Project ref. no.	FP6-507609
Project acronym	SIMILAR
Deliverable status	R
Contractual date of delivery	30 November 2004
Actual date of delivery	27 October 2004
Deliverable number	D17
Deliverable title	Descriptions of SIMILAR Applications
Nature	Report
Status & version	Final
Number of pages	71
WP contributing to the deliverable	SIG7
WP / Task responsible	NISLab
Editor	Laila Dybkjær
Author(s) (alphabetic order)	Niels Ole Bernsen, Silvia Berti, Laila Dybkjær, Monica Gemo, Enrique Gómez, Pedro Correa Hernandez, Pablo Lamata, Benoit Macq, Xavier Marichal, Vincent Nicolas, Georgios Nikolakis, Fabio Paternò, Samuel Rodríguez, Carmen Santoro, Michael G. Strintzis, Daniela G.Trevisan, Dimitrios Tzovaras, and Jean Vanderdonckt
EC Project Officer	Mats Ljungqvist
Keywords	Applications, functionality, usability, usability evaluation.
Abstract (for dissemination)	This report describes a set of applications made available by SIMILAR partners. Each application description follows the same template. Focus is on functionality and usability including evaluation.

Deliverable D17

Descriptions of SIMILAR Applications

HIIS Laboratory, ISTI – CNR
NISLab, University of Southern Denmark
Tele, Université catholique de Louvain, Alterface SA
ITI-CERTH
GBT, Polytechnic University of Madrid
TELE and BCHI – Université Catholique de Louvain

September 2004

SIMILAR Network of Excellence – Research in Human-Machine Multimodal Interfaces
www.similar.cc

Contents

1 Introduction

This report contains descriptions of a set of applications made available to the SIMILAR usability and evaluation SIG (SIG7) by SIMILAR partners. Since the focus of the SIG is on usability and evaluation, the focus in each description is on functionality and usability including how evaluation has been done.

In the following we briefly describe the approach we have used to produce the present report, and we provide an overview of the applications described and of the common description template used.

1.1 Approach

The approach adopted for producing the present report was to (i) first identify a set of applications which SIMILAR SIG7 members had access to and were willing to describe; then (ii) establish a common template for describing each application; and finally (iii) produce descriptions following the template agreed upon.

1.2 Applications

Seven applications were identified and described by SIMILAR SIG7 members. Figure 1 provides an overview of these applications. The last application (Medical Studio) is not described in as much detail as the other applications due to its present state.

Application	**Type of application**	**Description provider**
Portable Cicero Application	Portable museum guide	HIIS Laboratory, ISTI – CNR, Italy
NICE Hans Christian Andersen	Conversational edutainment system	NISLab, University of Southern Denmark
AlterStationTM System	Gesture-based interactive entertainment system	Tele, Université Catholique de Louvain, Alterface SA, Belgium
Training System for Blind People	Training system for blind people, e.g. cane simulation application	ITI-CERTH, Greece
SINERGIA laparoscopy virtual simulator	Surgery system	GBT, Polytechnic University of Madrid, Spain
Image Guided Surgery	Surgery system	TELE and BCHI – Université Catholique de Louvain, Belgium
Medical Studio	Surgery system	TELE - Université Catholique de Louvain, Belgium

Figure 1. Applications described.

1.3 Common description template

In order to obtain comparable descriptions and collect approximately the same information for each application we have used the structure or template shown in Figure 2 for each application description. The template has seven main entries. Each main entry subsumes a number of more specific information items.

Introduction

- Purpose of the application
- Input modalities
- Output modalities
- Target user group(s)
- Physical use environment
- Which domain does the application cover
- Which tasks (if any) does the application solve
- Is the application free or what is the price
- If not free, is a demo available?

Technical issues

- Platform(s) (operation system(s))
- Hardware requirements
- Implementation language(s)
- Architecture

Functionality

- Which functionality does the application offer
- Description of each main functionality

Interface and usability

- Description of interface design and possible design for usability
- Which user skills (if any) are assumed
- Is it walk-up-and-use? Is training foreseen? Is there a manual?
- Illustrative examples of use and interface
- Advantages and disadvantages of functionality and interface

Evaluation

- Who/how many have used the application so far
- How has the application been usability tested (detailed description of methods and criteria)
- Which evaluation results are available so far (including references to where they are documented)

Conclusion

- General assessment of usability and functionality of the application. Please make clear what is the basis for the assessment.

References

- Any references which provide more information about the application

Figure 2. Common description template.

1.4 Conclusions

This report simply presents the descriptions of the applications listed in Figure 1. It does not attempt to draw any conclusions regarding usability and evaluation across these applications. Deliverable D16 on state-of-the-art and current practice, however, provides an overview of usability evaluation issues in natural interactive and multimodal systems and includes a section on usability and evaluation in the applications presented in the present report. We therefore refer the reader to D16 for conclusive remarks regarding the applications included in D17 (this report).

Description of the Portable Cicero Application

Silvia Berti, Fabio Paternò and Carmen Santoro
HIIS Laboratory, ISTI - CNR

July 2004

1 Introduction

The growing availability of small devices (PDA or mobile phone) whose computational and interactive resources are continuously increasing in terms of power and capacity has raised an interesting discussion on how to exploit them to support users in various contexts of use. This document describes the Portable Cicero application which is a handheld museum guide being developed as case study of a research regarding interaction between human and mobile devices in which the mobility of the user, the new characteristics of the devices involved (small screen, no keyboard or mouse) and the new input devices (e.g. voice, gestures and location of the users) open new scenarios of use and urge the old paradigms of interaction to be rearranged for guaranteeing the usability of the new systems.

In this report we describe the implemented single-user version of the system. A cooperative version of the system is also available though it is still at a prototype level and supports cooperative visits including cooperative games aiming to improve the user experience.

1.1 Purpose of the Application

Portable Cicero is a location-aware handheld guide whose main goal is to support museum visitors during their visit to the Marble Museum of Carrara so as to improve their experience by providing users with information related to the current context by using both the visual and audio channels.

1.2 Input Modalities

Communication between the user and the PDA application is carried out through the pen. Moreover, the system receives as an input the signals from infrared emitters installed at the entrance of each room of the museum in order to localize the users in the Museum so as to provide them with context-aware information that takes into account the users' position.

Once the system has detected the room the user is in, then the user can freely activate audio comments regarding the artworks of interest.

1.3 Output Modalities

Audio and video file are provided with extra textual information to completely support the users in their visit. We decided to use text-to-speech synthesis for supporting audio comments. The application provides audio information in either Italian or English obtained differently (human voice for Italian visitors and text-to-speech engine for English) to verify if the different techniques for audio presentation can lead to a different quality of the visit. We used audio-recorded comments (human voice) for Italian visitors because the synthesized Italian voice was considered too unpleasant. The system was originally designed to provide audio information with headphones but the managers of the Museum noticed that visitors prefer to use the application sharing it in a small group (two or three visitors) without headphones, considering acceptable the mix of the audio presentation from different guide in the same section as this can happen rarely give the limited number of PDAs available.

