Archeologia e Calcolatori
Supplemento 1, 2007, 243-253

OMERO: A MULTIMODAL SYSTEM THAT IMPROVES ACCESS TO CULTURAL HERITAGE BY VISUALLY IMPAIRED PEOPLE

1. Introduction

The development of new technologies is continuously improving the interaction between people and digital systems, enabling new digital applications in several contexts of ordinary life.

New interfaces, like haptic interfaces (Salisbury, Conti, Barbagli 2004) or synthetic speaker or vocal command recognizers, make it possible to use more natural communication channels with respect to keyboard, mouse and monitor. These new human/machine interaction paradigms are not only simpler and more intuitive but also increase the amount and the kinds of information that can be conveyed to the user. Multimodal applications exploit different information channels (vision, touch, sound, language, etc.) in an integrated and redundant way (Jacobson 2002). Redundancy presents the same information in a polymorphous manner to match the specific abilities of the user and can make it easier for people with disabilities to cope with many aspects of real life (Alonso 2006). The use of haptic and acoustic interaction for the fruition of spatial data has been investigated in Van Scoy *et al.* 1999; Magnusson, Rassmus-Grön 2004; Murai *et al.* 2006, both concerned with urban and traffic environments.

The proposed system adds a haptic and an acoustic/vocal interaction to the common visual rendering. The haptic interface, a PHANToM device (Massie, Salisbury 1994), allows the users to "touch" 3D virtual models by applying to their hand a force feedback that realistically simulates the physical interaction with real geometries. The haptic use of 3D models enables the user to perceive environmental spatial organization as well as the distribution, shape and identity of objects. All these three-dimensional data are perceived inside a virtual space that provides more flexibility with respect to the real world.

In virtual space the user can interact with objects that could not be touched in reality (on account of their location, dimensions, vulnerability, etc.), can appreciate very small details by increasing their size up to the dimension of the user's fingertips, can acquire a mental scheme in a progressive and guided way. To this aim, the model can be organized into several levels associated with different semantic contents and proper forces can be applied to the user's hand along a predefined path that connects all the relevant parts of the model (Fig. 1). This experience in virtual space is meant to make the direct experience of complex environments more relevant and meaningful instead of being a mere substitute for the real perception experienced in the physical world.

Fig. 1 – A richly detailed 3D representation of an Italian historical building. The actraction forces (black arrows) bring the avatar towards the first target area along an axploration path connecting the relevant regions of the model. The grey arrows represent the contact between the probe and the model.

To simplify the perception of details that would be too small in a complete view of the environment, parts of it can be shown on a larger scale by changing the relative size of the models with respect to the avatar (representing the user's fingertip). The haptic/acoustic interaction with the virtual model makes extensive use of active objects belonging to three different categories: haptic objects (associated with one of the available tactile effects, such as vibration, viscosity, forces of attraction, etc.), acoustic objects (connected only with acoustic effects or vocal messages) and haptic/acoustic objects (associated with more complex effects involving both sensory channels to reach the goal of a simpler and more effective comprehension).

Vocal messages that inform the user about the type and identity of the object at hand can be associated with all the components of the scene: they quickly provide all the information that can be expressed synthetically and clearly in terms of words, leaving to the touch the task of revealing spatial data that could hardly be appreciated by a textual description.

Moreover, objects can be dynamic (associated to a state that can change over time as a door that can be open or closed) or static. The user can use the available interfaces (button on the haptic interface, keyboard, mouse and, in the future, vocal commands) to control the system and to change the state of dynamic components of the scene: this would allow use of the system as a way to control complex environments properly equipped with a network of sensors and actuators.

The applications described in this paper show the use of the system as a way to improve access to cultural heritage by blind people. All the properties described would enable the user to enjoy each artistic object in a properly represented and rendered context (involving historical, geographical and cultural information) to make the understanding of its value truly meaningful and complete. The system does not simply provide haptic and acoustic access to virtual data, it provides the tools to design and create a more powerful cognitive path to cultural heritage that will greatly enhance the user's understanding and appreciation.

