Multi Sensory 3D Tours for Cultural Heritage: the Palazzo Grassi Experience

Fabio Pittarello (*)

(*) Università Ca' Foscari di Venezia, Dipartimento di Informatica, Mestre (VE), Italia
E-mail: pitt@unive.it

ABSTRACT

Internet has brought to cultural institutions the chance to use new communication paradigms.

In particular the introduction of 3D technology for the network promises to be a powerful means for the diffusion of scientific and humanistic culture, especially for novices that can find in the 3D metaphor a more intuitive paradigm to access cultural information.

This work analyzes the current points of weakness of navigation in 3D worlds and proposes new guidelines to improve the quality of interaction for cultural heritage applications. A particular reference is made to naives end users which may be experts in their special application domain but are generally not capable of orientating, navigating and exploring within virtual information and knowledge environments, due to the lack of support provided for these activities by nowadays information systems and their 3D-visual front end architectures and tools.

The paper proposes a number of guidelines to overcome the current limits in 3D interaction and to help authors to build more effective 3D simulations.

These guidelines have been applied to a number of implementations made for the Palazzo Grassi web site (Palazzo Grassi is the world wide known institution that promotes cultural exhibitions and events in Venice); these realizations have been precious for receiving feedback from users and for refining the proposals, according to the principle of iterative design.

KEYWORDS: 3D guided tours, multimodality, navigation, Palazzo Grassi, VRML

INTRODUCTION

The advent of Internet has brought to cultural institutions the opportunity to experiment innovative methods to communicate cultural and scientific contents.

In particular the introduction of 3D technology for the network promises to be a powerful means for the diffusion of scientific and humanistic culture.

The introduction of 3D metaphors for cultural heritage can be a complementary or alternative approach both for experienced users, that may find stimulating and more informative the experiential approach offered by 3D, and for novices that may find more intuitive and familiar the 3D metaphor.

The latter class has grown with the recent massive access to Internet and it is composed by users coming from a variety of experiences, not necessarily trained in computer science; it is a priority to offer them more intuitive ways to access information.

Unfortunately, 3D interaction, in spite of the evolution in hardware technology, has not reached a satisfactory level of development.

So we examined the causes that prevent a full development of 3D worlds, in particular the difficulties in performing the basic tasks such as identification of the scene structure, orientation and navigation.

The formalization of the navigation problem and the analysis of real life experience helped us to identify the points of weakness and to find solutions to improve the usability of 3D interfaces for virtual environments.

The guidelines resulting from the analysis were applied to a number of implementations made for the Palazzo Grassi web site [14]; in this paper we present the *Artist's Eyes* and the *Einstein Tower* worlds, that show two different ways to implement the 3D metaphor: the first implementation is characterized by an abstract 3D universe that the user can explore to find information about objects and works realized by Picasso between 1917 and 1924; the second world is characterized by the presence of a complex building, the Einstein Tower, used as a site for a virtual cultural exhibition.

The user feedback has been very useful to refine our proposals and to model enhanced versions of our implementations, according to the principle of iterative design.

COMPUTER HUMAN INTERACTION: A BRIEF SUMMARY

In a few decades the domain of computer human interaction has rapidly shifted from a situation characterized by punch cards and batch processing to a scenery defined by a strong augment of computing power and high quality devices for interactive input and output. In this rapidly evolving scenery, the study of computer human interaction has been largely influenced by the availability of computing power.

Graphic bi-dimensional interfaces are today the most used metaphors to interact with a personal computer; research on visual interfaces has led to a number of solutions characterized by stability and maturity; in particular a number of visual and grammar rules to mediate the interaction has been formalized and widely accepted by a wide range of users [1] [9].

3D interaction has a different story: hardware constraints and lack of standardization have limited till recent times the use of 3D scenes for interactivity. The most part of research about 3D has been directed toward rendering engines to improve their quality and efficiency [7]; besides, a complementary effort has been dedicated to build systems for the manipulation of wireframe 3D elements, mainly for mechanical and architectural fields. In the majority of these situations, 3D representation is the final result of a modeling process rather than a starting point for interaction; besides, in general, the final result is presented using a number of rendered static views; in this sense the communication strategy of architects that use digital tools has not evolved too much from that one used by their colleagues during the Renaissance age.

In other words, we could say that interaction on a 3D environment is an open field for research, and a considerable effort should be spent in order to achieve a full exploitation of 3D interaction potentialities. In spite of this undefined scenery, the situation is rapidly evolving and new low-cost accelerated video cards are bringing to a broad range of users the power of 3D representation.