The system exploits both the audio and the visual channels by providing two kinds of feedback:

- *Audio feedback*: in our application, we decided to highlight the automatic detection of the room where users are entering, obtaining two results: the first it to signal this event to the users who might not be looking at the PDA display; the second reason is to assure the visitors that the system is aware that the context has changed and that the information is related to the new section. Moreover, to signal the event of entering new room, a sound is generated as the association of a sound to the section change helps them to get oriented in the museum. In order to reduce the difficulties for the users to interact using both hands with the mobile devices while they are moving, in our system as soon as the user enters a new section the application immediately starts a vocal comment to present the new section. Each page is also associated with a voice comment automatically started the first time it is accessed. Furthermore, to improve the support to the users during the visits, we have also added audio feedback when the users select artworks on the section map to assure users that the system has received their input. We adopted this solution because we noticed that, after selecting an artwork, users often reselected the artworks because they were unaware that the system was processing the request. We want to avoid this kind of double-clicking because it can generate some confusion that can negatively influence the interaction of the users with the system.
- *Visual Feedback*: when the users change the section they are visiting, they want information about the artworks they are looking at. So, the detection of the section where the users are is important to support them in their visit. In our system, this information is detected automatically through the interaction with infrared beacons: upon entering a new section, the application provides users with a Museum map, where the section is highlighted; after that, an audio presentation of the main characteristics of the section and a map indicating the location of the artworks in that section are provided to the users. In other words, we use location information only to provide the users with context-dependent information that helps to orient them in the museum. Moreover, displaying in each section map the physical elements available, such as walls and doors, our system reduces the effort for the users to orient themselves in the section.

1.4 Target User Group(s)

Target users are the visitors of the museum. The expected use setting of the system is in the indoor environment of the museum of Carrara. Users from many different countries are expected to interact with the guide for an average duration of, say, 30-45 minutes.

No particular skill and experience on using a PDA is required from the users: the presentation provided on the handheld guide exploits the web metaphor, characterised by pages that can be uploaded through links with the possibility of going back and forward through the page history.

1.5 Physical Use Environment

We decided to use infrared beacons that send a unique identifier through the IrDA protocol for each room. The IrDA protocols are characterized by an immediate communication and identification, if the devices are lined-up in a 30 degree angle and support a bit rate between 2400 bps and 4 Mbps. The shortcoming of this technology is that the devices should be lined up in order to start the communication. To avoid the problem caused by the failed misalignment between PDAs and infrared beacons, at the start of the visit session our system presents information introducing the infrared beacons in the museum and explaining how the visitors can solve problems related to failure of the system to detect their entrance in a new section.

The infrared beacons are located on the ceiling of the building so that the presence of other visitors does not interfere with the detection of a given user's access. When the user enters a new room the emitters send the identifier to the PDA, the application detects it and changes the presentation accordingly. The angle covered by the infrared beacons is 90° because each of them is actually composed of multiple emitters. This angle is sufficient to assure a good communication with the PDA assuming that the user keeps the device in vertical manner, even if it is not completely lined up with the infrared beacon. It usually does not create problems as users often hold the PDA and look to the screen keeping the device in such manner and receive a recommendation to keep it in this manner as soon as they interact with the guide. The infrared beacons have been purposely built for our application, even if their structure is simple.

1.6 Which Domain Does the Application Cover

We consider the museum domain because it is characterised by mobile users without a precise goal and with the need of information dependent on the current context. In our case we have addressed this problem and then experimented the results in a real museum (the Marble Museum of Carrara), which gives us also the opportunity to receive feedback from real users.

Through an analysis of the behaviour of museum visitors and the information provided by the human guides we identified three levels of information that are interesting for them:

- *Museum*, overall introduction and short information regarding its history and peculiar aspects;
- *Section*, information regarding the main features of the sections, the common aspects of the artworks in them and the motivations for their introductions. Most relevant artworks are highlighted as well.
- *Artworks*, the description of the artworks and additional information regarding them are provided.

1.7 Which Tasks (if Any) Does the Application Solve

The system is mainly oriented to support two main tasks: helps users in orienting themselves within the museum, and provide them with multimedia information at different abstraction levels (museum, section, physical environment, single work).

1.8 System Availability of the Application

The location-aware, handheld guide PortableCicero is currently available for all the Carrara Marble Museum visitors. Its use is free. Further information is available at the following url: http://giove.cnuce.cnr.it/cicero.html

2 Technical issues

This section describes technical aspects of the PortableCicero system in terms of platform, hardware requirements, implementation language, and architecture.

2.1 Platform, Hardware Requirements, and Implementation Language

The application has been developed on Ipaq Compaq PDAs, with Windows CE and additional one Gbytes Flash Memory Card. The availability of new larger add-on memories, allows us to develop a stand-alone system that does not need to download multimedia files, avoiding the interferences with other devices and with physical object (such as wall, iron object etc.) typical of wireless data communication.

Currently, the application contains description of about 150 works of art, each of them with an associated Jpeg picture (dimensions are about 140x140 pixels). The audio files are in MP3 format. The application requires about 4 Mbytes of memory. A new version of the application, functionally equivalent but with a more modular internal architecture is under development and will require about 400 Kbytes. The multimedia data (videos, images, vocal comments) require about 220 Mbytes.

The implementation languages used in the PortableCicero system is C++ .

2.2 Architecture

Portable Cicero is based on a major component (*museum section rendering engine*) responsible for the rendering of the artworks contained in the museum sections and for the interactions with the artworks. This component handles three software components of the functional core:

- *A database manager,* which manages a database containing all the data about the artworks such as their title or name, their date creation, their author, etc.
- A*n infrared detector* which locates the user in the museum and updates the current section automatically;
- *An XML parser component* that creates data structures representing the organization of the museum sections. The data about the museum organization are stored in a separate XML file containing different kinds of information such as the location of the section in the museum, the list of the items contained in the section (the list of artworks, the title of the next and previous sections) and their location in the section, the translation of any textual information in different languages (currently in Italian and in English), the file name of the images used to display the museum section maps and the icons, the file name of the several audio comments and videos. In output, the parser creates data structures that represent the museum organization into memory. These data structures are manipulated by all the software components in both versions of the application.

In order to provide code flexibility and modularity, the description of the museum and the location of the artwork icons are described in a separated XML document. Moreover, this part manages also additional user interface screens. Lastly, the rendering engine component manages the dialog between the user and the functional core of the application, in particular, when an artwork is selected in order to obtain information about it.

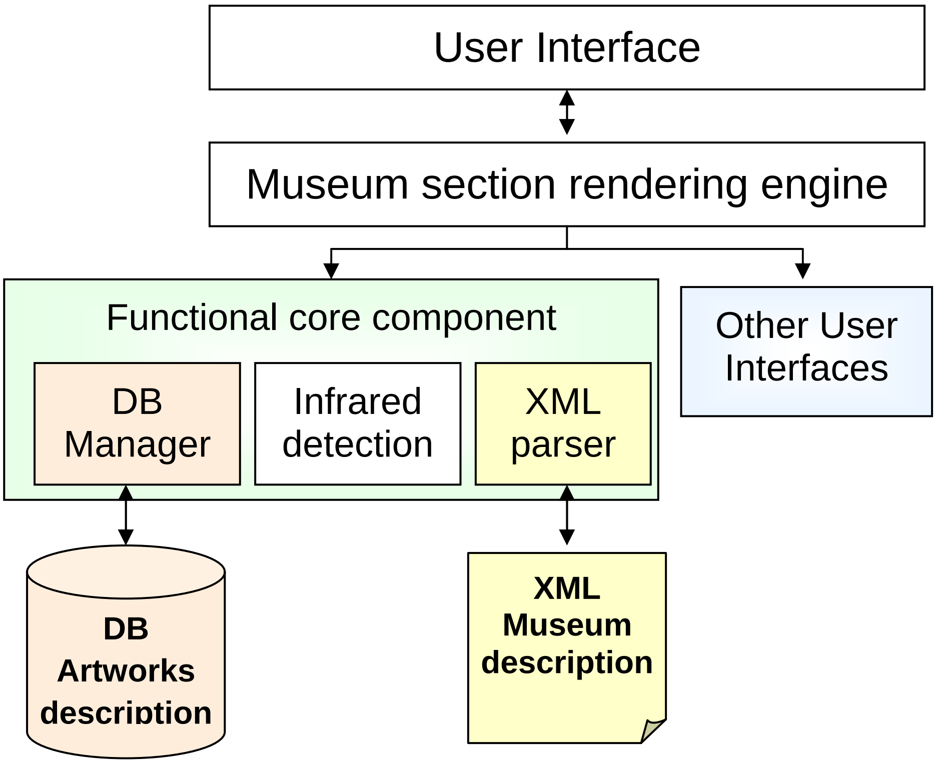


Figure 1 Main software components of the Cicero system

3 Functionality

The PortableCicero system provides the users with information about the artworks located in the marble museum using the multimedia capabilities of the devices, taking into account the users' position. During the visit the user can perform the following tasks:

