocap. As Vicon, they offer both software and hardware products that cover all the width of the motion capture industry, from entertainment to military application.
[VICONIQ] The IQ technical sheet: http://www.vicon.com/proddetail.jsp?id=164
IQ is one of the latest innovations from Vicon. This is a software that offers powerful post-processing and motion editing tools for efficiently and accurately treating mocap data.
[ASCENSION] The Ascension technologies corp. web site: http://www.ascension-tech.com
Ascension is one of the leaders in magnetic mocap and offers a wide range of hardware products. They not only manufacture magnetic trackers, but also eye trackers and other by-products, but always in the motion capture area.
[POLHEMUS] The Polhemus web site: http://www.polhemus.com
Polhemus is another company that manufactures and sells magnetic motion capture hardware products. They offer a comparable range of products as Ascension, but they also make the FastSCAN which is a 3D scanner for acquiring 3D models from real objects.
[IMMERSION] The immersion corp. web site: http://www.immersion.com
Immersion corp. is a haptic devices oriented company. Thye make well known products such as the CyberGlove, plus many other ones as the CyberTouch and the CyberGrasp. They also develop and sell mocap oriented softwares such as the "VirtualHand for MOCAP"
[METAMOTION] The Meta Motion corp. web site: http://www.metamotion.com
MetaMotion is a company that manufactures and sells affordable (comparatively to other systems) motion mechanical motion capture systems, plus other ones such as eye tracker and dataGloves for capturing hand movements.

[SHIN] Hyun Joon Shin, Lucas Kovar, Michael Gleicher, *Physical Touch-up of human motions*, 11th Pacific conference on computer graphics and applications, 2003
This paper addresses the issues of physics while manipulating an existing clip of animation of a character. It proposes a solution for editing and modifying existing clips of motion while preserving their physical correctness.

[CHOI] Kwang-Jin Choi; Hyeong-Seok Ko. *Online motion retargeting*, The Journal of Visualization and Computer Animation 11(5):223-235, 2000
This paper presents a method for adapting existing sequences of human motion onto skeletons of different topology in real-time.

[POPO] Zoran Popovic and Andrew Witkin. *Physically Based motion transformation*, proceedings of SIGGRAPH 99, 1999.
This paper proposes a solution for preserving the physical correctness of the motion while retargeting it onto another character.

[ALEX]: Alexander, E.J., Bregler C., Andriacchi, T.P.: *Nonrigid Modeling of Body Segments for Improved Skeletal Motion Estimation*, Computer Modeling in Engineering and Science, Vol 4, Number 3 & 4, pp. 351-364, 2003.
This paper uses deformable bodies for modeling the segments of a human body, in order to improve the quality of the estimated motion.

[CHU]: Chi-Wei Chu, Odest Chadwicke Jenkins, Maja J Matari´c, *Markerless Kinematic Model and Motion Capture from Volume Sequences,* Proceedings of IEEE computer vision and pattern recognition, Madison, Wisconsin, USA, June 16-22, 2003
This paper estimates a series of reconstructed volumes from video sequences and then estimates an overall shape plus a skeletal hierarchy that is then reanimated from the volumes reconstructions.

[LUCK]: Jason Luck, Dan Small, Charles Q. Little, *Real-Time Tracking of Articulated Human Models Using a 3D Shape-from-Silhouette Method*, Proceedings of the International Workshop RobVis 2001, Auckland, New Zealand, February 2001.
This paper describes a method for reconstructing and animating a full human body in real time.

6 Commented references in 3D sounds

[Allen1979]
J.B. Allen and D.A. Berkeley: Image method for efficiently simulating small-room acoustics. *Journal of Acoustics Society of America,* vol.65, pp.943-950, 1979
Sound propagates through several reflection, transmission and diffraction paths. This paper describes how to simulate your auditorium design using Image Source method. Visualization is performed similarly as mirror image sources are visualized in the image-source method.