Besides, the advent of Internet technology has brought a great impulse towards the definition of open common standards; HTML [17] is the most outstanding representative of this tendency; but this general trend has involved also the field of 3D interaction, and has produced VRML (Virtual Reality Modeling Language [16]) as a result of this effort. VRML is a language that describes both geometric objects and interaction primitives and it has been conceived to create and distribute 3D interactive worlds on the net. At present it is really one of the few (or perhaps the only) non proprietary language for desktop users that allows to describe both geometry and behaviors in 3D worlds; besides, integration with HTML and its evolution (XML [18]) is allowed and promises to be more effective with the new version of this language (X3D [16]); this seems an interesting field to explore in order to augment the efficiency and the usability of 3D user interfaces. This is the reason why we chose VRML as the basis for all the implementations accompanying this work.

The User Interaction: an Exploration Based Approach

There is a number of differences between the visual structure of 2d and 3D interfaces: the most prominent one is probably that in the most part of situations, 3D interfaces don't offer a unique, comprehensive synthetic view; the user needs to move through the scene to perform the given tasks. In other words, 3D interfaces generally offer a peculiar approach, based on exploration rather than on synthesis; therefore orientation and navigation are issues of fundamental importance for 3D interfaces.

We can roughly distinguish between two basic categories of the 3D representation: the first category is related to the metaphorical or symbolic representation of qualitative and quantitative data; the second category is related to the representation of real 3D scenes that are the primary focus of the investigation.

In the first situation metaphors based on abstract or real objects can be used to access punctual information about a certain record of a dataset or to visualize certain properties common to the dataset itself. For example a simulation of an office environment can be used as an intuitive metaphor to have access to particular classes of activities (i.e. folders to represent information about current projects, phones to access the phone book, etc.). Besides, symbolic representations are used also in game environments that represent a consistent part of 3D worlds.

In the second situation the 3D scene represents a real environment for exploration; there are a variety of situations related to this category, ranging from the visualization of urban landscapes to simulations of the industrial plants. Again 3D game environments are often interactive experiences based on simulation of real environments.

3D interfaces often present a mix of elements belonging to the categories described above, for example in the case of simulations of real environments which contain objects that acts as metaphors to perform actions or to access information.

In both cases orientation and navigation are crucial issues, because they are the primary functions of the interaction model or they are prerequisites functions for the execution of more sophisticated behaviors. These issues represent a challenge for the human interaction research field; the following discussion will analyze the state of art, identifying points of weakness and suggesting a range of solutions.

ORIENTATION AND NAVIGATION IN 3D INTERFACES: POINTS OF WEAKNESS

As stated in the beginning, the considerations will be limited to the so called desktop virtual reality, i.e. virtual reality based on the interaction with a screen rather than on full immersion obtained with head mounted displays or other devices. In particular we will analyze the state of art for VRML browsers, applications designed for visualizing virtual worlds built according to the VRML standard.

In general we can identify some major drawbacks in the use of current navigation interfaces for VRML browsers:

- the paradigm of interaction in many cases is not easy to understand; the last generation consoles for navigation have too many options for interaction and this generally creates confusion;
- the default interface of VRML browsers usually lacks tools for orientation;
- the visual field for desktop virtual reality is narrow, and this brings additional difficulties for orientation and navigation; the scene representation has to be simplified due to performance limits on average personal computers, even if this situation is rapidly changing thanks to the new generations of graphic cards.

What shall we do to increase the usability of navigation in 3D worlds? To give a correct answer to these questions we'll try to supply a better formalization of the navigation problem, distinguishing three related primary issues: identification of the scene structure, orientation and navigation.

THE REAL WORLD SCENE

Before suggesting enhancements to current interfaces, we will consider the three different navigation issues in real world interaction, in order to extrapolate some useful hints for artificial environments.

The observation of the real world scene seems to be a good starting point for building 3D worlds in which the experiential processes typical of reality are replicated; a similar approach characterizes a recent paper that defines the requisites for 3D stores for e-commerce [6]; in this paper the discussion goes through the analysis of guidelines existing for stores in real world, in order to capture the features that attract customers and to replicate them in virtual reality.

Ultimately this approach should lead to an extensive set of guidelines for 3D worlds; these are already available for 2d interfaces, but they are definitely a major and still unexplored issue for 3D interfaces.

Identification of the Structure of the Scene

In the real world people usually remember a particular scene where the visual relationships between the objects that compose the environment are evident.

For example, it is easy to remember a square with a church placed on the main axis, with a theory of columns on one side and a tower on the other side. The spatial relations between these objects are evident and constitute the main contribution to the scene identification.

On the contrary, there are situations where the spatial relations between objects are not evident (i.e. a degraded suburb), where objects are not sufficiently characterized (i.e. a district characterized by an endless repetition of the same inhabited module), or even particular conditions where there are no recognizable objects (a wide desert or even a foggy zone). Usually people associate to all these situations a sense of uncertainty and frustration.

This situation is not peculiar of the real world, but it is frequent also in synthetic worlds, for instance in 3D simulations characterized by a low number of geometries, in 3D worlds where the narrow views don't permit to recognize the spatial relationships between objects or even in 3D representations where objects with the same morphology activate different categories of actions.

Therefore even in virtual worlds we should avoid to generate spatial configurations that cause confusion and disorientation.