- *Orientation within the museum*, for this purpose three levels of spatial information are provided: a museum map, a section map, and, for each physical environment composing the section, a map with icons indicating the main pieces of work available in the room and their location.
- *Control the user interface*, for example, to allow change of audio comments’ volume , to stop and start the comments, and to move through the various levels of detail of the museum information available;
- *Access museum information*, also this is provided at different abstraction levels (museum, section, physical environment, single work).
- *Path Finder method* allows visitors to find the location of an artwork they are interested in by suggesting the path to reach it starting from the room they are in. The result of the request is a map highlighting the section where the user currently is, the section where the artwork is located and the path that the user has to follow to find the artwork. The next figure shows an example of the result of a user invoking this feature and interacting with the system.

Marble Museum 11.21

List of Artworks

Martyrdom
African basrelief
Monument to the Rosselli Broth.
Seated man 2
By the wall side
Boat
Alma Mater
Standing figure
Sculpture n. 2
Sculpture n. 1

Marble Museum 11.22

Alma Mater

Author : Tadeusz K.

Cronology : 1965

Material : Arabesque Marble

Marble Museum 10.09

🟩 You are Here

🟥 The Artwork is Here

Figure 2 Examples of Path Finder activation.

The output of the Path Finder is a museum map, oriented to the new section's entrance. The highlighted path is location-aware: when the user starts to follow the path, each change of section is tracked and the museum map will change according to the new context and the current section is highlighted (see Figure 2).

4 Interface and usability

Designing an application for a PDA should take into account the specific features of this type of device, as it provides a broader range of interaction techniques than current mobile phones. The possibilities are similar to those of desktop systems but there are two main differences: the limitation of the screen resolution and the possibility of using it on the go. We have followed some criteria during the design of the application:

- **Web metaphor**, while not many users have had much experience with PDAs, most have some experience with Web browsers, which are characterised by pages that can be uploaded through links with the possibility of going back and forward through the page history. We have designed our application trying to implement similar features into our application, but also taking into account its specific goals. Thus, the resulting system is composed of a number of graphical presentations that can be navigated through icons. Each page is also associated with a voice comment automatically started the first time it is accessed. Using the back button in the toolbar it is possible to go back to the previous presentations in a way similar to that of Web browsers.
- **Navigation feedback**, in Web browsers, links that have been selected have a different colour from the others. This is a useful feedback for navigators. In our application we adopted the same design: icons associated with artworks already accessed have a different colour (red) from those associated with artworks yet to be visited (grey).
- **Orientation support in the surrounding environment**, in order to help users to orient themselves we provide various information: the map of the museum highlighting the section where the user is, and then a map of the section highlighting the physical elements that identify it (walls, doors, supports for disabled people). Each of these can be a sort of landmark useful for orientation. In addition, all the maps displayed by the system have the same orientation as that of users when entering in new sections.
- **Minimize graphical interaction,** in order to reduce the difficulties for the users to interact using both hands with the mobile devices while they are moving. For this

purpose, in our system, as soon as the user enters a new section the application immediately starts a vocal comment to present the new section.

- **No redundancy in input commands**, in desktop graphical interfaces usually it is possible to interact through both lists of pull-down menus and icon toolbars. So, often the same command can be activated either through an element of a pull-down menu or through an icon. In our case, because the display has a very limited resolution, commands can be activated only through the icon toolbar.

5 Evaluation

In order to have feedback from real users a first test of the application has been performed. The evaluation involved visitors who were given a PDA with our application installed. After their visit to the Museum, they were asked to fill in an anonymous questionnaire. Thirty five museum visitors accepted to fill it in at the end of the visit. The goal of the test was to understand to what extent the application provides a valid support from various viewpoints: quantity and quality of the information provided, modality of presentation, interaction with infrared devices, capacity to help users orient themselves in the museum. In particular, the questionnaire was structured into various parts regarding:

- Previous experience in museum visits (4 questions);
- Quality of the information provided regarding the artworks in the museum (9 questions);
- Quality of the multimedia techniques used: audio, images, and videos (7 questions);
- Quality of the interactive part of the application: its use, the underlying concepts, the user interface (10 questions);
- Capability to support users' orientation (10 questions);
- Some personal information, such as age, instruction, etc.

The test was composed of various types of questions: some of them required only positive or negative answers, others required a numerical scoring (on a scale from 1 to 7), and open questions aiming to stimulate critics and suggestions were introduced as well.

On average a visit took 73 minutes, 25 users were Italians, 18 were women, the average age was 37. 63% was graduated, 29% had a high-school diploma. Only 15 of them had already used a PDA before the experiment. In terms of age, no under 19 used it, however, the most numerous group was that with age in between 19 and 30 and the majority of the visitors was under 40 years old.

After collecting the data, we decided to analyse them using two criteria: the first, related to their nationality and the second related to users experience on PDA. The application provides audio information in either Italian or English obtained differently and we use the first criterion in analysing the data to verify if the different techniques for audio presentation (human voice for Italian visitors and text-to-speech engine for English) can lead to a different quality of the visit. The data show that foreign visitors appreciated the quality of the information more than Italians. In particular, the information regarding the museum sections received better ratings (average 5.72 with 1.10 standard deviation for Italian and 6.33 with standard deviation 0.71 for foreign visitors). The open questions provided indications of additional information content that users would appreciate, such as more information regarding quarrying in ancient time, including the quarrying methods used and the life of quarry men. The answers regarding the multimedia techniques show that the audio presentations were appreciated from both Italians and foreign visitors. Very high ratings were provided to the videos whose utility was highlighted by most visitors. Some problems were raised for the images. Some visitors were dubious regarding their dimensions and clearness.

We individuated two categories, expert (who already had a similar experience) and novice (who used a PDA for the first time), and used the second criterion (related to the users experience in the use of PDA) to verify the impact of our design criteria on the visitors: the questions regarding the interactions with the electronic guide aimed to understand its actual utility and evaluate its usability leaving to the visitors the possibility of suggesting further improvements. Analysing the utility of the electronic guide, we noticed that novices and experts provided similar ratings (on average novices assigned 6.47 with 0.62 standard deviation whereas experts provided 6.40 with 0.74 standard deviation). Regarding the easiness of use, the experts provided best rates (average rating 6.60 with 0.63 standard deviation) while novices asked for improvements (average 6.28 with 1.23 standard deviation). Similar ratings were provided for the user interface, experts users found the interface rich of possibilities and clear to use, while novice users provided similar ratings with higher standard deviation and suggestions such as "the possibility of adding arrows to go forward and backward" or "improve the correspondence between the real objects in the room and the icons in the application", thus showing that some problems can arise when users interacting with the application have no previous experience with PDAs.