2. The architecture

The architecture of the multimodal application, shown in Fig. 2, was developed to be independent from the haptic device being used. Each module in the application communicates with the underlying haptic libraries by means of a suitable wrapper that encapsulates the particular data types relative to the sdk being used, leaving the interface with the rest of the application unchanged.

The virtual scene is described by a structure called a Scene Graph, that consists of one or more nodes, each of which can represent a geometry, a property, or a grouping object. Hierarchical scenes can be created by adding nodes as children of grouping nodes, resulting in a directed acyclic graph. The scene is described by a VRML file that is provided in input to the application. This scene structure generates two internal representations (scene graphs): the Graphic Scene Graph (GSG) used for the graphic rendering and the Haptic Scene Graph (HSG) used for the haptic rendering.

The VRML file describes mainly the geometrical structure of the scene while a corresponding XML-based description file details how the objects must be associated with haptic and/or acoustic effects. Decoupling the virtual models from their translation in terms of haptic and auditory perceptions allows the same model to be dynamically associated with different multimodal behaviors.

The system is based on three main blocks, controlling respectively the haptic, acoustic and visual interaction with the virtual scene. The

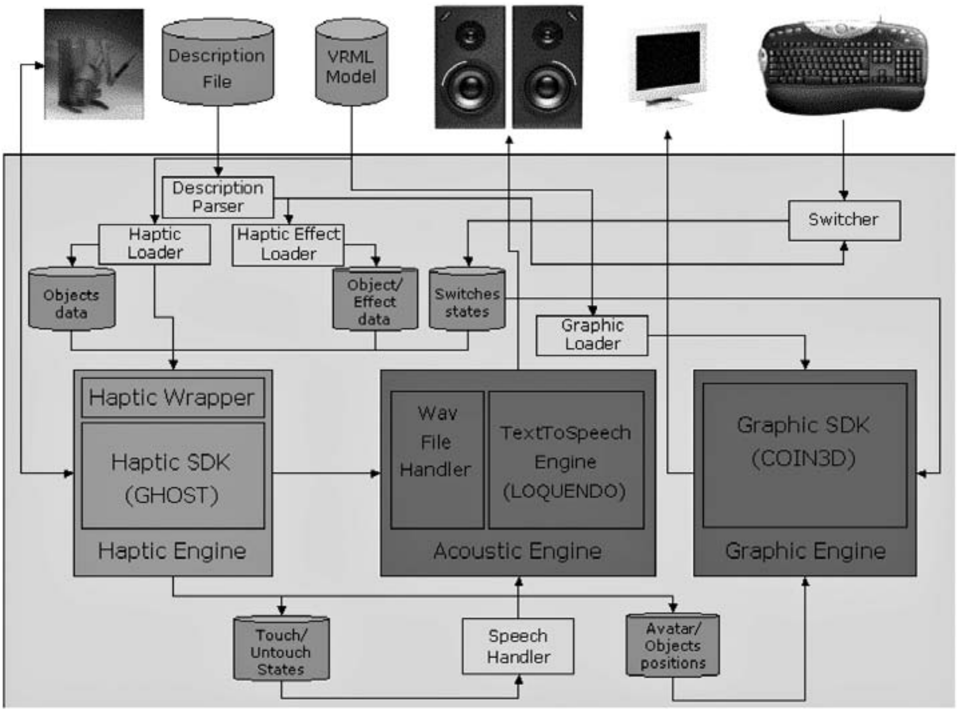

Fig. 2 – System architecture.

Haptic Engine, responsible for haptic interaction, creates the HSG using the data provided by the **Haptic Loader** and feeds it to the haptic loop that runs continuously, managing the haptic effects associated with special objects in the scene on the basis of information that the **Haptic Loader** and the **Haptic Effect Loader** extract from the input files. A particular haptic effect (vibration, viscous effect, force of attraction, etc.) linked to an object in the virtual scene is activated whenever the avatar touches it. These effects help the user to identify specific types of objects inside the whole scene. Finally, the **Haptic Engine** can enable/disable the perception of the different components of the model depending on the user's needs.

