Multimodal virtual navigation of a Cultural Heritage site: the medieval ceiling of Steri in Palermo

S. Andolina[1], A. Santangelo[1], M. Cannella[2], A. Gentile[1], F. Agnello[2] and B.Villa[2]

[1]Dipartimento di Ingegneria Informatica, University of Palermo,
Viale delle Scienze Ed. 6, Palermo, Italy
santangelo@dinfo.unipa.it, gentile@unipa.it
[2]Dipartimento di Rappresentazione, University of Palermo,
Via Cavour 118, Palermo, Italy
fabrizio.agnello@unipa.it, bevilla@unipa.it

***Abstract* — The advance of information technology has enabled in recent years new fruition scenarios for cultural heritage sites. Multidisciplinary approaches integrate survey techniques with multimodal interfaces to allow enhanced fruition for larger group of users. In this paper we propose a multimodal interface to a virtual representation of a medieval ceiling, built in the XIV century, which covers the "Sala Magna" of Steri, the historical headquarters of the University of Palermo, in Italy. This research deals with the definition of a process for the integration of surveying techniques, modelling processes and communication technologies for the documentation of such artifacts. This is a two-stage process: in the first stage, a 3D digital document is produced that describes the artifact; in the second stage, a multimodal guide interacting with the 3D model is developed. A prototype multimodal guide has been implemented to narrate the Trojan Cycle, depicted on two of the rafters in the ceiling. This prototype was demonstrated to the general public and is detailed in the paper.**

***Keywords* — Cultural heritage fruition, survey techniques, multimodal interaction, human-computer interaction.**

I. Introduction

RECENT studies in the field of artificial intelligence and human-computer interaction have enabled the development of systems that can interact with the user more naturally in a variety of modes; with the use of these systems it is possible to generate, navigate and explore reconstructed environments of cultural interest, enabling the extended fruition of works of art that are not physically accessible to the user. Putting the user at the focus of attention leads up to the creation of cultural guides that can be easily consulted through a friendly interaction. In Cultural Heritage fruition systems, the use of virtuality and multimodality (i.e. using multiple modes to interact with the system) offers a trade-off between the need to preserve unchanged the authenticity of the heritage and the demand to make it aesthetically comprehensible and enjoyable. Furthermore this solution is useful to researchers in Cultural Heritage field, as it provides tools to study work of art otherwise not easily accessible.

Virtual Cultural Heritage studies thus deal with the use of tools and techniques to generate, navigate and explore reconstructed environments that are of cultural interest. Additionally, these techniques enable the extended fruition of works of art which are not physically accessible to the user (the object is no longer existent or not easy accessible, the user can only perform a remote access via internet, etc.).

Among systems that combine virtual representation with multimodal interaction, Cabral in 2007 created a virtual reconstruction of an old historic city in Brazil of the year 1911, which enables to explore, visualize and hear narrations about such site [1].

The aim of this work is to present a system that visualizes a 3D environment and provide the user with a multimodal interactive guide which assists the user during the navigation. Our goal is to enable access to visual representation and information contents about a cultural heritage site, providing visitors with a multimodal user-friendly virtual guide.

The paper the process to build such a virtual guide, from the initial survey to the final fruition interface. A prototype virtual guide has been implemented to assist visitors during their visit at the Sala Magna of the Steri, the historical headquarters of the University of Palermo. Specifically, a multimodal interface allows visitors to navigate the virtual representation of an ancient wood ceiling with tempera paintings and interact vocally with the system, which narrates artistic and historical descriptions of painted scenes, also answering visitors' questions. The entire system development process is herein described, from the initial survey, to the 3D photorealistic model production, and the multimodal interface realization.