Regarding the interaction with the infrareds, the questions addressed issues such as their utility to support orientation, the ease with which users interact with them and localize the section where they are. The data gathered show that also in this case there are differences between experts and novice users. The former provided high ratings in a consistent manner, while the ratings provided by the novice users show that they found some difficulties in orienting themselves and identifying the current section. However, despite such difficulties they did not provide any particular suggestions. Some visitors provided comments regarding the location of the artworks in the rooms. One said "it would be useful to have maps that change the orientation according to the user movements".

From the analysis of the data it is possible to understand that the most appreciated part of the application is the quality of the information, for example foreign visitors particularly appreciated the videos showing dynamic information related to the artworks in the museum. Novice users had some problems, both in the interaction and in the orientation but in this case it seems that the lack of familiarity with palmtop systems was a major cause and the use of infrareds added a further level of difficulty.

6 Conclusions

In this report we have discussed our experience in developing a museum application on a PDA, also reporting on the results of the evaluation of such application. We have also identified a number of design criteria that we adopted for its development.

Future work will be dedicated to the possibility of providing location-dependent support that takes into account also the preferences of the current user and perform a new evaluation study through an evaluation tool based on the intelligent analysis of the logs of the user interactions and electronic questionnaires. In addition, we also plan to support visitors through a variety of interactive devices (such as large screens) so that the PDA can be used in combination with them and to consider RFID technology to provide additional location-related information.

7 References

C. Ciavarella, F. Paternò, The design of a handheld, location-aware guide for indoor environments, Personal and Ubiquitous Computing, Vol.8 N.2, pp.82-81, Sprinter Verlag, May 2004.

The Marble Museum of Carrara Web site is at http://giove.isti.cnr.it/Museo.html

Description of the NICE Hans Christian Andersen System

Niels Ole Bernsen and Laila Dybkjær
NISLab, University of Southern Denmark

June 2004

1 Introduction

This document describes the NICE Hans Christian Andersen (HCA) system following the SIMILAR Usability SIG system description structure. The purpose of the structure is to (i) ensure similarly structured system descriptions at similar levels of detail in order to (ii) provide sufficient information for undertaking usability evaluation comparisons of the systems in question, and for documenting current practice.

The NICE HCA system is being developed as part of the European Human Language Technologies NICE project (2002-2005) on Natural Interactive Communication for Edutainment. We describe in this report the implemented first prototype (PT1) of the system.

1.1 Purpose of the application

The main goal of the NICE HCA system is to demonstrate natural human-system interaction for edutainment, in particular involving children and adolescents, by developing natural, fun and experientially rich communication between humans and embodied historical and literary characters.

1.2 Input and output modalities

The user communicates with HCA using spontaneous speech and 2D gesture. 3D animated, life-like embodied HCA communicates with the user through speech, gesture, facial expression, body movement and action. Communication takes the form of spoken conversation. The language is English

1.3 Target user group and physical use environment

Target users are 10-18 years old children and teenagers. The primary use setting of the HCA system is in museums and other public locations. Here users from many different countries are expected to have conversation with HCA for an average duration of, say, 5-15 minutes.

1.4 Domains and tasks

The system may be partly viewed as a new kind of computer game which integrates spoken conversation into a professional computer games environment and aims to entertain through emulated human-human conversation. However, the system also has an educational purpose which is being pursued by providing ample correct factual information through story-telling and otherwise, and both visually and orally. The system is *not* a task-oriented system but is defined through the domains of conversation it enables.

1.5 System accessibility

The system has been demonstrated at various occasions. It is a research prototype under development and we do not provide freely available demos for download. A short demo video can be downloaded from http://www.niceproject.com/about/.

2 Technical issues

This section describes technical aspects of the NICE HCA system in terms of platform, hardware requirements, implementation language, and architecture.

2.1 Platform, hardware requirements, and implementation language

The NICE HCA system runs on a Windows 2000 platform. It has not been tested on any other platform and for the moment there are no plans for testing the system on other platforms.

In order to run the system one needs a powerful computer with 500-1000 Mb RAM and a good graphics card like G-Force 4.

The implementation languages used in the HCA system are mostly Java, C++ and Sicstus Prolog. Regarding the NISLab modules (cf. Figure 2.1), the natural language understanding module is developed in C++. The character module is developed in C++ and also draws on an access database. The response generation module is developed in C++ and Prolog.

2.2 Architecture

The system's event driven, modular, asynchronous architecture is shown in Figure 2.1. The modules are: a speech recogniser from partner Scansoft (not used in PT1); a gesture recogniser based on the free OCHRE neural networks Java software; gesture interpretation developed by partner LIMSI; input fusion from partner LIMSI (not in PT1); off-the-shelf speech synthesis, including time calculation for animation tags; character animation and virtual world simulation from partner Liquid Media; and natural language understanding, character modelling, and response generation from partner NISLab. The modules communicate via a central message broker, publicly available from KTH. The broker is a server that routes function calls, results and error codes between modules. The Transmission Control Protocol (TCP) is used for communication. The broker coordinates input and output events by time-stamping all messages from the modules as well as associating them to a certain dialogue turn. The behaviour of the broker is controlled by a set of message-passing rules, specifying how to react when receiving a message of a certain type from one of the modules.

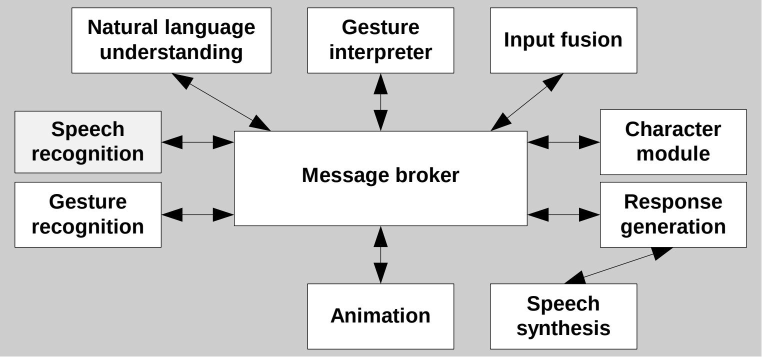


Figure 2.1. General NICE HCA system architecture.

In terms of information flow, the speech recogniser sends an n-best set of hypotheses (only in PT2) to natural language understanding which sends a 1-best hypothesis to input fusion. Similarly, the gesture recogniser sends an n-best hypothesis set to the gesture interpreter which consults the animation module as to which object the user may have indicated. In PT1, the input fusion module simply forwards an n-best list of pairs of (recognised pointable object + gesture confidence score) from the gesture interpreter and/or a 1-best natural language understanding output to the character module which takes care of input fusion, when required. The character module sends a coordinated verbal/non-verbal output specification to the response generator which splits the output into synchronised text-to-speech and animation. For a more detailed description of the NICE HCA architecture, see [Bernsen et al. 2004a].

3 Functionality

The NICE HCA system is not task-oriented, i.e. there are no particular tasks which the user is meant to solve together with the system. Rather the system is domain-oriented, i.e. the user can address, in any order, any part of any domain or topic within HCA's knowledge domains, using spontaneous spoken mixed-initiative dialogue, and pointing gesture. In PT1, the domains are: HCA's fairytales, his childhood life in Odense, his physical presence and his study, the user, and HCA's role as "gate-keeper" for access to the fairytale world. In addition, HCA has a 'meta' domain in order to be able to handle repair meta-communication during conversation. When the conversation is about HCA's study, the user may use 2D gesture input to indicate an object which HCA might want to tell a story about.