[Begault1994]
D.R. Begault: *3D sound for virtual reality and multimedia*. Academic Press Professional, 1994
This book intends to be a reference in term of basic knowledge of simulation and rendering of sound in virtual environments. Fundamental elements for complete spatial audio system are overviewed: transmission, reflections, reverberation, Head Related Transfer Function (HRTF), diffraction, refraction….Knowledge contained in this book are still relevant for virtual reality but also for multi-channel "surround sound" formats as well.
Category: C3, pdf available at UNIGE

[Casier2003]
G. Casier, P. Plenacoste, C. Chaillou, B. Semail: The DigiHaptic, a new three degrees of freedom multifinger haptic device. *In Virtual Reality International Conference,* pp. 35, 2003

[Fisch2003]
A. Fisch, C. Mavroidis, J. Melli-Huber, Y. Bar-Cohen: Chapter 4: Haptic Devices for Virtual Reality, Telepresence, and Human-Assistive Robotics. *In Biologically-Inspired Intelligent Robots. SPIE Press,* 2003

This paper presents a state of the art of haptic technology and research. They describe here an overview of existent haptic devices and some applications used for medical purposes (Virtual Reality for surgical training), space-related purposes (Mars Pathfinder Mission's Sojourner Rover on the surface of Mars inspecting a boulder) or tele-robotics (example of Tele-Presence Control of Robonaut, NASA).
Category: C2, pdf available at UNIGE

[Fong2003]
T. Fong, C. Thorpe, B. Glass: PDADriver: A handheld system for remote driving. *In Proceedings of IEEE International Conference on Advanced Robotics, Coimbra,* 2003
This paper describes a Personal Digital Assistant (PDA) system for vehicle teleoperation.
Category: C2, pdf available at UNIGE

[Frisoli2002]
A. Frisoli, F. Simoncini, M. Bergamasco : Mechanical Design of a Haptic Interface for the Hand, *ASME International DETC- 27th Biennial Mechanisms and Robotics Conference,* 2002
This paper is a kinematic study of a haptic interface for the hand. The study of hand and mechanism kinematics allows a complete evaluation of the performance of the mechanism seeing the influence of hand movement on the mechanism.
Category: C2, pdf available at UNIGE

[Funkhouser2002]
T. Funkhouser, JM. Jot and N. Tsingos: "Sounds good to me" computational sound for graphics, virtual reality and interactive systems. *In SIGGRAPH 2002 Conference Proceedings,* 2002
This course can be used as a support for the simulation and rendering of sound in virtual environments. Some concepts, methods and models are overviewed and real-time methods for spatializing sounds are described.
Category: C2, C3, pdf available at UNIGE

[Funkhouser2004]
T. Funkhouser, N. Tsingos, I. Carlbom, G. Elko, M. Sondhi, J.E. West, G. Pingali, P. Min and A. Ngan: A beam tracing method for interactive architectural acoustics. *Journal of the Acoustical Society of America,* 2004
This paper describes a beam tracing method allowing computing propagation paths from sound sources to receivers fast enough to obtain interactive systems. They also allow the receiver to move around. The goal of this method is to obtain real-time auralization in large architectural environments.
Category: C2, pdf available at UNIGE

[Gardner1994]
B. Gardner and K. Martin: HRTF measurements of a KEMAR dummy-head microphone. *MIT Media Lab Perceptual Computing, technical report 280,* 1994
You can find on an html file a set of head-related transfer function (HRTF) measurements of a KEMAT dummy head microphone. The measurements represent the left and right ear impulse responses from the KEMAR. These data are available to the research community on the Internet and allow the researchers to simulate the surrounding sound thanks to headphones.
Category: C2, html file: http://sound.media.mit.edu/KEMAR.html

[Gomez1995]
D. Gomez, G. Burdea, N. Lagrana: Integration of the Rutgers Master II in a virtual reality simulation. *Proceedings of Virtual Reality Annual International Symposium '95, IEEE Computer Society Press,* pp. 198-202, 1995
This paper presents a new compact hand master device with force feedback. It allows the user to avoid wearing sensing gloves.

[Gregory2000]

A. Gregory, S. Ehmann, M.C. Lin: In-Touch:interactive multiresolution modeling and 3d painting with haptic interface. ***In Proceedings of IEEE International Conference on Virtual Reality 2000*, 2000**

The creation of a new intuitive haptic interface is presented here. This interface intends to help artists and designers to create and define a three-dimensional polygonal mesh, not only by the use of visual feedback but also by the use of sense of touch.

Category: C2, pdf available at UNIGE

[Huopaniemi1999]

J. Huopaniemi. Virtual acoustics and 3D sound in multimedia signal processing. *Thesis*, 1999.

During his thesis, Jyri Huopaniemi studied real-time modeling and synthesis of three-dimensional sound in the context of digital audio, multimedia and virtual environments. Three main domains have to be studied to obtain a complete spatial audio interface: the sound sources, the room acoustics and spatial hearing.