This doesn't mean that we can't use a low number of geometric primitives to characterize a 3D environment, but that we have to define spatial relations which allow the user to recognize the visual structure of the scene.

Orientation

Recognition of a particular place is only the first requirement for navigation; there is also orientation, which involves relations between venues; there is a number of elements that people use to find their position in a structured space:

- artifacts inserted into the scene that use iconic or textual languages to communicate directions; in some cases there are various parallel levels of sign languages referring to different uses of the same physical space (i.e. visual signs used in a complex building, such as an hospital or a public palace);
- tools that identify some properties of the current position in relation to some natural or artificial invariable properties (compasses, GPS receivers);
- artifacts that give various levels of synthesis for navigation (city plans, maps, etc.).

Navigation

Finally the third element to be considered is navigation itself; in the real world this may be an implicit issue if we consider the act of walking through an environment; nevertheless there are a lot of situations in which people use a specific language to navigate (this happens when people drive a car, use a lift or operate an artifact built to overcome physical disabilities).

Multimodality

After having analyzed the three key elements for navigation, we must emphasize a parallel aspect characterizing the real world: the everyday experience is based on multimodality, which means giving and receiving different information simultaneously from different senses [13].

Vision, hearing and touch are all part of the multimodal experience; the efficiency of navigation and the task completion are often related to the simultaneous stimulation of them. This apparent redundancy is often useful in many situations and, in some cases, it can be vital.

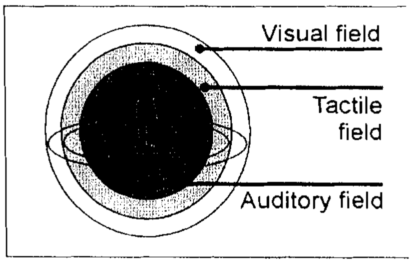

Fig. 1 - Multimodality plays a key role in everyday experience; visual, auditory and tactile fields are simultaneously stimulated by information coming from the environment.

GUIDELINES FOR VIRTUAL ENVIRONMENTS

Our thesis is that also navigation in a 3D synthetic environment requires a satisfactory resolution of all the issues emphasized for the real world: scene identification, orientation and navigation; besides, we suggest the importance of the multimodal approach even for synthetic scenes.

Therefore in the following paragraphs we'll describe a number of guidelines for virtual worlds critically extrapolated from the analysis of the real world and we'll show how we used them in the making of a number of experimental 3D interactive worlds.

Multimodality for the Scene Structure

As stated before, the scene structure is more schematic in synthetic 3D environments and the narrow optical cone is an additional point of weakness for scene recognition.

We propose to enforce the strength of architectonic language with a parallel multimodal approach; in particular, using the VRML standard it is easy to superimpose additional layers of 3D auditory information to the 3D visual scenes.

A number of studies supported the introduction of a multimodal approach in the human-computer communication [12]; for what concerns the addition of a parallel audio component, Meera Blattner et al. [2] [3] [8] [10] proposed to add an auditory counterpart of icons to graphic bi-dimensional interfaces, defining the *earcons* as *non-verbal audio messages used in the computer/user interfaces to provide information to users about some computer object, operation or interaction.*

What we suggest (and what we contributed to experiment in the prototypes described below) is:

Guideline 1: to extend the use of earcons to identify the scene structure in 3D scenes.

In particular the *Einstein Tower* world, which will be discussed in the following paragraphs, shows the use of *earcons* as enforcement for the identification of the scene nodal places, an additional auditory mark-up of the place the visitor is currently walking in.

Earcons for Orientation
VRML browsers usually don't have a built-in standardized support for user orientation; tools for orientation are implemented by world builders and, in the most part of situations, they consist in visual signs inserted into the scene.

Again the use of multimodality can bear significant enhancements over the current situation; therefore what we propose is:

Guideline 2: to extend the use of earcons for orientation in 3D scenes.

For example, a number of *earcons* placed at fixed locations can give information about the position of important places in the 3D world or about the relative position of the user (see Fig. 2A).

Another option can be the use of the *earcons* to give *dynamic* hints about the correctness of a certain direction; in this case the *earcons* can be used as signs dynamically changing position to suggest direction, activated by the user motion (see Fig. 2B).

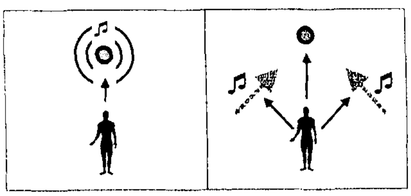

Fig. 2 - Orientation can benefit from the implementation of *earcons*; the figure on the left shows an example of a static *earcon* signaling a key position; the figure on the right shows an alternative approach, based on the use of dynamic earcons; if the user deviates from a given target, a sequence of tones moving toward the focus of navigation is played, suggesting the right direction.

A first attempt to carry out the implementation of auditory artifacts for orientation has been proposed in the *Artist's Eyes* world (see below); in that case these earcons have a function as location markers, and in addition they suggest the relative position of the visitor respect to the starting and arriving points.