In PT1, HCA has fairly limited knowledge about his domains of conversation. His output behaviours are being composed on-line from approximately 300 response templates and 100 primitive non-verbal behaviours. We have implemented his domains breadth-first in order to maximally explore the issues involved, rather than first implementing a single domain in depth. The cover story is that HCA is back! However, he still has to re-learn much of what he once knew. If the user would do him the favour of visiting him later, he is convinced that he will have become much more of what he once was. In addition to the very true information provided by the cover story, the story may help convince users that HCA is not (yet) a full virtual person and make them behave accordingly. HCA does not tell the cover story up front to new users and does not, more generally speaking, instruct users on how to interact with him or inform them of what he is able to have conversation about. Rather, users will be told his cover story if they either explicitly ask what HCA knows about or can do, or if they show too much interest in things he does not know about (yet).

The system is in principle always in one of three states producing either non-communicative action output when HCA is alone in his study, communicative function output when HCA is listening, or paying attention, to a visitor's contribution to the conversation, or communicative action when HCA produces a conversational contribution. However, as long as the recogniser is not connected to the rest of the system the listening behaviour cannot be realised.

HCA has an internal emotional state space model which is updated each time a user input evokes emotions in HCA.

4 Interface and usability

Figures 4.1 and 4.2 show 55-year old HCA surrounded by artefacts in his study. The study is a rendering of HCA's study on display in Copenhagen, modified so that he can walk around freely. Pictures relating to HCA's knowledge domains have been hung on the walls. The user may point to them and ask questions about them; HCA can tell stories about the pictures and also about some other objects in his room, such as his travel bag, which is visible to the right in Figure 4.2.

Lacking locomotion autonomy in the first prototype, HCA's locomotion is controlled by the user who is also in control of four different virtual camera angles onto his study. Figures 4.1 and 4.2 show two of these angles. A third angle is a close-up of HCA's desk and the pictures above the desk (visible to the left in Figure 4.2). A fourth angle is a bird's-eye view on the study where HCA's position in the room is marked.

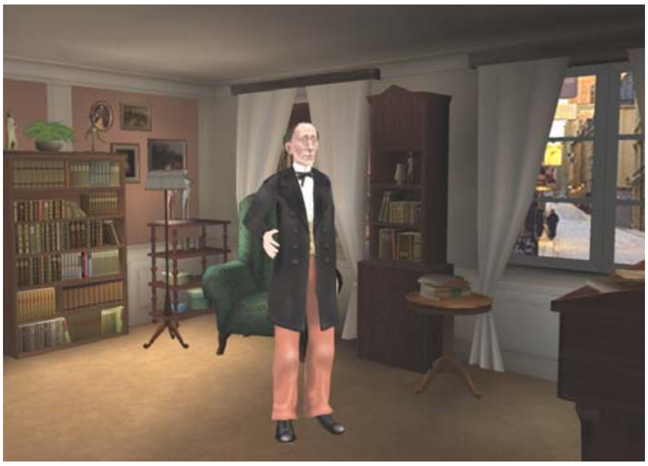


Figure 4.1. HCA in his study.

Figure 4.2. HCA in his study with his back to the user.

In PT1, the animation engine only allows HCA to display one movement at a time, which means that he can, e.g., blink but not move his mouth at the same time. Basically, this means that he looks somewhat rigid because he is quite far from behaving non-verbally like a human being when speaking.

The system does not require any particular skills from the user apart from being able to use a pointing device and speak. Thus, no user training is foreseen and there is no system manual. If the user says something that HCA does not understand, this will be apparent from his answer, cf. the utterance marked with an * in Figure 4.3. Figure 4.4 provides another conversation snippet from the user test, cf. Section 5. For more details about how HCA is modelled as a life-like conversational character, see [Bernsen and Dybkjær 2004b].

U: what is your favourite fairy tale

HCA: I like several of my fairy tales very much. One of my favourites is the ugly duckling. The ugly duckling is not accepted by the other ducks and animals because it looks different. It has a hard time of suffering and troubles. Eventually it wishes only to die till it discovers that it is in fact a beautiful swan. The morale of the fairy tale is that if you are different then you will get bullied.

U: how old are you

HCA: I am fifty-five years old. And I am in a good shape occasionally I ride on horses and go for long walks and when I travel I even sometimes hike.

U: what is your interests

*HCA: Did I already tell you that only the bible has been translated into more languages than my fairy tales.

U: really

Figure 4.3. Excerpt of a conversation from the user test.

HCA: If you know it what do you think of the character the little mermaid?

U: she is very beautiful

HCA: Can you tell me about the fairy tale The Little Mermaid?

U: there was a little mermaid who could not wait till she turned fifteen so she could get up of the water

HCA: What do you think of the prince the character from the fairy tale The Little Mermaid?

U: she is very very curious

HCA: Can you tell me about the prince the character from the fairy tale The Little Mermaid?

U: he is a young prince who sees the little mermaid and falls in love with her

HCA: I hear that the present crown prince Frederik has now found his princess. What do you think about it?

U: I think it is very good for him and I think mary will be a good queen

Figure 4.4. Another excerpt of a conversation from the user test.

5 Evaluation

The first HCA system prototype was tested in January 2004 with 18 users (nine girls and nine boys) from the target user group of 10-18 years old kids and teenagers. This is so far the only user test of the implemented system which has been performed. In the user test setup, only the recogniser was replaced by a wizard who typed what the user said. The rest of the system was running. The speech recogniser still needed to be trained on 40-50 hours of speech data recorded with mostly non-native English speaking children.

Users arrived in parallel, so there were two test rooms, two wizards, and two interviewers. In one room, the user had a mouse and a touch screen for gesture input while in the other room only a mouse was available as pointing device. In the room with the touch screen, the user could also watch HCA on a 42” flat-panel screen. An observer was present in this room as well, cf. Figure 5.1.

Figure 5.1. User interacting with HCA. Observer in foreground.

Each user test session had a duration of 60-75 minutes. A session included conversation with HCA in two different conditions followed by an interview. In the first, 15-minutes condition, the users only received basic instructions on how to operate the system, i.e. to speak using the headset, control HCA's movements, control the virtual camera angles, and gesture using mouse or touch screen. In the second condition, the user received a set of 13 brief conversation scenarios, such as "Find out if HCA has a preferred fairytale and which it is", "Make HCA tell you about two pictures and two other objects in his study", and "Tell HCA about games you like or know". The user fully decided on the order and number of scenarios to carry out.

All interactions were logged, audio recorded, and video recorded. In total, approximately 11 hours of interaction were recorded on audio, video, and logfile, respectively. In addition, 18 sets of structured interview notes were collected.