The second main module is the **Acoustic Engine** that is composed by two sub-modules: the **Wav File Handler** and the **TextToSpeech Engine**. The **Haptic Engine** directly calls the **Wav File Handler** whenever the activated haptic effect involves a corresponding acoustic effect. Instead the **TextToSpeech Engine** is called every time the user requests more information about the object being currently touched in the virtual scene, which is provided in the form of a synthesized vocal message, the text of which is selected by the **Speech Handler** module.

The visual interaction increases the versatility of the system which can be used contemporarily by blind and sighted users, each using the sensory channels in which they feel the most comfortable. The synchronization between the haptic and the visual loop is maintained via the Avatar/Objects positions data structure. The **Graphic Loader** prepares the data (Graphic Scene Graph) that are used by the **Graphic Engine** module to generate the pictorial rendering of the current scene.

At runtime the **Switcher** module creates and manages the mechanism to switch between the different semantic views defined for the model at hand. In a preliminary phase it modifies the scene graphs structure in order to make parts switchable. At runtime it checks the keyboard to intercept user input that determines the objects that need to be haptically and visually rendered. This function allows the user to access the information content of the model in a progressive and controlled way, simplifying their acquisition and comprehension.

The implementation of the system used in the experimental sessions described in this paper uses the PHANToM haptic device, with its GHOST haptic sdk[1], the COIN3D graphical library[2] and the LOQUENDO TextToSpeech sdk[3].

3. The experiment with the Svevian Castle in Bari

The VRML plant of the visitable part of the ground floor of the Norman-Svevian Castle located in Bari (Italy) was constructed from its detailed planimetry. The model includes: the entry area (including the ticket office), the external and internal courtyards, the gallery of plaster casts, the chapel and the bathrooms.

Two sessions of tests were run on two different groups of users. Each group was composed of four visually impaired people without any previous knowledge about the castle.

During the first test session, the virtual model of the castle presented to the users included most of the objects located inside the modelled environments. These objects (such as trees, hedges, pots) were represented by simple solid shapes and defined as acoustic active objects (the users received vocal messages clarifying their identity). Some doors could be opened and were defined as haptic/acoustic dynamic objects while doors to inaccessible environments were modelled as static objects with associated vocal explanatory messages.

[1] www.sensable.com/.
[2] www.sim.no/.
[3] www.loquendo.com/.

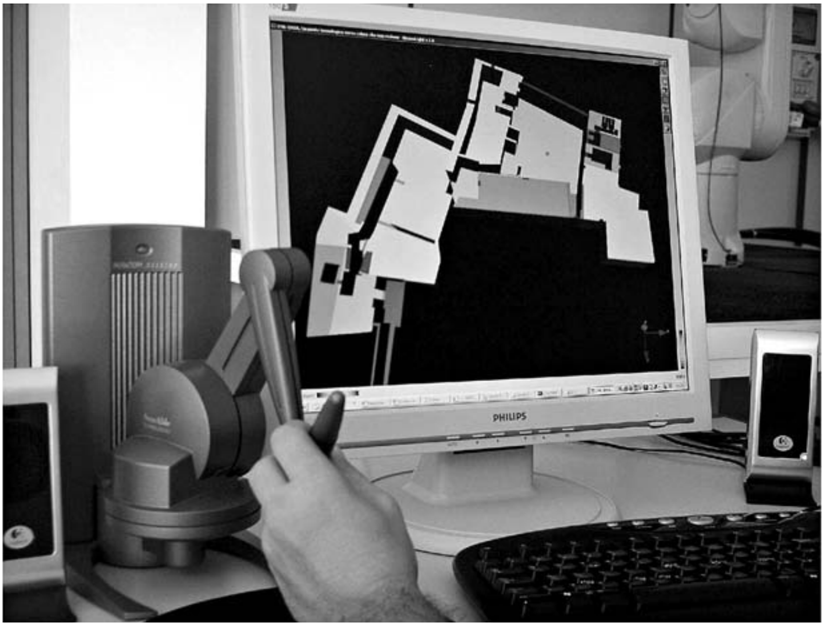

Fig. 3 – A blind person finds the well in the internal courtyard of the real castle after he had explored its corresponding zoomed virtual model.