Simplification of the Navigation Interface
There is a range of solutions for navigation in 3D virtual worlds; we can distinguish three main categories:

- console-style visual interfaces, based on the composition of a number of navigation tools placed below the window representing the 3D scene;
- visual interfaces based on artifacts inserted into the scene;
- haptic devices interfaced with the scene; there is a number of devices for interaction: mouse, keyboard, game-pads, head-mounted trackers, etc.

Implementations are various and include a number of solutions that uses a mix of interfaces taken from the different categories given above.

What we propose, as a general approach to the problem, is

Guideline 3: to simplify or to modularize the navigation interface

The majority of solutions are characterized by the use of a huge number of artifacts, even if the user needs only a part of them to navigate. Developers often forget that the interface should be more intuitive for the non-expert user: the last generation of VRML browsers proposes interfaces taken from the realm of 3D modeling, more familiar to a CAD designer than to a generic user.

If the interaction is particularly sophisticated in terms of quality and number of actions to perform in order to obtain a certain task, we should think in terms of artifacts that automatically trigger a coordinated sequence of actions; besides, the activation of a specific sequence could depend on the context (i.e. a particular position of the user or the localization of the user in a specific point of the scene); in this way it would be possible to have a low number of artifacts for interaction while providing support for sophisticated needs.

This approach is used in some video game interfaces where the same user gestures lead to different results in the scene according to the change of context; this modality could be proficiently extended to more serious applications.

THE IMPLEMENTATIONS

After having analyzed the weaknesses of the current user interfaces and having proposed some guidelines for a better navigation, we will present two different practical implementations; the first approach is related to the access to a set of data, the second one is related to the representation of a real environment. These 3D metaphors were commissioned and built in occasion of two important cultural events in Venice promoted by Palazzo Grassi, the worldwide known cultural institution promoting exhibitions ranging from Archaeology to Modern Arts.

The presentation order of these 3D worlds is not chronological, but it reflects the different complexity of guidelines implementation, starting with the simpler one.

These implementations have been conceived for the network user and are based on VRML language [5] [11].

THE ARTIST'S EYES

The *Artist's Eyes* world is conceived as a didactic environment through Picasso's life and works between 1917 and 1924, realized in occasion of the Picasso 1917-1924 exhibition held in 1998 at Palazzo Grassi in Venice.

3D is used here as a metaphor to access hypertextual and pictorial information: a number of pavilions (one for each year) is placed on an abstract plane dominated by the artist's eyes and by the works painted by Picasso during this period.

The visitor uses simplified 3D cognitive artifacts integrated into the scene to move through the pavilions, to access information about the specific year and to start animations related to the artist's work.

The *Artist's Eyes* world, characterized by the integration of hypertext and 3D representation, implements the concepts of *interaction locus* and multimodality to increase efficiency and usability.

3D as a Metaphor to Access Information

Pavilions placed on an infinite plane are the key elements in the *Artist's Eyes* environment; each panel is built using two panels placed side by side: a photograph about Picasso's life is exposed on one panel; an artist's work, chosen to represent the artist's activity for each year between 1917 and 1924, is exposed on the other panel; hypertextual information about the artist appears on the left side of the environment each time the user approaches to the pavilion (see Fig. 3).

The exploration of this world is

conceived as a temporal travel through Picasso's life; this means that even if the 3D geometry offers the chance to be visited in any direction, a particular path is privileged; that is, this experience can be catalogued in the so called *guided tours*: the starting point and the following stations are chosen by the world author in a precise order to maximize the user satisfaction in relation to the world exploration.

For what concerns the objects populating this world, we can point out two main categories: information objects (photographs and works) that the user can explore just clicking over them; artifacts to trigger movements and complex actions.

In this experience we use a simple object (the *Harlequin's hat*) as a trigger to start a number of animations concerning Picasso's work; each pavilion contains a *Harlequin's hat* and becomes a portal evoking different aspects of the artist's work exposed on the panel.

For example, the 1917 pavilion focuses on the curtain, painted by Picasso for *Parade*, a ballet written by Jean Cocteau. When the visitor clicks over the *Harlequin's hat* displaced in the centre of the pavilion, the pavilion's panel vanishes and the visitor is introduced into a theatre adorned by the curtain painted by the artist, while the notes composed by Erik Satie for the ballet are played.

In this case the visitor has the opportunity to compare the dimensions of the painted curtains with a scale factor, the theatre's grades. Besides, the music composed for the ballet adds a further level of information to the user experience.

Interaction Locus

Following the guidelines given above, we used multimodality to overcome some limits of vision introducing the concept of *interaction locus*; it is an extension of the architectonic concept of *venue*, obtained associating to it both hypertextual and auditory components; more formally we could define *the interaction locus as the basic element to give a structure to 3D space, summa of coordinated 3D representation, auditory signs and hypertextual information.*

For what concerns the *Artist's Eyes*, the architectonic language has been used to build a serialized artifact (the pavilion) as the place deputed for interaction; the complementary languages involved in the definition of interaction *locus* (text and auditory language) have been used to give an identity to each serialized element: each time a visitor approaches to a different pavilion, textual information and a bi-tonal individual auditory signal inform the visitor about the *locus* identity.