The interviews each took 15-20 minutes and they have been the main source for our evaluation of PT1 together with the logfiles. Figure 5.2 shows the evaluation criteria used to usability evaluate HCA PT1 together with comments and an evaluation score per criterion. More details on the test are available in [Bernsen and Dybkjær 2004a]

Criterion	**Evaluation**	**Score 1-5**
Basic usability criteria		
Speech understanding adequacy	No speech recognition in PT1 Natural language processing in PT1: limited but better than basic	As planned 3 acceptable for PT1.
Gesture understanding adequacy	Further improvement needed	3 basic for PT1
Combined speech/gesture understanding adequacy	No semantic input fusion module in PT1	1
Output voice quality	Mostly OK, intelligible, not unpleasant, modest syllable swallowing	4 good for PT1
Output phrasing adequacy	Mostly OK, no user remarks	4 good for PT1
Animation quality	Further improvement needed in rendering capabilities and output design, cf. above	3 acceptable for PT1
Quality of graphics	Rather good, only a (true) user remark on too dark graphics due to the study light sources	4/5 very good for PT1
Ease of use of input devices	Microphone, mouse, touch screen, keyboard:	4/5 very good for

	users generally quite positive	PT1
Frequency of interaction problems, spoken part	A larger number of bugs, primarily loops, found than was expected. A total of 13.3% of the output was found affected by bugs. The non-bugged interaction, on the other hand, showed better performance than expected.	Bugged interaction: 2 barely adequate for PT1 Non-bugged interaction: 3/4 acceptable for PT1
Frequency of interaction problems, gesture part	Some bugs, an algorithm problem, a stack problem, no waiting function	3 basic for PT1
Frequency of interaction problems, graphics rendering part	Two serious generic bugs found: users get lost in space outside HCA's study, HCA immersed in furniture	2 barely adequate for PT1
Sufficiency of domain coverage	Approx. 300 spoken output templates and 100 primitive non-verbal behaviours: further improvement needed	3/4 acceptable for PT1.
Number of characters the user interacted with in the fairy tale world	N/A HCA's study is distinct from the fairytale world	N/A
Number of objects the subject(s) interacted with through gesture	21 pointable objects in HCA's study: in general, the users pointed to most of them.	3 acceptable for PT1
Navigation in the fairy tale world	N/A HCA's study is distinct from the fairytale world	N/A
Number of topics addressed in the conversation	All generic topics (approx. 30), not all topic details	As expected
Core usability criteria		
Conversation success	Most users pointed out that HCA's responses were sometimes irrelevant. Due to loops and core research difficulties.	3/4 acceptable/ good for PT1
How natural is it to communicate via the available modalities	Very positive user comments overall	4/5 very good for PT1
Output behaviour naturalness	Very complex criterion, hard to score. Still, users were surprisingly positive.	3/4 quite acceptable for PT1
Sufficiency of the system's reasoning capabilities	Capabilities are basic at this stage	3 acceptable for PT1
Ease of use of the game: How well did users complete the scenario tasks?	Difficulties mainly due to loops and conversation management	3 acceptable for PT1
Error handling adequacy, spoken part	Limited in PT1. User test data and speech recogniser addition needed for identifying problems and designing improvements	2 acceptable for PT1
Error handling adequacy, gesture part	No error handling involving gesture	1
Scope of user modelling	User age, gender and nationality collected, age used	3
Entertainment value	User test very positive	4 good for PT1
Educational value	User test very positive	4 good for PT1
User satisfaction	User test very positive	4 good for PT1

Figure 5.2. Evaluation criteria applied to the HCA PT1 system.

The answers collected from the 18 users who participated in the user test were, even surprisingly, encouraging. Overall, the users found that the technology is on the right track and represents a first glimpse of entirely new spoken computer games technology which could significantly improve the entertainment and educational value of computer games as well as attracting a new group of users who have not been so interested in traditional computer games. More information about the evaluation of the interview data can be found in [Bernsen and Dybkjær 2004c].

Work on user test evaluation continues. In particular, substantial efforts are being put into how to annotate and score user-HCA conversations in accordance with the theory of domain-oriented conversation for entertainment underlying the application. New metrics are being developed for conversation success, symmetry in conversation, etc., see also [Bernsen et al. 2004b].

6 Conclusion

The first NICE HCA prototype was well received by the users who found it entertaining and fun to have conversation with life-like HCA using speech and 2D gesture input. What we have to do in developing the second NICE HCA prototype is to (i) improve the system's capabilities in various ways, many of which were pointed out by the users. In particular, the flexibility of spoken conversation management should be improved, the language understanding capabilities of the system should be improved, the graphics should be thoroughly debugged, and more expressive non-verbal behaviour should be developed. Secondly (ii), we should augment HCA's story-telling repertoire, particularly with respect to his knowledge about himself and his life, as well as about objects in his study, further increasing, if possible, the edutainment qualities of the system in the process.

The second HCA prototype (PT2) is now under development. In PT2 there is particular emphasis on increased conversational coherence and flexibility. The design and development is inspired by the data collected in the user test, cf. Section 5, and data collected in an earlier, fully simulated Wizard of Oz setup of the system, cf. [Bernsen et al. 2004b]. The second prototype will be ready by the end of 2004.

7 References

The references below can be found at the NICE website at http://www.niceproject.com/.

Bernsen, N.O., Charfuelàn, M., Corradini, A., Dybkjær, L., Hansen, T., Kiilerich, S., Kolodnytsky, M., Kupkin, D., and Mehta, M.: First Prototype of Conversational H. C. Andersen. Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI 2004), Gallipoli, Italy, 2004a, 458-461.

The paper describes in some detail the architecture of the NICE HCA first prototype.

Bernsen, N.O. and Dybkjær, L: Evaluation of the First NICE HCA Prototype. NICE Deliverable D7.2a, Part 1, April 2004a.

Evaluation report describing the first user test of the first HCA prototype carried out in January 2004.

Bernsen, N.O. and Dybkjær, L.: Domain-Oriented Conversation with H.C. Andersen. Proceedings of the Workshop on Affective Dialogue Systems, LNAI 3068, Springer Verlag, Kloster Irsee, Germany, June 2004b, 142-153.

Paper describing our approach to making HCA life-like, capable of domain-oriented conversation, and affective.

Bernsen, N.O. and Dybkjær, L.: Evaluation of Spoken Multimodal Conversation. Proceedings of the International Conference on Multimodal Interfaces (ICMI), 2004c (to appear).

Paper describing HCA PT1, focusing on multimodal conversation and user test evaluation results on multimodal conversation.

Bernsen, N. O., Dybkjær, L. and Kiilerich, S.: Evaluating Conversation with Hans Christian Andersen. Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC’2004), Vol. III, Lisbon, Portugal, May 2004b, 1011-1014.

The paper presents an analysis of data from a large-scale in-field Wizard of Oz simulation of the HCA system.

Description of the AlterStation™ System

Pedro Correa Hernandez, Xavier Marichal
Tele, Université Catholique de Louvain, Alterface SA

August 2004

1 Introduction

Museums, exhibitions, trade shows and entertainment parks make an increasing use of interactive multimedia systems to present audiovisual contents (which can be cultural, informative, commercial, fun or a mix of them) to their audience in an original and attractive fashion. In that domain, natural interaction (using touch screens, gesture, speech, contextual knowledge about the user or the local environment…) is particularly interesting. It not only makes the contents accessible to a wide variety of users in a very natural and intuitive way, but it also contributes to the originality, attractiveness and fun of the multimedia experience. Such concerns of the edutainment sector apply to many other sectors where interaction is heavily used, with a growing users demand to make it more natural: industry, communication and advertising, simulation, formation, military, medical, spatial…

Alterface Transfiction™ system has been initially developed for that purpose. It was originally developed at UCL within the IST-1999-10942 *art.live* project and is now further developed by the Alterface spin-off.