Transit areas without doors between environments were modelled by bumps, defined as haptic/acoustic static objects. Users were warned of their presence by the vibrations they felt crossing these areas: an associated vocal message explained the rooms they connected. The synthesized messages were managed by a series of triggers associated with the touch of objects and with explicit user requests made by pressing the button on the haptic interface.

Users freely explored the model, providing their impressions during the whole visit. After a complete exploration of the model, they were asked to describe the spatial arrangement of the environments and to reach specific rooms. All four people in the first group were able to accomplish the proposed tasks in a satisfactory way. The main difficulties they had to deal with arose from the large number of objects placed around the model, the vocal messages frequently activated by touching them and the small size of some of the environments in the castle (Fig. 3).

During the second test session, the model was modified to represent only the location of the environments, keeping doors and bumps but removing any other object. In this way, users were able to explore the model focusing on the shape and arrangement of the rooms and on the haptic/acoustic interaction with doors and bumps without being annoyed by the objects and the related vocal messages. After that, users could focus their attention on the zoomed model of the environments they required, with all the corresponding objects. Again, all the users were able to fulfil the requests which were made to check their comprehension of the environment.

Both the test sessions with the haptic/acoustic system were followed by a real visit; in this way the visually impaired people could visit the castle making use of the information previously acquired and verify their mental scheme of the environments they had constructed during the virtual experience. Each blind person accomplished the real visit assisted by their preferred type of support (companion, cane or guide dog) to ensure their safety; in all cases, the blind person was able to autonomously determine the path through the castle without any help.

The main strategy used to explore the virtual environment was to follow its borders, in order to understand the related shape. Some users mentioned that objects located along the borders of the environments can act as effective reference points, while the most central objects are not important for this purpose. However, these objects are necessary for a complete comprehension of the site and the most active users were really intrigued to find them both in the virtual and in the real visit (Plate XI, a).

Users had some difficulties during the real exploration of the largest areas in the castle due to the fact that in this case it is not simple to recognize references and correctly match real and virtual environments in a short time. In any case, they commented that all the implemented features (haptic/vocal synchronization, change of the scale of the model, insertion/removal of details) allowed them to construct an effective mental representation of the environment and to explore it effectively.

Data show that users during the second test session were able to explore the virtual model in a faster and more effective way with respect to those of the first session; in fact, their capacity to recognize the location of the rooms during the real visits was also enhanced. This might be due to the more intuitive and direct interaction with a simpler model, characterized by a much smaller number of details and automatic vocal messages. Furthermore, the users made very frequent use of on-demand vocal messages, which are quite useful for improving the comprehension of the overall scene.

Some special cases occurred. A user, after the exploration of an enlarged version of the internal courtyard, returned to the global model to integrate the new pieces of information into the overall scheme. He made a very thorough visit of the real castle (for example he looked for and found the well in the internal courtyard) suggesting that a proper use of the system could really stimulate the curiosity of the most interested visitors. Some other users stated that they would be interested in a new virtual visit (following the real one) to refine and definitively assess their mental idea of the castle.

Blind users gave a very positive opinion of the system for exploring the layout of a building; having a complete and organic knowledge of the environment, acquiring an *a priori* understanding of what you can find, and having the possibility of autonomously planning the real visit was of real value to them.

4. The experiment with the Apulia model

Sites that are important from a historical and cultural point of view are scattered around the territory; therefore it is useful to acquire a proper knowledge of the territorial context in which the sites are located. To test the possibility of achieving this goal, a virtual model of the Apulia region was built and presented to visually impaired users.