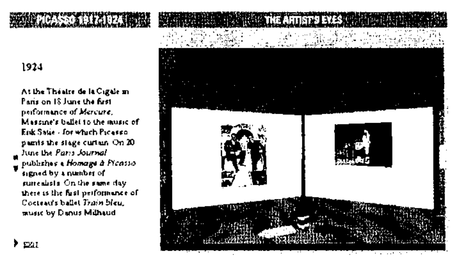

Fig. 3 - A screen shot from the *Artist's Eyes* environment: the 1924 Pavilion. Hypertextual information about the artist's life characterizes the left part of the screen. 3D information is shown on the right. The mouse clicks over the photograph (left panel) and the artist's work (right panel) allow to zoom and to have additional information about the chosen item.

Navigation icons are visible on the lower part of the screen: the blue and yellow arrows enable the user to go forth and back, along the navigation path; the *Harlequin's hat* starts the 3D animation.

Earcons for Orientation

In the *Artist's Eyes* world we use earcons to identify the *locus* the visitor is walking in. Besides, they work also as an auditory sign for orientation, informing the visitor about his/her relative position: each time the visitor walks into a particular *locus*, two different sounds are played in fixed sequence: the first tone volume is proportional to the proximity to the starting point; the second tone volume is proportional to the proximity to the end of the world, measured long the suggested navigation path. Fig. 4 sketches the relationships between the user position and the tone volumes.

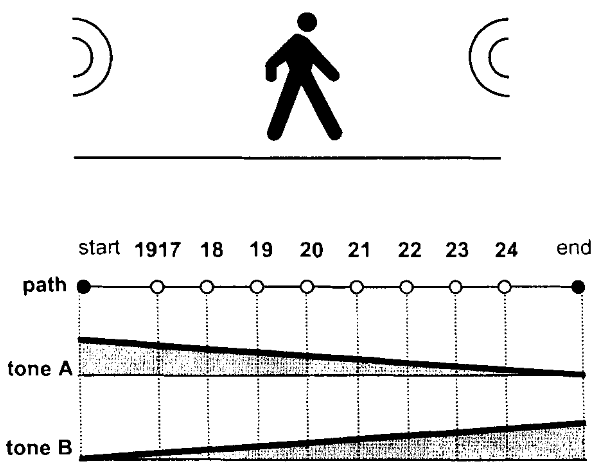

Fig. 4 - The navigation *earcon* is a two-tone sequence that gives information about the position of the visitor along the navigation path; volumes of tone A and tone B are proportional to the user proximity to the starting point and to the end of the navigation path. Vertical dot lines show tone volumes when the user enters a specific *locus* (for example, when the user enters the 1924 *locus*, tone A is nearly inaudible and tone B is nearly at full volume).

The Navigation Interface: Less is More

As stated before, the guided tour is the peculiar navigation modality provided for the *Artist's Eyes* world.

A small number of geometric artifacts has been conceived for this purpose as a replacement for the interaction objects contained into the default navigation console, that in general are excessive in number and not intuitive for users not acquainted with CAD graphical tools.

The cognitive artifacts we used in this case are simply arrows to move forth (the yellow arrow) and back (the blue arrow), along the navigation path.

A great number of options is not generally useful for the majority of interactive experiences; if the 3D experience needs the activation of complex interactive behaviors, the complexity can be encapsulated and activated by additional simple artifacts, like the *Harlequin's hat*, that we placed between the arrows; this artifact, when pressed, activates a guided tour that focuses on the themes related to the artist's activity.

This is all the user needs to interact in this world, characterized by a simple geometric structure and linear navigation modalities.

It is a first step towards simplification and standardization of the interface; we can conclude saying that even for navigation artifacts, the old *motto* suggested by one of the most famous German architects, Mies van der Rohe, is valid: *less is more*, that is, simplicity of expression can lead to more valuable results than complex solutions; in our specific case: a reduced number of rationalized navigation artifacts is more helpful and valuable for clarity and efficiency than a complex interface.

User Feedback

We registered the users' impressions about the usability of the environment. In general people using the simplified interface had no problem in navigating through the world even with no preliminary training: users just looked at the basic help appearing at the beginning of the tour and rapidly took confidence with the interaction objects.

People appreciated that an articulate tour could be controlled with only a little initial cognitive effort, using objects easily recognizable in the 3D scene.

Besides, people appreciated that even the animations sequences were introduced and concluded by the same recurring visual signal: the panel shifting up and down the original position. Some people considered the directional arrows, the hat and the panel's motion, as different signs of a language that they could easily learn while performing the interaction.