Andolina, S.; Santangelo, A.; Cannella, M.; Gentile, A.; Agnello, F.; Villa, B., "Multimodal virtual navigation of a cultural heritage site: The medieval ceiling of Steri in Palermo," in Human System Interactions, 2009. HSI '09. 2nd Conference on , pp.562-567, 21-23 May 2009, http://dx.doi.org/10.1109/HSI.2009.5091039

The rest of the paper is organized as follows: section II deals presents a brief historical note about the Steri; section III gives an overview of multimodal interfaces in the cultural heritage field; in section IV, background information are sketched about the adopted technologies; section V describes the survey phase; in section VI.A, details of the overall system implementation are given; and section 0 describe the multimodal virtual guide. Lastly, some conclusions and future work are drawn in section 0.

II. An historical note about the site

The implemented prototype aims at assisting a user during a visit to the Sala Magna inside Palazzo Steri.

Historical headquarters of the University of Palermo, in Italy, the Steri was initially the residence of the Chiaramonte family, built starting from 1320 on the eastern edge of Piazza Marina, near the city's ancient harbour. The Chiaramonte family was chased away from the palace in 1392, and in the two following centuries the Steri was a royal and vice-royal residence. From 1605 to 1782, the Palace was the headquarters of the Inquisition Tribunal, a place of detention and torture. Graffitos, paintings and verses carved by prisoners on the cell walls have remained. These documents were discovered in 1906 during some restoration works, followed by the careful and patient transcription and interpretation work by Giuseppe Pitrè. Since 1984 the Steri has been the headquarter of the Rector's Office of the University of Palermo, that has commissioned the restoration works that have returned the building to its current state.

The largest room located at the second floor of the building is called Sala Magna and stands out on account of its beautiful tempera paintings. These fine painted decorations are made on wooden panels hooked to the vertical and horizontal faces of the beams of the ceiling of the room and are known as one of the most important masterpiece of Sicilian paintings of the fourteenth century, see Fig. 1.

Fig. 1. The wooden ceiling of the Sala Magna.

Some inscriptions in gothic characters report the dates of beginning and end of the decoration works, and the names of the painters.

Some paintings have geometric and floral patterns, whilst others portray scenes inspired to religious, mythological and heroic episodes. Scenes related to the same subject are developed on vertical beam faces, all heading to the same direction, in a sequence that goes, according to the point of view, from left to right, and from the background to the foreground. The partitioning of the panels of each beam was determined by the articulation of the narration into distinct scenes, according to a technique resembling strip cartoons. The implemented prototype has been developed to navigate some episodes of the Trojan cycle, see Fig. 2, on the vertical faces of beams VIII and IX, heading towards the South- Western wall. Each of the right and left portions of beam VIII are subdivided into three scenes, whilst the faces of beam IX are divided in two; the narration is thus articulated in ten scenes.

Fig. 2. Scenes from the paintings. Elena and Paride come to Troy.

The information about the context has been adapted from historical sources [6] by field specialists of the Regional Institute for the Catalogue and Documentation of Cultural Heritage.

III. Multimodal interfaces in cultural heritage

Multimodal access to information contents enables the creation of systems with an higher degree of accessibility and usability. In 2001, Cinotti presents MUSE (MUseums and Sites Explorer) [2], an experimental mobile multimedia system for information retrieval in cultural heritage sites, offering virtual reality and multimedia tours on e-book terminals.

A multimodal interaction may involve several human senses, such as touch and voice to navigate contents, or gestures to activate controls. The evolution of speech synthesis techniques supplies a new multimedia content, based on text-to-speech technology, which artificially generates human speech to provide the user with services accessible by telephone. This solution tries to overcome the well-known limits of static pre-recorded audio guides. For example, using this technology, Nickerson proposes the system named History Calls [3]. He developed an experimental system using VXML (Voice eXtensible Markup Language) technology capable to deliver an automated audio museum tour directly to cell phones.

The main target of some research groups is to make the visit within a cultural heritage site as more natural as possible for the largest part of users. In 2006, authors [4] propose a multimodal approach for virtual guides in cultural heritage sites that enables more natural interaction modes using off-the-shelf devices. The system, called MAGA, that integrates intelligent conversational agents (based on Chat-Bots) with speech recognition/synthesis, in a system framework using RFID based location Wi-Fi based data exchange.