Category: C2, pdf available at UNIGE

[Kawasaki2003]

H. Kawasaki, J. Takai, Y. Tanaka, C. Mrad, T. Mouri: Control of multi-fingered haptic interface opposite to human hand. ***In Proceedings of the International Conference on Intelligent Robots and Systems*, 2003**

“This paper presents a control architecture of the developed multi-fingered haptic interface, named Gifu Haptic Interface.”

Category: C2, pdf available at UNIGE

[Kleiner1993]

M. Kleiner, B.-I. Dalenbäck and P. Svensson: Auralization – an overview. ***Journal of the Audio Engineering Society vol.41, pp.861-875*, 1993**

This Journal article is a good introduction to auralization and provides us a good reference list. Different approaches to model the early reflections are described here.

[Kulowski1984]

A. Kulowski: Algorithmic representation of the ray tracing technique. ***Applied. Acoustics, vol.18, pp.449-469*, 1984**

This paper presents the classical ray-tracing algorithm. This technique is often used for room acoustics modelling. The detection of sound ray emission from the source is an important problem we need to resolve.

[Mendoza2001]

C. Mendoza, C. Laugier: Realistic haptic rendering for highly deformable virtual objects. ***In Proceedings of IEEE International Conference on Virtual Reality 2001*, pp. 264–269, 2001**

This paper studies the methods which provide solutions “for stability problems arising from the difference between the sampling rate requirements for haptic devices (about 1 KHz) and the update rates of the physical objects being simulated (about 10 Hz)” but which fail when the object is deformable.

[Nguyen2001]

L.A. Nguyen, M. Bualat, L.J. Edwards, L. Flueckiger, C. Neveu, K. Schwehr, M.D Wagner, E. Zbinden: Virtual reality interfaces for visualization and control of remote vehicles. ***In Autonomous Robots*, Vol. 11, 2001**

Results of the Autonomy and Robotics Area (ARA) at the NASA Ames Research Center are presented here as well as the advantages and issues of using Virtual Reality interfaces. ARA intends to explore the use of autonomous robotic systems. VR interfaces are used to control complex robotic mechanism.

Category: C2, pdf available at UNIGE

[Savioja1996]

L. Savioja, J. Huopaniemi, L. Lokki and R. Vnnen: Virtual environment simulation – Advances in the DIVA project. ***Proceedings of ICAD’96*, 1996**

This paper presents the creation of a real-time virtual audio reality model called Digital Interactive Virtual Acoustics (DIVA). This system aims to simulate a virtual symphony orchestra. It includes sound rendering, room acoustics modelling, binaural auralization for headphone and loudspeaker listening. In order to improve the real-time computation of auralization, authors used Head-Related Transfer Function (HRTF), late reverberation and an image source method.
Category: C2, pdf available at UNIGE

[Snibbe2001]
S.S Snibbe, K.E. MacLean, R. Shaw, J. Roderick, W.L. Verplank, M. Schee®: Haptic techniques for media control. ***In Proceedings of the 14th annual ACM symposium on User interface software and technology, ACM Press*****, 2001**
This paper presents different techniques to use haptics in digital media. They demonstrate the power of haptic compared to the use of button and key presses.
Category: C2, pdf available at UNIGE

[Takala1992]
T. Takala, J. Hahn: Sound rendering. ***In SIGGRAPH 1992*****, 26(2):211-220, 1992.**
This paper presents a methodology to produce synchronized soundtracks for animations. Their approach consists in attaching specific sound to objects and using behavioral or physically-base motion control.
Category: C3, pdf available at UNIGE

[Tarrin2003]
N. Tarrin, S. Coquillart, S. Hasegawa, L. Bouguila, M. Sato: The stringed haptic workbench: a new haptic workbench solution. ***In Proceedings of Eurographics*****, 2003**
The use of an arm feedback device on one-screen workbrenches leads to drawbacks. This paper tries to find solution for these problems: perturbation of the stereoscopic display, cross virtual objects, hide parts of the visualization space. They decided to integrate a stringed haptic device on a workbrench.

[Williams1998]
H.R.L. Williams, D. Noorth, M. Murphy, J. Berlin, M. Krier: Kinesthetic Force/Moment Feedback via Active Exoskeleton. ***Proceedings of the Image Society Conference*****, 1998**
This paper describes a general control architecture "for real-time, sensor-based, rate-based, shared control of general telerobotic systems including force-reflecting hand controllers".
Category: C2, pdf available at UNIGE