Other people hinted that the interaction paradigm was so essential that it rapidly *disappeared*, leaving all the concentration on the cultural contents of the 3D world.

For what concerns sound, people in general appreciated the additional level of information for orientation given by the earcons, even if in a number of cases they had to be informed about the meaning of the auditory signals to take full advantage of this feature.

Users suggested also a more articulated use of music, to avoid the alienating effect typical of early mute cinema: sound is part of the real life, and experiences that simulate it can't be complete without the stimulation of ears.

THE EINSTEIN TOWER

The *Einstein Tower* is another implementation of a number of suggestions to improve navigability in 3D worlds, conceived in occasion of the *German Expressionism* exhibition held at Palazzo Grassi in Venice.

In this case we proposed to model a virtual counterpart of the Einstein Tower, a sun observatory built in Potsdam from 1919 to1923 by Erich Mendelsohn and chosen as a symbol of the exhibition in Venice; the VRML version (see Fig. 5), conceived for the Internet user, becomes the focus of a virtual exhibition where architecture, paintings, manifestos, cinema fragments and music melt into a unique composition, a small account of what Expressionists called *Gesamtkunstwerk*, an integrated aesthetic experience achieved by eliminating the divisions between architecture, music and visual arts [15].

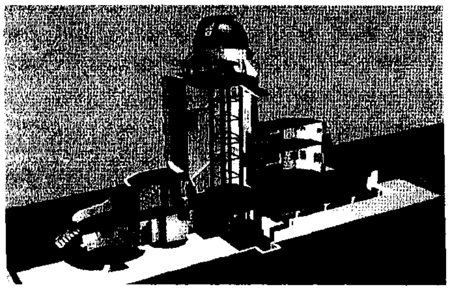

Fig. 5 – A cross-section of the Einstein Tower model.

A selection of the expressionist works of the real exhibition are displayed in the frame of the VRML model; the research rooms and the laboratories of the tower are converted into exhibition rooms exploring the different themes and the different phases of the artistic movement.

At any time, for each piece of art exposed in the tower, the visitor has the chance to ask for information, just clicking the mouse over the corresponding image; moreover, textual information links to other information, in the usual way provided by hypertextual tools.

We must point out that, in this case, the 3D simulation and the navigation needs are rather different from the ones conceived for the *Artist's Eyes* world. In the former situation there was no architecture to explore and the navigation along a linear path was the only significant way to move across this world. In this case a preferential path (but not unique) is suggested, choosing among a number of hypertextual nodes disposed on a real architecture; the user has the chance to follow the path, but he can also move freely through the world. For example, an architect or a scientist could be interested to explore the structure of the tower, both an architectonic work and a functional building for experiments about the solar spectrum.

Enforce the Scene Structure: the *Interaction Locus*

The concept of *interaction locus* was applied also in the *Einstein Tower* world. The scene is rather different from the previous implementation; in this case a real complex building, both a container for the exhibition and a work of art itself, is presented for exploration. All the *locus* components (geometry, text and audio) are differentiated in order to help the user to identify his/her location and to prevent him/her from wandering lost in the tower.

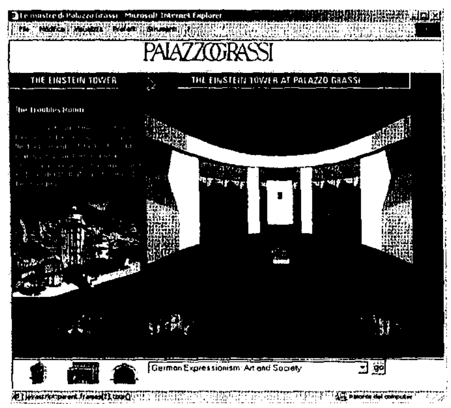

Fig. 6 – A screen shot from the Einstein Tower world: *The Troubles room*; hypertextual information is visualized on the left; 3D representation is shown on the right; every time the user enters the room a background auditory loop is played to help *locus* recognition.

In the Einstein Tower world the auditory component is implemented in the form of a background auditory loop univocally associated to each significant *locus*. Therefore in this case *earcons* work as *memoria loci*, auditory reminders automatically activated each time the visitor enters a particular venue. The textual component of the *locus* is implemented in the form of a hypertextual area visualized on the left of the 3D scene; every time the user enters a *locus*, textual data are offered to help scene identification (see Fig. 6).

Navigate through a Collection of *Loci*

The whole Einstein Tower world is a collection of *loci*, each of them defined as a coordinated set of three-dimensional, textual and auditory information (for example the *locus First Impressionism Room* is defined as the *summa* of the three-dimensional representation, the hypertext automatically linked to describe the contents of the room and the Schönberg music audible only when the visitor walks inside it).

As stated before, the opportunity for the user to move in 3D spaces without constrains has been complemented with the implementation of a preferential path to visit the virtual exhibition: a simple mechanism gives the visitor the chance to be guided through the rooms of the tower. Yet, at any time, the visitor has the opportunity to choose a particular node (point of view) among the hypertextual network of loci to re-start exploration or to freely move in the 3D world (see Fig. 7).

