	IST-2002-002114 – ENACTIVE NETWORK OF EXCELLENCE **WP4b** **STAR in Action and vision fusion**
Reference	EI_WP4b_STAR_04096_INPG
Title	***Action-Vision fusion in CG, HCI, VR, teleoperation, telecommunication*** Cornerstones and trends
Object	*D4b.1 : State of the art in technologies for fusing action and vision* **Deliverable D4b.1, due to 2004, 30th september**
Writer	Luciani Annie, WP4b leader
Participants	WP4b participants
Dissemination	WP4b participants
Nber of Pages	21 pages
Date of the version	2004, September 6th
Confidentiality	ENACTIVE Internal Restricted Use

Analytic Part
Annie Luciani
INPG
September 6th, 2004

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.

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.

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.

Note of the author:

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.

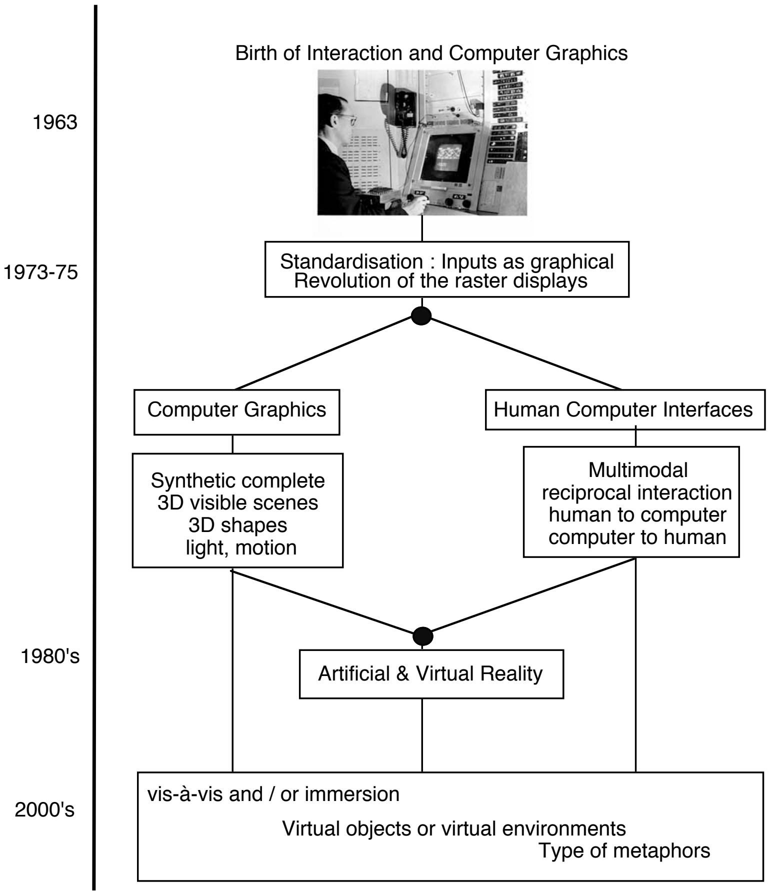
4. The second post-pioneer phase: From CG and HCI to VR

The new convergence between CG and HCI, represented in Figure 1, 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 [16], 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

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

meaning of AR or VR is often synonymous of Virtual Environments in the sense of worlds surrounding the user and being explored by him.



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 1 – The CG and HCI convergence

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

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].

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 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.

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,

[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)).

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.).

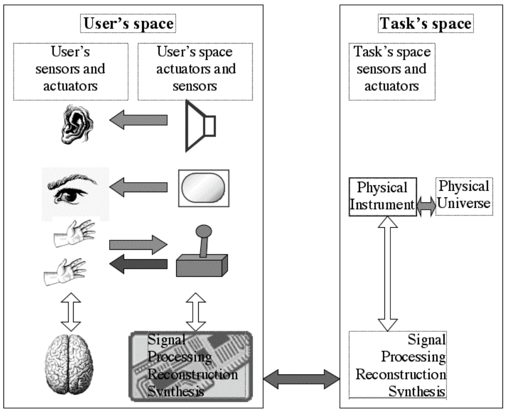

User's space
User's sensors and actuators
User's space actuators and sensors
Signal Processing Reconstruction Synthesis
Task's space
Task's space sensors and actuators
Physical Instrument
Physical Universe
Signal Processing Reconstruction Synthesis

Figure 2 - 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.

