Improving access of elderly people to real environments: a semantic based approach

Fabio Pittarello
Università Ca' Foscari di Venezia
Dipartimento di Informatica
Via Torino, 155
+39 041 2348435
pitt@unive.it

Alessandro De Faveri
Università Ca' Foscari di Venezia
Dipartimento di Informatica
Via Torino, 155
+39 041 2348435
adefaver@dsi.unive.it

ABSTRACT
Access to real environments is often conditioned by a number of issues, including the skills of the user (i.e. affected by aging, physical and psychological deficiencies, etc.) and the complexity of the real environment itself. This work proposes an approach for helping users with different skills, including elderly people, to navigate through complex real scenes; such approach is based on the semantic description of the objects and zones that characterize the environment itself and takes advantage of an implementation architecture based on web standards for generating navigational support. A case study related to the creation of a guided tour through the indoor and outdoor locations of the city of Venice, accessible through a multimodal web browser, is presented.

Categories and Subject Descriptors
H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing – *Indexing Methods*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *Graphical User Interfaces (GUI), Voice I/O*; I.3.7 [**Computer Graphics**]: Three-Dimensional Graphics and Realism – *Virtual Reality*; K.4.2 [**Computers and Society**]: Social Issues – *Assistive Technologies for Persons with Disabilities.*

General Terms
Design, Human Factors.

Keywords
Elderly People, Multimodality, Navigation, Semantic 3D Environments, XHTML + Voice Profile.

1. INTRODUCTION
Everyday experience teaches us that access to 3D environments is not always straightforward, but can be conditioned by a variety of issues, ranging from the skills of the user to the complexity of the environment itself. In particular there are categories of users whose access to 3D environments is conditioned by physical and cognitive deficiencies that prevent them from recognizing easily locations, objects and their relations; such situation often leads them to the inability of having information about objects of interest, recognizing situations of dangers or taking advantage of support tools (e.g., parapets, traffic-lights, pedestrian crossings, etc.) for moving across the environment.

That is the reason why this work proposes an approach for helping people to navigate the real environments, considering different levels of assistance in relation to the skills of the users: people with normal physical and cognitive faculties and elderly users with visual and cognitive deficiencies.

The solution proposed takes advantage of the semantic description of the environment, mapped on a virtual 3D counterpart, for presenting to the users a multilevel and multimodal description of the objects and zones that characterize the real environment. Such multilevel description is useful for navigational needs (i.e. it allows the users to understand the relations of a given object or zone with the environment), but it gives also an answer to specific information needs (e.g., a user might be interested on investigating the architectural details of a building or the technical features of a work of art contained inside an exhibition room); besides, such description is coupled with the accurate underlining of the dangers, the navigational aids and the steps' sequence for moving through the environment itself.

The rest of the paper is organized as follows: Section 2 considers related works; Section 3 describes the user requirements guiding our assistive solution; Section 4 focuses on the definition of the semantics for a 3D virtual counterpart mapping the real environment; such description is used by an implementation architecture for extracting and presenting navigational information in the form of a multimodal guided tour; Section 5 focuses on a case study related to the urban environment of Venice and draws the final conclusions.

2. RELATED WORKS
The issue of pedestrian navigation through a complex real environment has been considered by a large number of research works that investigate the different aspects of the problem.

Some works focus on the general requirements for navigational aids: an interesting study [6] analyzes the nature of the navigation task and the information needed by users within an urban navigation context, showing that landmarks (i.e. visible objects of different sizes that help the user to orient herself /himself inside the environment) are the most important navigation cue.

AVI'06, May 23–26, 2006, Venezia, Italy.

Other papers focus on the specific needs of different categories of people. Hub et al. [3] point out the importance of informing blind people navigating a real environment about the specific features of the scene that may help the exploration (e.g., the size of rooms) or situations of danger that may cause severe injuries (e.g., stairs, revolving doors, etc.). Touchable landmarks, which can be any kind of object including light switches or thresholds, are considered as an important aid for helping such category of users. Zajicek [14] investigates interface design issues for older adults; while her work is not directly concerned with navigation in real spaces it puts in evidence specific disabilities that typically affect older adults over 70, including memory and visual impairment; such analysis will be considered in our work for deriving some important user requirements.