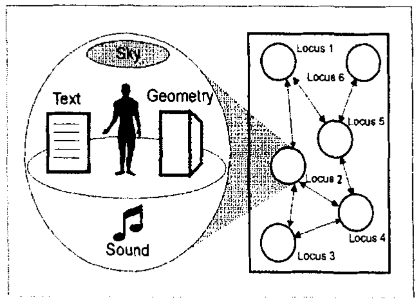

Fig. 7 -Navigation in a 3D environment: a guided tour through an ordered collection of loci.

User Feedback

Even for the *Einstein Tower* world, the users appreciated the simplicity of interaction mechanisms, even though some of them stated that the artifacts used inside the 3D scene, as in the *Artist's Eyes* world, were more intuitive; anyway, they agreed that the additional opportunity given for the *Einstein Tower* to move without constraints could have brought some problems for the identification of navigation artifacts disseminated in the 3D scene; in fact particular views on the scene resulting from free navigation could hide the navigation artifacts, preventing the users to return to the guided modality.

The users that tested also the *Artist's Eyes* world stated that the *Einstein Tower* was much more complex because of the morphology of the building, but at the same time the mixture of architecture, works of art and music gave a true feeling of what the Expressionism really was, even to people who were not informed about this artistic movement.

In other words, the subjective immersion in a 3D scene, where heterogeneous elements were merged, was felt much more appealing than alternate educational methods, for example an hypertext about Expressionism.

Users appreciated even the freedom to wander freely in the *Einstein Tower*, but only as a second opportunity to focus on some aspects of the building after a first guided tour.

More precisely, some users suggested to implement also an automated guided tour; in their opinion, visitors should have the chance to activate, at their will, the automatic modality and to stop the tour any time they would find some objects to investigate on. This suggestion has been implemented in an improved version of the Einstein Tower world and it is already available on the Palazzo Grassi web site.

For what concerned textual information, we performed a test with a group of users, giving them the chance to experiment both a VRML-only version and the subsequent HTML-VRML integration; everybody appreciated the additional textual information; they were pleased to find qualitative information about the loci they were visiting; besides, they all agreed on the usefulness of auditory information to help identification of venues.

Finally some people appreciated that only the information about *loci* was given automatically; they appreciated that the information about the works of art, felt as a further level of understanding, was given only with an explicit user click.

A FURTHER STEP:
ADDING STANDARDIZATION TO INTERFACE

As stated in the introductory chapter, standardization has been one of the main keys for the success of bi-dimensional graphic interfaces.

Lessons learned from pioneer Xerox studies on graphical interfaces led the major producers of hardware and software to develop a number of standardized classes of visual objects for interaction called *widgets* [1] [9].

This standardization has demonstrated to be a powerful opportunity for the development and the interaction with bi-dimensional user interfaces [4].

Standardization is a feature that currently lacks in 3D scenes and this situation has led to longer development cycles for the absence of pre-defined objects to use as building blocks for 3D worlds and to longer learning periods for users interacting with different 3D environments; so, as a further step for our investigation, we tried to add it to 3D scenes. In particular we tried to standardize a part of the guidelines given before, concentrating on the auditory artifacts.

A Parallel Effort: the Universal Media Approach

The Web 3D Consortium [16], responsible for the development of the VRML standard, has recently proposed a first step towards the creation of a standard library of objects for 3D worlds: the *Universal Media* (UM) initiative.

In particular the Universal Media Working Group has proposed to increase the realism of VRML worlds and decrease network downloads by defining a small, cross-platform library of locally resident media elements (textures, sounds and VRML objects) and a mechanism by which VRML content creators can incorporate these media elements into their worlds.

Universal Earcons: A Proposal for a Class of Informative Auditory Objects

Our contribution to the standardization of 3D interaction elements fits into the existing architecture of VRML and shares some features with the Universal Media initiative, in particular the creation of a high quality standard library of objects shared by all world builders and available to all final users, both in the form of a packed library to install in the local hard drive and as a direct link to remote repositories on the Internet.

But there are also important differences; the most significant is that our library adds a semantic connotation to objects; we can imagine it as part of a standardized collection of interaction objects available for world builders, to provide uniform interaction mechanisms to final users.

Name	Geom. Properties	Audio Properties	Additional Features
***Locus* Identifier**	(center, radius)	(volume, tone, loop)	Positional static earcon
Relative Position Identifier	(center, radius)	(volume, tone1, tone2, loop)	Positional static earcon
Audio Arrow	(start position, radius, path)	(volume, tone, loop)	Positional dynamic earcon
Trigger Object	(center, radius)	(volume, tone)	Positional static earcon
Informative Object	(center, radius)	(volume, tone)	Positional static earcon
Object Selection	(center, radius)/none	(volume, tone)	Positional/ Non positional earcon
Confirm. for Action Execution	(center, radius)/none	(volume, tone)	Positional/ Non positional earcon
Generic Error	(center, radius)/none	(volume, tone)	Positional/ Non positional earcon
Incoming Message	(center, radius)/none	(volume, tone)	Positional/ Non positional earcon
Outgoing Message	(center, radius)/none	(volume, tone)	Positional/ Non pos. earcon

Table 1 – A class of auditory artifacts

Table 1 shows the components that are part of this new class of interaction objects.