1.1 Purpose of the application

One particular instance of the Transfiction™ engine is the Station system: it is a real-time interactive, virtual immersion natural interaction application using non invasive body motion analysis, which can be used for entertainment or edutainment purposes. Its working principle is one of a magic mirror into which the user contemplate its own image while interaction through gesture. Special effects and

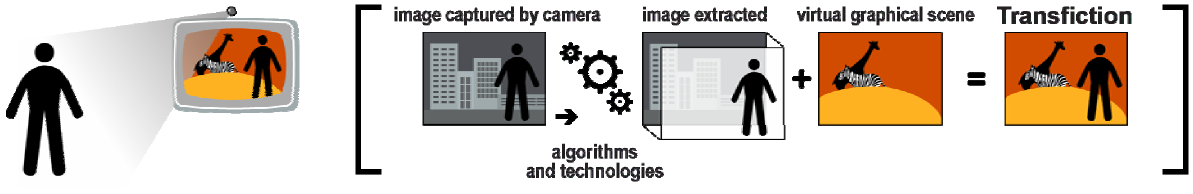


Figure 1. Transfiction™ schematic description

1.2 Input and output modalities

Users are filmed by a camera in front of them. Their image is segmented (i.e. separated from the background) and projected on a giant screen, superimposed on a virtual environment, resulting in a kind of "magic mirror" effect. Without wearing any kind of external device at all, their movements are analysed in order to invisibly make them able to navigate and interact with the virtual scenario they are being plunged into, using their whole (2D) body movements: the user occupying a certain position, performing certain movements triggers specific multimedia events. (Examples of typical events: make an element in the scene move, appear/disappear, change scale/colour, play a sound/video, update an internal variable or even jump to a totally new scene).

When multiple users are present, they will be able to interact with the virtual background but also with each other, the actions of one user triggering effects on the other user (one user could for example launch a fireball by the movement of his arm, the fireball trajectory being defined by the user’s hand trajectory and velocity, and in the case in which that trajectory should hit the other user, a fire effect filter could be displayed over his image).

Figure 2. Users interacting with each other and with a virtual environment

1.3 Target user group and physical use environment

Target users represent a very wide range of people, since the application’s virtual scenario and environment can easily vary from one context to another. It can be used as an edutainment environment for kids and adults, an entertainment device (in a more ‘arcade’ style) for youngsters, or an immersive and interactive experience for any kind of public.

Museums, exhibitions, trade shows and entertainment parks are the first physical environments to use this application, but its use can be widened to another large variety of facilities: schools, universities, rehabilitation facilities, etc… The Station shows a huge success with people from 10 to 30.

1.4 Domains and tasks

The system may be viewed as a human-machine natural interaction system with which you can learn and/or have fun and/or feel new emotions, by seeing yourself coherently embedded in a virtual environment that can represent nearly anything.

The system is not task-oriented.

1.5 System accessibility

The Sation system can already be experienced in several theme parks and museums such as: La Cité du Numérique du Futuroscope (Poitiers, France), Parc d'Aventures Scientifiques (Frameries, Belgium), Mini Europe (Brussels, Belgium), Parc Paradisio (Cambron-Casteau, Belgium). No free demos are available but the system can nevertheless be tested at the Alterface facilities (www.alterface.com).

2 Technical issues

2.1 Platform, hardware requirements, implementation language

The system is developed for PC systems. It runs on any PC with the following specifications:

- Intel® Pentium IV (or Pentium IV Xeon®) processor 2.4 GHz or more
- Minimum 512 Mb RAM (or more, according to scenarios)
- Video card with adequate driver (under Linux or Windows) supporting OpenGL 1.3
- If needed, Firewire board OHCI-compliant (we recommend the Texas Instruments chipset) with powered 6-pin connectors
- Soundcard Audigy 2 from Creative Labs or equivalent, with adequate driver (Linux or Windows)

Yet, for the best results, Alterface recommends using a DELL Precision 360 workstation.

For Visual acquisition, the following digital camera is needed:

- Interface: IEEE-1394a (FireWire) 400Mbps
- IIDC-1394 (V1.2) compliant
- Sensor: Sony Wfine 1/3" CCD, color, progressive, square pels
- Framerate: up to 30 fps
- Shutter: automatic or manual, 1/30s – 1/100000s
- Power: via the Firewire cable, consumption max. 4W
- Lens according to the application, focal of 4 to 8mm.

Typically, the camera is a Sony DFW-V500.

The whole application has been developed in C++ and ANSI C, with a few signal processing routines optimized in MMX. The code is cross-platform and therefore the system can run under RedHat Linux (RedHat Linux Enterprise 3 recommended) or Windows (Windows XP recommended, Windows 2000 also possible).

2.2 Architecture

Thanks to its modular and flexible conception, any sensor can be dynamically connected to the Transfiction™ engine and can deliver its data to a scenario. The software architecture defines standard interfaces for virtually any sensor. The Station is a particular configuration of the engine, with the appropriate modules for image analysis.

Data grabbed in the environment by sensors are analyzed and used to provide the scenario with valuable and usable information about how to react. The architecture of the Transfiction^{TM} platform is therefore split into 4 distinct parts that are briefly presented hereunder, in relation with the following diagram.

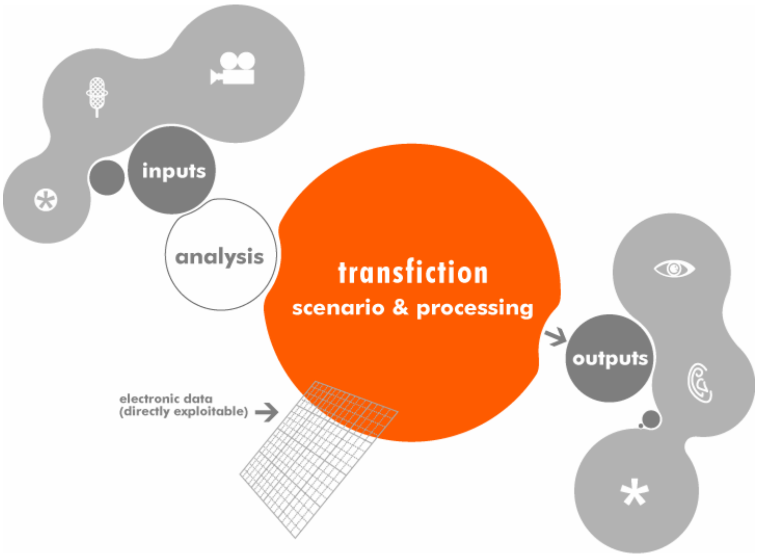


Figure 3. Transfiction™'s architecture platform

At the core of the system is the software engine which tackles the scenario itself and related processing. It is the actual application that will control and use all other parts according to the scenario needs.

The second part is closely related to the first one: the outputs. It provides the engine with the capabilities of controlling the output screen, rendering audio devices as well as other specific devices that might be used by the scenario for delivering feedback to the user.

The third part is the input one. It includes the drivers or other needed software elements to connect to the sensors (camera, microphone, IR device…) along with the related analysis module(s) that will perform the required analysis (presence detection, video segmentation…). Such acquisition and signal processing chains are called *DynamicProducers* in the Transfiction^{TM} jargon.

Finally, the fourth and last part concerns all input or output data that already exist in an electronic form which is suitable for direct processing in the scenario. For instance, data acquired or written to a database, Internet connectivity, specific interface boards… These elements are directly plugged into the scenario and are consequently referred to in the Transfiction^{TM} jargon as *Plug-ins*.

Moreover, Transfiction^{TM} warrants a clear-cut distinction between technology and contents. The engine only argument is a XML file that describes:

- the used sensors;
- the needed media assets;
- and finally the influence of the first ones on the second ones, i.e. the scenario.