The past two years see the advent of ambient integrated guides that are based on mixed reality, where real physical elements interact with virtual reality ones. In [1] Liarokapis and Newman focus their work on the design issues of high-level user-centered Mixed Reality (MR) interfaces. They propose a framework of a tangible MR interface that contains an Augmented Reality, a Virtual Reality and a Cyber Reality rendering modes.

Damala et al. in [6] present a prototype of an augmented reality enabled mobile multimedia museum guide, also addressing the full development cycle of the guide, from conception, to implementation, testing and assessment of the final system. They use last-generation ultra mobile PCs, which allow designers to exploit some of the most powerful software technologies, such as OpenCV for video acquisition, ARToolkitPlus for tracking the paintings, OGRE3D for the insertion of the virtual objects, Open AL for the audio output and XERCES for xml document parsing.

IV. Technological background

A. Survey techniques

Surveying is the technique and science of accurately determining three-dimensional space position of points and the distances between them. In this work three type of survey methods have been adopted, according to the application domain features: topographic, photogrammetric, and laser scanning methods.

Topographic method has been used to measure the coordinates of targets used in photogrammetric restitution and in laser scanning data processing. Photogrammetric method to obtain reliable information through photographic images. Laser scanning method to create a dense point cloud of the surface of the object and point clouds are used to extrapolate the shape of the subject.

B. The X3D format and the Xj3D project

The interactive virtual environments has been designed in VRML (Virtual Reality Modeling Language) and consequently translated in X3D (eXtensible 3D) [9]. The X3D is a new Open Standard developed by the Web 3D Consortium as evolution of the VRML, which provide developers with a component architecture more easily extensible. It is a format non-proprietary and open-source, based on XML and standardized by the ISO in 2004.

The X3D was born in the laboratories of the WEB3D consortium (downbranch of the W3C) which set itself the target of creating a language to address the issue of the interaction in multimedia applications, providing developers with a ad-hoc data structures.

The library used to support X3D technology is Xj3D. It is an open source (LGPL) project of the Web3D Consortium Source Working Group focused on creating a toolkit for VRML97 and X3D content written completely in Java. It serves a dual purpose of being an experimental codebase for trying out new areas of the X3D specification and as a library for application developers that want to build its own X3D-based project [Web3D site]. The X3D technology and the Xj3D project provides an open source framework to easily developed new applications with an embedded virtual world.

C. Vocal interaction

The "conversational computer" has been the scope of the research in the speech technology field supported by the artificial intelligence (AI). The goal is to allow users to interact with computer-based applications by using natural spoken language, transferring in computer interaction a specific skill of the human beings [8].

Today there are many technologies that allow the creation of multimodal interfaces, integrating multiple input (visual, vocal, touch etc.). Among these the voice interaction task is performed with Automatic Speech Recognition (ASR) and the Text-To-Speech (TTS) engines. Research on automatic speech recognition has been going on from the'50s.

The major division between the different types of ASR systems is made according to the dependence from speaker. The speaker-dependent systems are designed to meet the needs of an individual user but a training phase is always needed. In speaker-independent systems there is no training to recognize a particular speaker, but performance are less than in the speaker-dependent systems and the speech recognition process is carried out through an ad-hoc built-in speech grammar. The application domain of the presented work suggested the selection of a speaker-independent system.

TTS task refers to the artificial production of human speech. In contrast to ASR, the research on TTS is very close to its ultimate goal that is to convert ordinary orthographic text into an acoustic signal that is indistinguishable from human speech.

In this work, voice interaction tasks are based on the *Speaky®* technology platform, provided by *Mediavoice S.r.l.* [9] as part of the an on-going research collaboration program between it and the Dipartimento di Ingegneria Informatica.