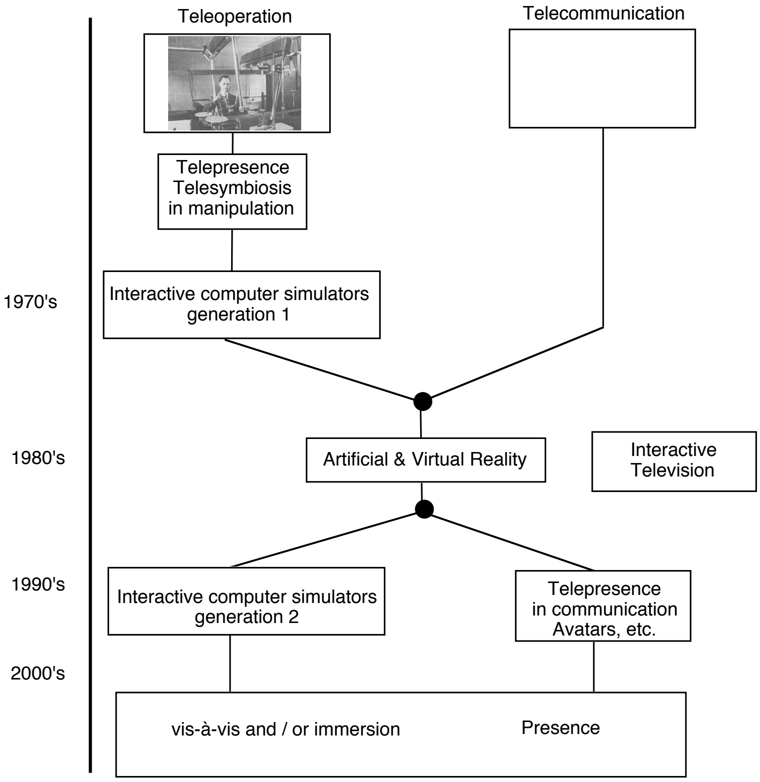

Teleoperation
Telecommunication
Telepresence Telesymbiosis in manipulation
1970's
Interactive computer simulators generation 1
1980's
Artificial & Virtual Reality
Interactive Television
1990's
Interactive computer simulators generation 2
Telepresence in communication Avatars, etc.
2000's
vis-à-vis and / or immersion
Presence

Figure 3 - 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 (Figure 3), both stamped by the question of Presence of one world in the distant (physically, sensorially, cognitively) other.

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].

7. 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.

8. 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.

8.1. Immersion and vis-à-vis : similarities and differences

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 tele-operation: 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.

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.

8.2. 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.

<u>Examples of instrumental situations:</u>
- 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.

<u>Examples of non-instrumental situations</u>
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.

9. 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?

10. 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.

11. 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.

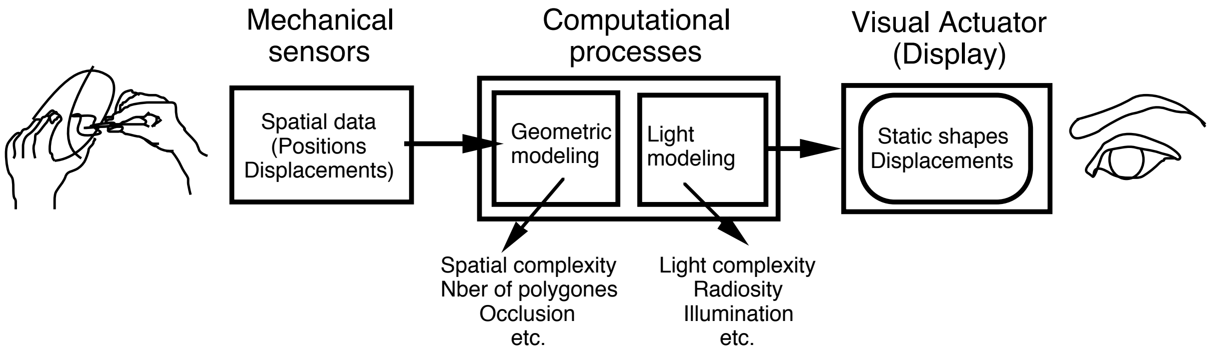


Figure 4. The hand-vision chain in conventional interaction

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).