Concerning interface solutions, several speech audio solutions have been proposed for blind users [3, 4, 11]. Ross et al. [11] propose a multimodal speech and tactile interface for blind people tested in an urban context that has shown great usability and flexibility for a wide number of people. Our proposal shares with this work the multimodal approach, even though different senses are used to convey information.

Many outdoor navigational applications for blind people use GPS for determining the position of the user and giving her/him appropriate information. Indoor solutions rely on different technologies (e.g., infrared beacons, WiFi, WLAN based location, etc.) to compensate for the lack of the GPS signal. An interesting indoor solution [3] uses a portable sensor equipped with positioning and direction devices for identifying the shape of a given objects inside the room and then match it with the components of a virtual model that maps the environment. A speech interface presents to the user descriptive information associated to the virtual object mapping the real one. This solution doesn’t offer to the user explicit direction for navigating through the environment (i.e. it is more focused on the identification of objects), but it shares with our approach the idea of mapping the real environment with a virtual counterpart including also high-level descriptions of the objects contained into the scene.

Although semantic description of the 3D scene can be useful for deriving high-level information for navigational purposes, it is not a standard feature for 3D environments: often 3D worlds are modeled using low-level primitives (e.g., spheres, cones, etc.) that don’t contain any reference to the semantics of the objects they represent. Besides, the scene graph used for storing a 3D scene in a hierarchical order is not suited to store semantic information that often has a complex structure. Recent approaches [2, 5] resolve this problem by taking advantage of MPEG-7 standard for storing semantic information related to a 3D world in an external data structure. The current limit of such solutions is that the ontologies used for describing the semantic objects and their relations are scene-dependent. Such feature prevents the progressive building of a set of urban scenes that can be merged and queried using a shared semantics. That is the reason why our work includes the definition of a scene-independent solution.

Besides, most proposals for the semantic annotation of 3D environments focus on the definition of semantic objects rather than on semantic zones. While a high-level description of the geometry of a given environment is effective for cataloging the parts the 3D scene is made of and can be very useful for searching semantic objects at different levels of granularity, such annotation is not complete because it doesn't take explicitly into account the use of space. The Interaction Locus (IL) concept [8], defined by one of the authors of this work, satisfies the need for an explicit definition of the partitions of space that are designed and recognized as morphologic and functional units, and will be considered in Section 4 as a component of the scene-independent ontology.

3. DEFINING THE USER REQUIREMENTS

Our proposal aims to support the users through the different phases of the navigation task, including position identification, recognition of the relations with the surroundings and navigation properly said. Besides, the proposal aims to satisfy additional information needs about the objects that populate the environment, giving the user the opportunity to have more details.

The specific requirements for the presentation of navigational information are derived from the analysis of related works and take into account both the needs of normally sighted people and elderly users:

- description of semantic objects that characterize the environment, in order to give the users the opportunity to make a match with what they see and use them as local landmarks; only objects local to the current zone are described, avoiding to inform the users about objects belonging to other locations (e.g., a skyscraper visible at a distance) that probably can hardly be seen because of the reduced visual acuity of some users;
- multilevel description, for giving the users the opportunity to satisfy detailed information needs (e.g., details about a work of art or architectural elements) or to relate the semantic zone they are visiting with higher-level semantic components of the environment (e.g., the building the current room belongs to);
- availability of the output both through visual and speech interfaces and possibility for the user to switch seamlessly between the two different communication channels (i.e. elderly people may begin the tour by reading information and then switch to the speech engine when they are tired or it becomes less clear);
- possibility to communicate the input both through forms on the graphical interface and a voice-recognition engine, in order to meet the different preferences or to allow to overcome difficulties due to the specific environment or other conditions (e.g., traffic noise or difficulty to use hands in some circumstances);
- timely identification of dangers (e.g., staircases, sliding doors, etc.) and navigational aids (e.g., rails, chairs for having some rest, etc.);
- description of the navigation steps inside the current zone;
- possibility to repeat relevant information; such requirement is introduced for the speech interface to compensate both for the limits of the speech output and the reduced capacity of the elderly people to retain information in the short term memory;

- concerning voice interface, clear specification of the words that the user has to say when s/he has to choose among different opportunities, such as having more information about semantic objects X, Y and Z or proceeding with navigation; such requirement reduces greatly the uncertainty related to the speech input.