V. From the survey to the X3D model

Metric data were acquired with topographic, photogrammetric, and laser scanning methods. The topographic survey, was used for different purposes: general dimensioning of the ceiling and survey of the hall; measurement of control points for photogrammetric restitution; measurement of targets used for registration and orientation of point clouds. Photogrammetric surveying was aimed at producing rectified images of the vertical and horizontal faces of the beams, and of an orthophoto of the whole ceiling. Laser scans were used for the acquisition of a remarkable number of 3-D points of the ceiling. The scanning operations were conditioned by the complex geometry of the ceiling. In order to reduce the holes corresponding to areas not reached by the laser ray, numerous scans were performed from different points inside the hall.

A first digital model of the ceiling was obtained from the laser point cloud through the automatic extraction of a triangularmesh. The distribution of triangles, conditioned by the step adopted in the scanning process, results quite homogeneous. Therefore, it happens that a plain surface is described with the same triangle density of a complex surface. This generates a model that is too detailed for simple geometry surfaces, and at the same time inadequate for the description of complex surfaces.

As in this work each face of a beam was substituted with the plane that best approximates the corresponding portion of the point cloud. Of course, this process involves a loss in definition. The adopted simplification has been consider compatible with an activity aimed at defining a single digital representation of the ceiling with its painted decorations.

The research on new forms of enjoyment of cultural heritage has set off ‘mapping’ the three-dimensional model. In the “Magna” hall, the paintings are not less important than the elements defining structure and morphology of the ceiling; for this reason the mapping process was cared not less than surveying and modelling.

The 3-D model thus obtained allows to observe the paintings in their spatial collocation; it gathers what has always been distinguished in literature, i.e. paintings – described with photos – and the physical support - described by graphic or physical models. The transformation of the model into a generic format for visualization (VRML, X3D) makes 3D exploration accessible also to users who do not have specific software, and who are not familiar with the CAD environment. In this format, the model can be manipulated and embedded in graphical interface aimed at simplify the user navigation and fruition task.

VI. The Multimodal Virtual Guide of the Ceiling

A. System implementation

The architecture of the proposed system is client-server. Fig. 3 shows its main blocks and their interconnections. The server side part of the system embeds the Automatic Speech Recognition (ASR) and Text to Speech (TTS) functionalities, implemented using the Speaky® technology.

The recognition task is based on a speech grammar, a text file which contains the expressions the system is able to identify. The client part is responsible for the control logic of the whole system and furthermore provides user with the User Interface, whose attributes are formalized into an xml file loaded in a xml database. The user can interact with the system using two methods: visual and vocal. In the first case his spoken input is processed by ASR and the result is sent to the client part of the system for subsequent operations. In the second case (visual input through traditional pointing devices) the user interacts directly with the client avoiding the recognition step.

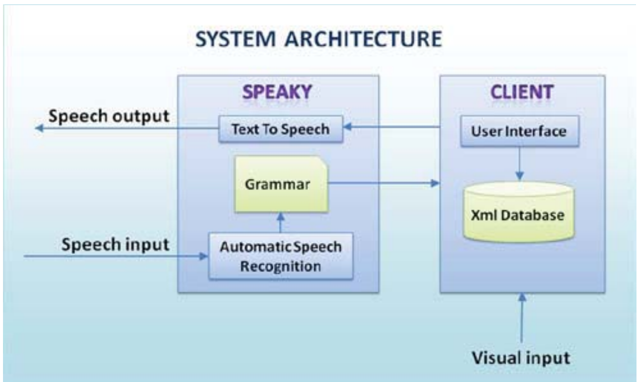

Fig. 3. System architecture.

B. Interaction paradigm

Below we will describe in detail the interaction paradigm. To accomplish this task it is useful to start from the description of the user interface, see Fig. 4. The buttons in the left panel are links to viewpoints corresponding to scenes of the Trojan cycle. The bottom panel shows the description associated to current scene, which is also pronounced by the TTS.