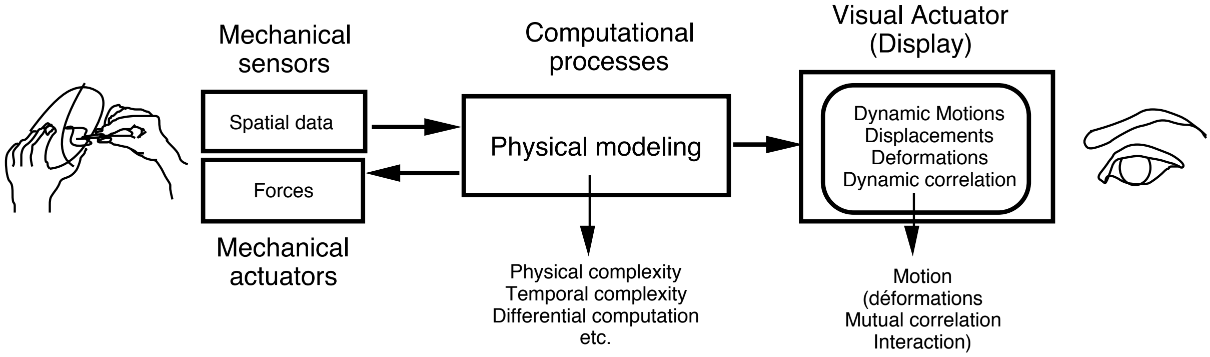


Figure 5. The manipulatory relationship between action and vision during instrumental interaction

12. 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.

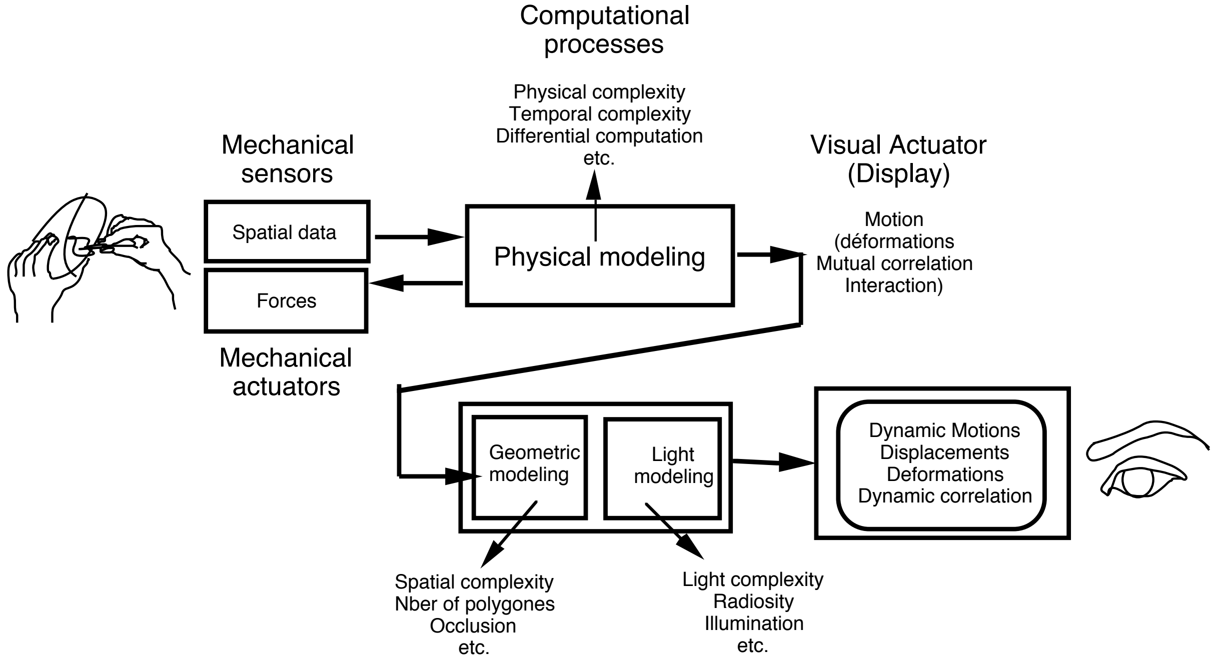


Figure 6. The complete action-vision chain

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).

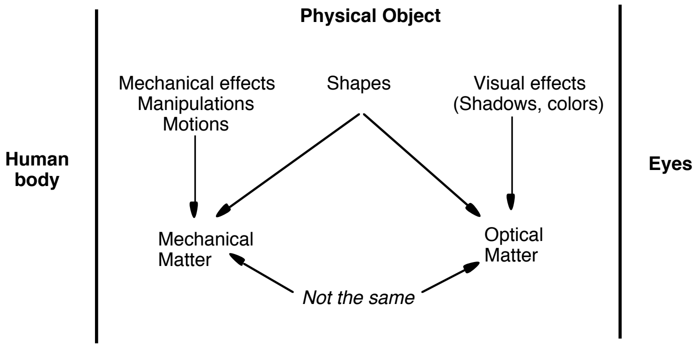


Figure 7. Geometry vs. Physics: the ambivalence of the shape.

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.

13. 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.

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.

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

[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

[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.