Our approach for supporting navigation doesn't use sophisticated methods for monitoring the position of the user, but relies on the (even residual) sight of the user herself/himself for moving through the environment following the indications of the guided tour. Therefore our approach is not suited to blind persons that don't have the possibility to match information received from the system with what they see. Such simplicity, while requires a more active behavior from the users, helps its application in scenarios, like the one described in the case study, characterized by navigation both inside indoor and outdoor locations; in such cases precise tracking would require a costly and complex technological framework coordinating GPS monitoring with ad-hoc solutions for the interior of the buildings. Besides, the specific urban situation analyzed in the case study (i.e. the city of Venice) is characterized by narrow pedestrian streets were the GPS signal can't be perceived; a precise tracking of the user would augment greatly the technological requirements, requesting the implementation of an additional costly urban network of sensors placed in public outdoor spaces.

4. EXTRACTING NAVIGATIONAL INFORMATION

As stated in Section 1, our approach for improving access to real environments is based on the accurate mapping of the real scene with a virtual model enhanced with semantic information; such model will be queried by the components of an implementation architecture for deriving navigational information about the real scene.

The technical details of the semantic specification for virtual environments, complemented with a number of code examples, are fully described in [9]. Given the limits of space, we'll focus on its high-level features:

- the definition of a complete multi-level description including not only the 3D semantic objects but also the semantic zones that characterize the scene; such approach takes explicitly into account the use of space as an important element for a complete description of the scene and it is particularly useful for deriving navigational information;
- the definition of a scene-independent reusable ontology for easing the creation of a growing set of virtual models mapping the real world based on the same semantics;
- the reference to existing web standards (i.e. X3D [13] for 3D, RDF and OWL for the semantic web [12]), in order to maximize the generality of the solution, to minimize the effort related to the creation of tools for manipulating data and to ease the conversion of existing 3D environments mapping the real world.

Scene information is split across two different files:

- the X3D file, describing a specific environment: it contains the description of the geometry of the 3D world and the metadata for adding semantic information; also semantic zones are specified for completing the description of the environment; such semantic information is specified in relation to the ontology defined by the RDF Schema file (see below);
- the RDF Schema file, for storing the scene-independent reusable ontology, containing the relations between the different classes of semantic objects; such relations are valid for a specific domain (e.g., the built environment).

The user requirements defined in Section 3 have strongly influenced the definition of the scene-independent ontology for urban environments; such ontology is characterized by a variety of relations among the different semantic objects, including: *containment*, *sharing*, *bounding*, *subclass*, *synonymy*, *danger* and *aid*. Some relations, such as *danger* and *aid*, are specified in relation to one or more classes of users (i.e. the classes of users are specified in a separate ontology and represent profiles such as *elder*, *blind*, etc.).

The semantic description of the environment is used as the main input of an implementation architecture for the extraction, elaboration and presentation of navigational information. Users may specify the tour they are interested in, using a simple low-level path based on a polyline (i.e. different techniques can be used for doing that, from sketches on a map representing the environment to selections from a set of suggested paths authored by content experts), and the system will show the results in the form of a multimodal tour satisfying the user requirements defined in Section 3. Such architecture represents an enhancement of the proposal illustrated in [9], because it takes into account also the skills characterizing the different classes of users for generating appropriate information presentation. Given the limits of space, we'll focus on the multimodal output provided by such architecture. A specific module implemented for this work permits to output the content in the form of a set of files compliant with the XHTML + Voice profile (X+V) [1]. X+V is a standard designed for creating multimodal dialogs that combine the visual input and output, represented by XHTML, and the speech input and output, represented by a subset of VoiceXML 2.0. The user may take advantage of both communication modalities, according to her/his preferences. Such files are published on a web server and can be browsed by any browser compliant with the X+V specification. In our case study we used the Opera multimodal browser [7], requiring a standard laptop (or sub-notebook) running Windows XP with a sound card for voice I/O.