Fig. 4. User Interface

Clicking on a button the system starts the navigation through the virtual world moving from the current viewpoint to that selected by the user. Each command can also be given by voice, using sentences in natural language, and below we will refer to this latter case. Fig. 5 shows the main two states in which the system can be: *Navigation*, and *Tour*.

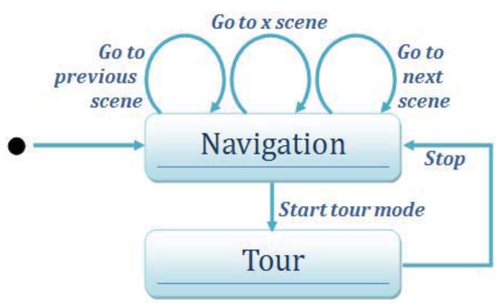

Fig. 5. State diagram

At the beginning the system is in the navigation state. In this state the user can perform three different operations:

- Ask the system to show the scene of interest (using the descriptions of the buttons on the left panel as a hint)
- Ask the system to show the next scene
- Ask the system to show the previous scene

The system will remain in the navigation state until the user says: “Start the tour” or the equivalent sentences in natural language. While the system is in the tour state the user can enjoy the virtual tour, moving from one scene to another, and receiving relevant information, without having to give any additional command. The utterance of a sentence such as: “Stop the tour” will result in the tour stop, the visual interface stops in the scene currently displayed and the system moves to the navigation state.

C. A typical scenario of the case study

In this section we present a typical scenario. Fig. 6 describes it using the common representation as a sequence diagram. In this scenario a user asks the system to show the “Helen’s kidnapping” scene. After the system’s response the user asks to view the next scene. A possible user-system conversation is what follow:

User: “Show me the Helen’s kidnapping scene”

System (*while browsing to new viewpoint)*: “Showing you Helen’s kidnapping.

System (*after the end of browse*): “The Trojans kidnap Helen and drag her to the ship. It follows a terrible battle in front of the temple…

User: “Go to next scene”.

System (*while browsing to new viewpoint*): “Now I expose to you the landing at the beach of Troy”.

System (*after the end of browse*): “After a storm, the ship landed at the beach of Troy, upheld by an emissary of King Priam. Paris and Helen fall from the catwalk…”.

VII. CONCLUSIONS

The aim of this work is the experimentation of innovative methodologies for the survey and fruition of Cultural Heritage. With reference to the problems of the survey, particularly, integration among different types of laser-scanning instruments and between laser-scanning and photogrammetric techniques has been proposed. With reference to the fruition task, a multimodal interface to access collateral information and navigate the virtual world by using voice interaction is presented. In the system, the accessibility is not only physical but also cognitive because the user can enjoy some informative contents related to the virtual environments that he/she is exploring.

A prototype virtual guide was presented to assist visitors during their visit of the Sala Magna of the Steri, in Palermo. The photorealistic 3D model of two of the rafters of the ancient ceiling was integrated with a visual and voice activated navigation interface, by which visitors navigate the painted scenes and see high-resolution close-up renditions of them, while hearing a narrated description.