The model, constructed on the basis of GIS data, is multi-layered; therefore it is possible to simply switch between several different versions, each containing a specific kind of information.

A view of this region describes the shape and the location of provinces, their borders and the borders between Apulia, the neighbouring regions and the sea. All the borders and the provincial areas were defined as acoustic active objects.

A second view reports the hydrographic network of the region: rivers and lakes were respectively represented as canyons and ditches in which the avatar can fall and move to provide perceptions about their course and shape. All these objects were defined as acoustic active objects associated with vocal messages telling their names.

Another view includes the location of the major towns (represented by little hexagonal prisms and defined as haptic active objects associated with an attraction effect that catches the user avatar whenever it falls in their vicinity: vocal messages inform the user about the name of the town he is touching). A support function has been introduced: by pressing the spacebar on the keyboard the avatar is pushed toward the nearest town, helping the user to find the closest place of interest.

The last view shows the main connections between towns. Roads are haptically represented as canyons connecting two towns and are acoustically active: they provide a vocal message about the name and kind of road as well as identification of the towns it connects. Vocal messages in this case are provided only on demand, made by pushing the stylus button.

This model was proposed to twenty visually impaired users, some of which did not have any previous knowledge about Apulia. They started the test exploring the first and simplest view of the model, in order to mentally acquire the shape of each province, of the whole region, of borders, names and relative positions of neighbouring regions. They then explored all the other views.

The spacebar was used very often, proving the usefulness of being driven by the system toward the nearest information point in the virtual scene whenever needed, especially during the first phases of the exploration (Plate XI, b).

Users who had had previous experience with the castle model appreciated the on-demand vocal messages, a modality which was also enthusiastically accepted by all the other users. Only two people were unable to complete the exploration of all the levels: one had very impaired manual dexterity while the other, with a low residual vision, was unable to use his vision since the monitor was not in the field of view but was not able to focus his attention on the haptic rendering and on the vocal messages. Another user experienced difficulties during the exploration of the various semantic levels of the region, due to his failure to identify valid reference points around which to build an efficient mental scheme of the territory.

Most of the users were able to correctly locate rivers, lakes and towns with respect to the regional and/or provincial territories and in relation to each other. Furthermore, after a complete and accurate exploration of all the semantic levels offered by the model, they were easily able to find any given object again.

The more curious and interested subjects were able to reach towns following a suitable pathway, even when they were very distant from each other. Moreover, people having some previous knowledge of Apulia easily found rivers, lakes, towns and paths between towns and were able to acquire new knowledge. All of them were able to learn new information about the region and found the haptic/acoustic interaction very stimulating.

5. Conclusions and future work

The paper presents a multimodal system (using visual, haptic and acoustic/vocal rendering of three-dimensional models) that allows the users to experience a virtual world involving different sensory channels; this means a richer and broader perception by sighted people but also an effective knowledge by visually impaired users. Several experimental sessions proved its effectiveness. Blind people virtually explored the ground floor of the castle in order to prepare a real visit: after the virtual use of the layout, were able to autonomously organize a walk through the real site, choosing freely which rooms to visit and the path to follow to reach them without requiring any planning or navigational support by sighted people. The possibility of adding or removing details and of increasing or decreasing the scale of the model has turned out to be useful for making the perception of complex environments easier.

The system was also used to explore a virtual model of the Apulia region. The model was organized in several different views, each providing different semantic contents (shape of provinces, hydrographic network, principal towns, main communication roads). The experiments confirmed that dividing the perception of complex entities allows a gradual and progressive cognition and helps the blind to create a mental scheme from their perceptions.

A few conclusions can be drawn. The multimodal system enables an effective interaction of blind people with virtual contents. The first result is to provide access to many digital contents that are already available but use of which was considered to be only for sighted people. Another result is to enable a more flexible interaction (varying scale, varying kind and amount of details, offering different semantic views of the same scene) that makes the cognitive process by the users easier and more effective.