5. A CASE STUDY: NAVIGATING THROUGH VENICE

This section describes the application of our approach to a case study related to the navigation inside the city of Venice, an urban environment including indoor locations with a complex organization and a variety of outdoor spaces including squares, streets, bridges and channels. Figure 1 shows a scheme illustrating the fragment of the real city that has been considered for the case study: a Venetian square connecting different buildings, pedestrian streets and water channels. The red line starting on the upper left corner of Figure 1 and ending on the water boat stop represents the sample low-level path that has been used in this case study, together with the semantic description of the environment, for the generation of the guided tour.

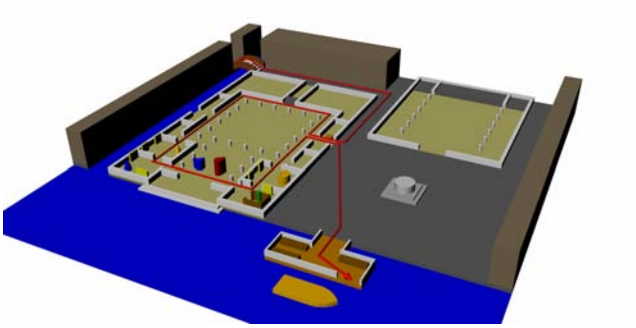

Figure 1. A scheme representing the Venetian environment.

Each X+V file generated by the implementation architecture for the presentation on the multimodal browser is associated to a specific semantic zone or object. Navigation information is presented in a sequence compliant with the user requirements and includes:

- a declaration of the semantic zone the user is visiting;
- a list of the dangers that characterize the current zone and the adjacent locations (the user must be aware of all the local dangers, even though the surroundings are not directly traversed by the path);
- a list of the navigation aids for the current semantic zone;
- a multi-level description of the zone; the description can be *zoomed out* to describe the spatial context that contains the current location or *zoomed in* leading to additional X+V files describing the objects contained inside the zone and their semantic components in depth;
- the navigation steps for traversing the location; when a given navigation step leads the user near to a specific danger, the description of such step comes after a warning repeating information about such danger and local aids, if available; such repetition corresponds to the criteria of supporting, whenever possible, the reduced cognitive faculties of elderly people;
- a link to information related to the next zone (i.e. such information is put also at the beginning of the output to allow users not interested in the current zone to switch to information related to the next location).

Besides, the interface includes functions that apply only to specific output channels; for example the visual output includes the possibility to switch to a high-contrast presentation, while the speech output includes the possibility of repeating the list of possible choices.

The first example, illustrated in Figure 2, focuses on navigation inside an exhibition room that is part of the Venetian building. The room includes works of art, a chair for sitting and it is connected to a corridor and to another room by two doors. Besides, the room is characterized by an irregular floor and a damaged threshold.

Figure 3 shows a snapshot of the browser illustrating the visual output for such room. The presentation of information is compliant with the sequence illustrated before. Link to information related to the next room is available at the beginning of the information blocks for being readily available to visitors not interested in the current room.

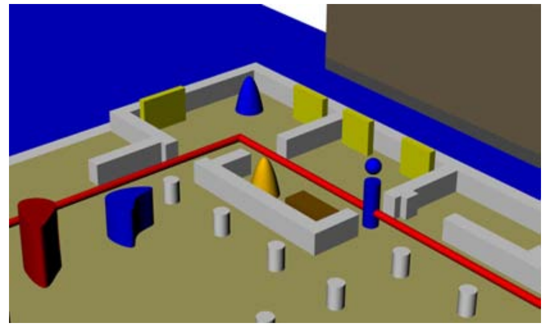

Figure 2. Navigation inside the Venetian palace.

The different information blocks are associated to specific icons and background colors in order to augment their identification. In particular the red background emphasizes the presence of dangers (e.g., the irregular floor and the ruined threshold) and navigational aids (i.e. a chair for having some rest). Nested information resources are linked using the standard XHTML underlining.