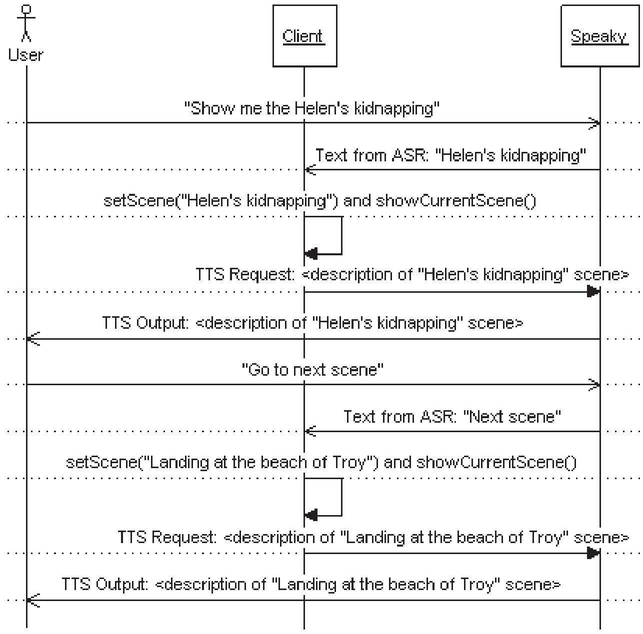

Fig. 6. A typical scenario

The proposed system can be easily adapted to generate, navigate and explore reconstructed environments of cultural interest, enabling the extended fruition of works of art. The next step of this work is to improve the interaction with the visitors, providing reasoning capabilities to the guide, extending the work presented in [4].

ACKNOWLEDGEMENTS

This work has been partially funded through the Programma di Rilevante Interesse Nazionale (PRIN) 2005, contract no. 2005103830_002, entitled: "*Artificial Intelligence Techniques for Processing Analysis, Preservation and Retrieval of Spoken Natural Language Archives*", and through University of Palermo funds, project ORPA06ZT43, entitled: “*Metodologie e tecniche per interazione multimodale in interfacce uomo-macchina*”.

The authors would like to thank Arch. Garofalo of Opera Officina per le Arti, an academic spin-off, for the insightful discussions and collaboration. Authors are also indebted to the specialists of the Regional Institute for the Catalogue and Documentation of Cultural Heritage, directed by Dr. Francesco Vergara Caffarelli, for the time devoted to build an enjoyable narration of the Trojan Cycle. Lastly, the voice recognition and synthesis has been enabled by Speaky® technologies, provided by Mediavoice S.r.l.

REFERENCES

[1] Marcio Cabral et all., “An experience using X3D for virtual cultural heritage”, *Proceedings of the twelfth international conference on 3D web technology*, (pp. 161 – 164), 2007, Perugia, Italy.

[2] Salmon Cinotti, T., Summa, S., Malavasi, M., Romagnoli, E., & Sforza F., (2001). MUSE: an Integrated System for Mobile Fruition and Site Management, *Proceedings of International Cultural Heritage Informatics Meeting, ICHIM01*, Milano September 2001.

[3] Nickerson, M., (2005). All the World is a Museum: Access to Cultural Heritage Information Anytime, Anywhere, *Proceedings of International Cultural Heritage Informatics Meeting*, ICHIM05, Parigi September 2005.

[4] Augello, A., Santangelo, A., Sorce, S., Pilato, G., Gentile, A., Genco, A. & Gaglio, S. (2006). MAGA: A Mobile Archaeological Guide at Agrigento, *Proceedings of the workshop "Giornata nazionale su Guide Mobili Virtuali 2006", ACM-SIGCHI* Italy, Torino, 18 ottobre 2006, online proceedings http://hcilab.uniud.it/sigchi/doc/Virtuality06/index.html.

[5] Liarokapis F., & Newman,Design, R. M. (2007). Experiences of Multimodal Mixed Reality Interfaces, *Proceedings of the 25th annual ACM international conference on Design of communication*, El Paso, Texas, USA, Pages: 34 – 41.

[6] Damala, A., Cubaud, P., Bationo, A., Houlier, P., & Marchal, I. (2008). Bridging the gap between the digital and the physical: design and evaluation of a mobile augmented reality guide for the museum visit, *Proceedings of the 3rd international conference on Digital Interactive Media in Entertainment and Arts* (DIMEA '08), Athens, Greece, Pages 120-127

[7] Bologna Ferdinando, *"Il soffitto della sala magna allo Steri di Palermo"*, ed. Flaccovio, 2002, (pp. 179-195)

[8] Michael F. and McTear *"Spoken Dialogue Technology Toward the Conversational User Interface."* (University of Ulster) London Springer-Verlag, 2004

[9] http://www.mediavoice.it/