The system can be very useful for improving the enjoyment of cultural heritage by blind people. The complete and profound comprehension of the cultural and artistic value of an object requires its connection with the proper context, composed by historical, geographic and cultural information. The system also allows visually impaired people to experience data involving large amounts of spatial information, which is expressed poorly and with difficulty using verbalized or written words.

Future work will address this specific point. Through closer cooperation with experts in the artistic and cultural fields we expect to create virtual models that can further improve the comprehension of the cultural treasures that are so widespread in our country. This will represent a significant step forward in the integration of the disabled into the community and offer an improvement in their quality of life.

Fabio De Felice, Floriana Renna,
Giovanni Attolico, Arcangelo Distante
Istituto di Studi sui Sistemi Intelligenti per l'Automazione
CNR – Bari

REFERENCES

Alonso F. 2006, *Design guidelines for audio-haptic immersive applications for people with visual disabilities*, in K. Miesenberger, J. Klaus, W. Zagler, A. Karshmer (eds.), *Computers Helping People with Special Needs. Proceedings of the 10th International Conference ICCHP (Linz, Austria, 2006)*, Berlin, Springer, 1071-1078.

Jacobson R.D. 2002, *Representing spatial information through multimodal interfaces*, in *Proceedings of the Sixth International Conference on Information Visualisation (IEEE IV '02)*, London, IEEE Press (http://doi.ieeecomputersociety.org/10.1109/IV.2002.1028858).

Magnusson C., Rassmus-Grön K. 2004, *A dynamic haptic-audio traffic environment*, in *Proceedings of Eurohaptics 2004 (Munich, Germany, 2004)*.

Massie T.H., Salisbury J.K. 1994, *The PHANTOM haptic interface: A device for probing virtual objects*, in *Proceedings of the ASME Winter Annual Meeting, Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (Chicago, Illinois, 1994)* (http://www.stanford.edu/class/cs277/papers/building1.pdf).

MURAI Y., TATSUMI H., NAGAI N., MIYAKAWA M. 2006, *A Haptic interface for an indoor-walk-guide simulator*, in K. MIESENBERGER, J. KLAUS, W. ZAGLER, A. KARSHMER (eds.), *Computers Helping People with Special Needs. Proceedings of the 10th International Conference ICCHP (Linz, Austria, 2006)*, Berlin, Springer, 1287-1293.

SALISBURY K., CONTI F., BARBAGLI F. 2004, *Haptic rendering: Introductory concepts*, «IEEE Computer Graphics and Applications», 24, 2, 24-32.

VAN SCOY F.L., BAKER V., GINGOLD C., MARTINO E., Burton D. 1999, *Mobility training using haptic interface: Initial plans*, in *Proceedings of Fourth Annual PHANToM User Group (Dedham, Massachusetts, 1999)*.

ABSTRACT

The paper describes a multimodal application, based on haptic/acoustic/visual interaction. A system of this kind offers two very interesting possibilities: it can be used to permit access and comprehension of cultural heritage by visually impaired people, for whom touch and hearing represent the main channels for interaction with the real world and it can also enhance the experience of cultural heritage by sighted people, making it possible to use the sense of touch, which is often forbidden in museum situations.

The system not only allows the experience of touching objects which, on account of their location, dimensions, and vulnerability cannot be offered for direct haptic contact, but, in addition, it makes it possible to experience a much more flexible and powerful interaction in complex situations. In fact, virtual models can change in a very dynamic and flexible way to match the needs of the specific user and to help his/her exploration and cognitive process. The system moves the haptic experience into the virtual world where the digital potentials can be used to make communication of the cultural content of each object more effective.

Multimodal interaction allows visually impaired people to access cultural heritage involving large spatial information content. The system makes it possible to interact with haptic/acoustic active objects and to select the information that must be shown on the basis of user requirements. Several tests, involving people with different types and levels of visual disabilities, were conducted. They showed that haptic/acoustic interaction and modular representation of information really do help blind people to cope with the serious and challenging task of acquiring and managing spatial data.