The description of the navigation steps includes values such as *go straight* followed by a measure of distance expressed in meters; such information is not directly available in the semantic description of the environment but it is computed by one of the implementation architecture components during the generation of X+V files, on the basis of the X3D geometry and the specific low-level path. The description of the semantic objects (e.g., work of art, wall, etc.) includes an annotation (e.g., left, in front of, etc.) about their position in relation to the user; again such annotation is computed by the implementation architecture on the basis of the user position and orientation along the low-level path.

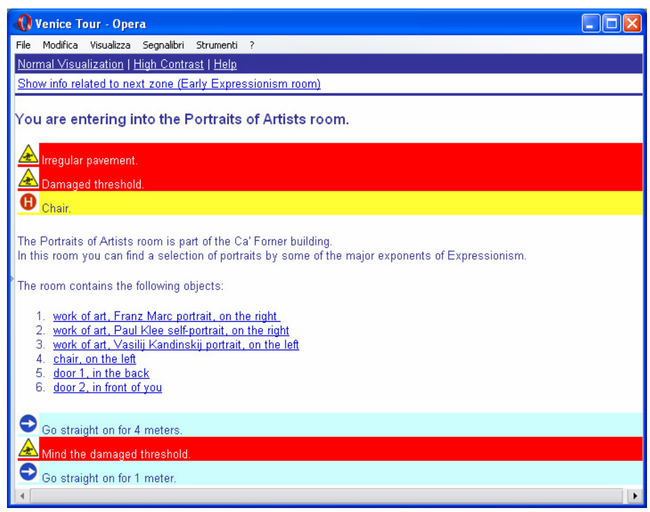

Figure 3. A snapshot of the interface for elderly people.

The user may take advantage of the speech interface as an alternative to the visual interface; a sample dialog, related to the example illustrated in Figure 2, is shown in the following:

```
Computer: You are entering into the Portraits of Artists room.
[bip 1] Irregular floor. ...
Would you like some information about this zone?. Say 'Yes' or 'No'.
Human (speaking): Yes
```

Computer: The Portraits of Artists room is part of the Ca' Forner building.

In this room you can find a selection of portraits by some of the major exponents of Expressionism.

The room contains the following objects:

- one, work of art, Franz Marc portrait, on the right ...

If you want information about an object of the list say the number of the object otherwise say 'No'; if you want to hear again the list of the objects say 'List'.

Human (speaking): One

Computer: (proceeds to marc03.vxml)

As you can see from the interactive dialogue, the audio version is characterized by the possibility of requesting a repetition of relevant information, in order to compensate for the sequentiality that characterizes the speech output. Besides, in order to attract the user attention about safety issues, the information items related to dangers and aids are presented after a non-speech audio signal (i.e. *bip1*) associated to such information categories.

Finally, the system will present a different visual output in association to the standard user profile; in such case the user doesn't need detailed navigation assistance and may be tired of the continuous references to safety issues. Therefore the interface will lack the information blocks related to danger and aids.

The second example focuses on outdoors navigation. The system supports the user accessing the square, navigating from the entrance of the building (see Figure 1) to the boat stop on the channel. The environment and its surroundings present a number of dangers: the irregular floor of the square, the mobile floor of the boat stop and the water channel. The user can count on a rail for accessing the boat stop. Figure 4 illustrates a fragment of the high-contrast version of the interface, characterized by the definition of larger font sizes and bigger icons. Such version can be selected clicking over the *High contrast* link placed in the interface toolbar.