The simplest of them are derived from their bi-dimensional counterparts and can be used by world builders who need to give information about events that do not have spatial locations; the second category has additional geometric coordinates to localize the event in space, for example an error or a confirmation referred to a specific geometrical object in the scene.

CONCLUSIONS AND RESULTS

In this paper we considered the role of 3D desktop technology in relation to cultural heritage.

This application field is very promising, both for experienced users and, most of all, for a new category of inexperienced Internet users that can find very useful an intuitive paradigm for approaching cultural information.

Unfortunately, in spite of the evolution in hardware technology, interaction in 3D worlds for the net has not reached a satisfactory level of development.

We have examined the causes that prevent a full exploitation of this new media; in particular we have formalized the navigation problem and analyzed the real life experience to identify the points of weakness and to improve the usability of 3D interfaces for virtual environments.

We have proposed the multimodal approach, typical of the real life experience, to achieve higher level of efficiency and usability in navigation and interactivity.

We have introduced multimodality both formulating the concept of *interaction locus* and extending the use of *earcons* for 3D as an additional help for navigating the scene.

Besides, we have suggested the simplification of the navigation artifacts, including also the introduction of simple artifacts to trigger complex actions.

We have applied successfully our proposals to a number of complex 3D worlds built for Palazzo Grassi in Venice; the feedback of the users has confirmed the effectiveness of our approach and has given us further suggestions that we have applied in enhanced versions of our worlds.

Finally the formalization of *earcons* as a set of auditory artifacts for general use has been a further significant contribution to smooth the learning curve of final users and to give authors a common language for interactivity.

ACKNOWLEDGMENTS

I would like to acknowledge Giuseppe Donegà and Pasquale Bonagura, heads of the Palazzo Grassi management, for their willingness to explore new methods to communicate cultural heritage contents. Thanks to Otello Martin, head of the C.T.A. (Università Ca' Foscari di Venezia), who gave me the opportunity to direct the design process of the Palazzo Grassi web site. His shrewd remarks were precious to refine the design process and the implementation efforts.

REFERENCES

1. Apple Computer. *Macintosh Human Interface Guidelines*, Addison-Wesley, Reading, 1992
2. M. Blattner, D. Sumikawa, and R. Greenberg. Earcons and icons: their structure and common design

principles. In *Human Computer Interaction*, 4, 1989, 11-44

3. S.A. Brewster. *Providing a structured method for integrating non-speech audio into human-computer interfaces*, PhD thesis, University of York, York, UK, 1994

4. E. Buie. Hci standards: a mixed blessing. In *Interactions*, 6(2), 1999, 36-42

5. [5] R. Carrey and G. Bell. *The Annotated Vrml 2.0 Reference Manual*, Addison-Wesley, Reading, 1997.

6. L. Chittaro and R. Ranon. Virtual reality stores for 1-to-1 commerce. In *Proceedings of the CHI2000 Workshop on Designing Interactive Systems for 1-to-1 E-Commerce*, The Hague, The Netherlands, 2000

7. J. Foley, A.van Dam, S. Feiner, and J. Hughes. *Computer Graphics: Principles and Practice*, Addison-Wesley, Reading, 2nd edition, 1996

8. W. W. Gaver. Auditory icons: Using sound in computer interfaces. In *Human-Computer Interaction*, 2, 1989, 167-177

9. Microsoft Corporation. *The Windows Interface: An Application Design Guide*, Microsoft Press, Redmond, 1992

10. E. D. Mynatt. Auditory presentations of graphical user interfaces. In G. Kramer, Ed., *Auditory Display: Sonification, Audification, and Auditory Interfaces*, Addison-Wesley, Reading, 1994, 533-556

11. D. R. Nadeau. Building virtual worlds with VRML. In *IEEE Computer Graphics and Applications*, March-April 1999, 18-29

12. L. Nigay, J. Coutaz. A design space for multimodal interfaces: concurrent processing and data fusion. In *Proceedings CHI93* (Amsterdam), ACM Press, 1993, 172-178

13. D.A. Norman. *Things that make us smart - Defending human attributes in the age of the machine*. Addison-Wesley, Reading, 1993

14. Palazzo Grassi,
http://www.palazzograssi.it

15. F. Pittarello, M. Pittarello, and G.F. Italiano. Architecture and digital exhibitions - The Einstein Tower world. In *Virtual Environments 1998*, Springer Verlag, Wien/New York, 1998, 162-171

16. Web 3D Consortium,
http://www.web3D.org/