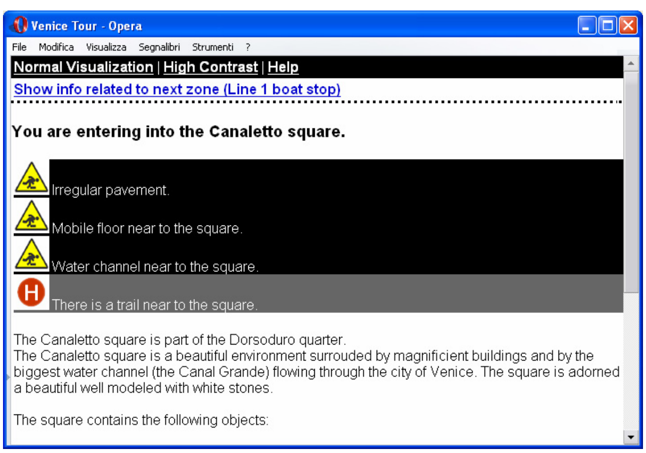

Figure 4. A snapshot of the high-contrast interface.

Concluding, the potentialities of the semantic annotation for helping navigation of real environments have been exploited in a case study related to the urban centre of Venice. While preliminary tests are encouraging, future work will include the evaluation of the solutions adopted; besides we'll consider also the possibility of porting the approach to different presentation platforms (e.g., palmtops and smart-phones) that, although are characterized by additional problems for the visual output (i.e. screen size), are more diffused and less expensive.

6. REFERENCES

[1] Axelsson, J., Cross C., Ferrans, J., McCobb, G., Raman, T. V., and Wilson, L. XHTML + Voice Profile, March 2004, http://www.voicexml.org/specs/multimodal/x+v/12/

[2] Halabala, P. Semantic Metadata Creation. In *Proceedings of CESCG 2003: 7th Central European Seminar on Computer Graphics*. 2003, 15-25.

[3] Hub, A., Diepstraten, J., and Ertl, T. Design and development of an indoor navigation and object identification system for the blind. In *Proceedings of the 6th international ACM SIGACCESS conference on Computers and accessibility (ASSETS'04)* (Atlanta, GA, USA, October 18-20, 2004). ACM Press, New York, NY, 2004, 147 – 152.

[4] Loomis, J. M., Golledge, R. G., and Klatzky, R. L. *GPS-based navigation systems for the visually impaired*. In: Barfield, W., and Caudell, T. (eds) Fundamentals of wearable computers and augmented reality. Lawrence Erlbaum, Mahwah, NJ, 2001, 429–446.

[5] Mansouri, H. *Using Semantic Descriptions for Building and Querying Virtual Environments*. MSc Thesis, Vrije Universiteit Brussel, 2005.

[6] May, A.,J., Ross, T., Bayer, S., H., and Tarkiainen, M., J. Pedestrian navigation aids: information requirements and design implications. *Personal and Ubiquitous Computing*, 7, 6 (December 2003), 331 – 338.

[7] Opera Multimodal Browser, http://www.opera.com/products/verticals/multimodal/

[8] Pittarello, F. Accessing information through multimodal 3D environments: towards universal access. *Universal Access in the Information Society Journal*, 2, 2 (2003), 189-204.

[9] Pittarello, F. and De Faveri, A. Semantic Description of 3D Environments: a Proposal Based on Web Standards. In *Proceedings of Web3D 2006: 11th International Symposium on 3D Web Technology* (Columbia, Maryland, USA, April 18-21, 2006). ACM Press, New York, 2006.

[10] Raggett, D. Getting Started with VoiceXML 2.0, http://www.w3.org/Voice/Guide/

[11] Ross, D., A., and Blasch, B., B. Wearable interfaces for orientation and wayfinding. In *Proceedings of the fourth international ACM conference on Assistive technologies (ASSETS'00)* (Arlington, Virginia, USA, November 13-15, 2000). ACM Press, New York, NY, 2000, 193 - 200.

[12] W3C. Semantic Web, http://www.w3.org/2001/sw/

[13] Web3D Consortium. Extensible 3D (X3D) ISO/IEC 19775:2004, http://www.web3d.org/x3d/specifications/ISO-IEC-19775-X3DAbstractSpecification/

[14] Zajicek, M., Interface Design for Older Adults. In *Proceedings of the 2001 EC/NSF workshop on Universal accessibility of ubiquitous computing: providing for the elderly (WUAUC'01)* (Alcácer do Sal, Portugal, May 22-25, 2001). ACM Press, New York, NY, 2001, 60 – 65.