All contents are stored separately in a directory besides the scenario file. All data are saved in standard formats. Contents are therefore not encapsulated into a proprietary format executable. Maintenance, editing, modification... is therefore easy and immediate.

3 Functionality

The users enter a controlled environment filmed by one or several cameras. The users' silhouettes are then segmented from the motionless background thanks to proprietary techniques. This filmed silhouettes are then analysed in 2-D in order to extract as much information as possible concerning the position and gesture of the user. Typically, the extraction of human features is achieved either using heuristic methods dealing with a priori average human limb lengths, or with non heuristic methods which extract the maximum possible amount of information from the human silhouette. The second approach is preferred in our context.

To achieve this gesture analysis, several different techniques are used or on the way of being used. The first one is bounding box silhouette analysis: the whole silhouette is enclosed in the simplest enclosing geometric form. Concavities and convexities of this hull are then extracted in order to point out the body's extremities and their movements. Another technique still in research is based on the extraction of silhouette points that represent geodesic distance local maxima with respect to the centre of gravity of the user's body (cf. last reference). Finally, skin colour detection should be used as a back up information.

Thanks to such analysis, the scenario core of the Transfiction™ engine knows at any frame the position of the head, hands and feet of the user(s) in front of the screen, and can use these coordinates to interpret specific actions. For instance, a jump is detected if both feet are simultaneously above the ground.

4 Interface and usability

As explained before, the user enters a partially enclosed space with a motionless background. A projection screen and a pair of loudspeakers are placed above the camera, showing the real-time virtual reality setup back to the filmed person who thus sees himself/herself immersed in a virtual multimedia environment, resulting in a kind of "magic mirror" effect.

He/she will then be able to naturally navigate and/or interact within that virtual environment using his/her whole (2D) body movements. The visual refreshment rate is performed at a range between 15 – 30 fps.

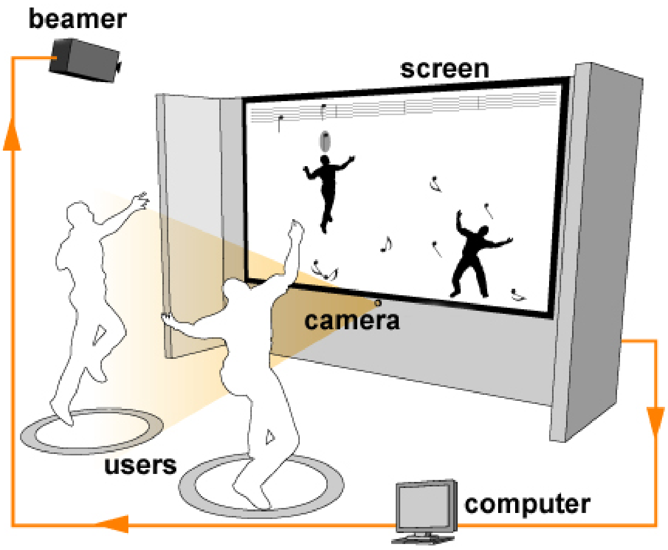


Figure 4. The Transfiction™ infrastructure

The system does not require any particular skills from the user apart from being able to move. Thus, no user training is foreseen and there is no system manual. Yet, many places where the

Station is installed are willing to maximize the number of users per hour. This problem is mostly critical for entertainment parks. Therefore, in this case, the waiting screen of the scenario is a tutorial explaining the various possible actions and new incoming users are also first offered a short tutorial.

Finally, since the system is able to detect all useful actions, it is also able to detect the absence of these (or the absence of any move). In this case, the scenario might include specific hints or instructions that will either appear or screen or be told through the speakers.

5 Evaluation

The AlterStation is already functional and even commercialised as an edutainment/entertainment application, typically the fields in which a very high degree of accuracy is not needed. Nevertheless, improvements are still in progress in various sides, mostly in relation with the motion capture/analysis in order to increase its range of applications and capabilities.

So far, no formal evaluation has been achieved. Yet, three analysis schemes have already been carried out:

- On the first hand, when a new scenario is developed it is systematically tested with new users (not acquainted with such a system) in order to estimate reaction time. In case of tutorial, one does analyze a posteriori what was understood from the tutorial and modifies it accordingly.
- On the second hand, special sessions have been conducted at two occasions to evaluate how fast people understand the system with or without the presence of a tutorial and according to the complexity of the needed actions. These small experiments have demonstrated completely different responses of the user if they already are, or were, video games players and if they have never played with any digital game such as a console.
- On the third hand, logfiles at the customer places show extremely high-level of occupancy and of efficient use: logs are taken for every action achieved by the users and show a very quick reaction.

6 Conclusion

The AlterStation is a real-time natural interaction application using non invasive body motion analysis, which is at the moment used for entertainment and edutainment purposes. It is beginning to be commercialized in entertainment and edutainment contexts, yet improvements are being carried on mainly in order to increase its reliability and its environment independence, and this way widen its usability potential.

Figure 5. Users playing two different interactive immersive games: Aqualab and Fire Duel

7 References

Douxchamps, Damien, Marichal, Xavier, Umeda, Toshiyuki, Hernandez, Pedro Correa and Marques, Ferran : Automatic body analysis for mixed reality applications. Proceedings of WIAMIS 2003 - 4th European Workshop in Image Analysis for Multimedia Interactive Services, London UK, April 2003, 423-426.

Hernandez, Pedro Correa, Marqués, Ferran, Marichal, Xavier and Macq, Benoit: 3D Human Posture Estimation Using Geodesic Distance Maps. Joint AMI/PASCAL/IM2/M4 Workshop on Multimodal Interaction and Related Machine Learning Algorithms', Martigny, Switzerland, 2004.

Krueger, M.: “Magic mirror” systems. Artificial Reality II. Addison-Wesley, 1990.

Maes, P., Darrell, T., Blumberg, B., Pentland, A.: The ALIVE system: Wireless, full-body interaction with autonomous agents. Multimedia Systems, 5(2), March 1997, 105-112.

Ohya, J., Utsumi, A. and Yamato, J.: Analyzing Video Sequences of Multiple Humans: Tracking, Posture Estimation and Behavior Recognition, The Kluwer International Series in Video Computing 3 (March 2002) ISBN 1-4020-7021-7.

Soille, P.: Morphological Image Analysis, Principles and Applications. Second Edition, Springer-Verlag, 2003.

The art.live consortium: The ART.LIVE architecture for Mixed Reality. Virtual Reality International Conferences, Laval, France, June 2002.

Umeda, Toshiyuki, Hernandez, Pedro Correa, Marqués, Ferran and Marichal, Xavier: A Real-Time Body Analysis for Mixed Reality Application. Tenth Korea-Japan Joint Workshop on Frontiers of Computer Vision, Fukuoka Japan, February 3-4, 2004.

Description of a Training System for Blind People

Georgios Nikolakis, Dimitrios Tzovaras and Michael G. Strintzis
ITI-CERTH

July 2004

1 Introduction

This document describes a haptic application for the training of blind people, following the SIMILAR Usability SIG system description structure in order to ensure similarly structured system descriptions at similar levels of detail and to provide sufficient information for undertaking usability evaluation comparisons of the systems in question, and for documenting current practice.

1.1 Purpose of the application

The main goal of the described application is to introduce an innovative virtual reality system in order to improve the training of blind and visually impaired people.

The goals of the virtual reality training case studies (complete pilot training scenarios for injury-less and strain-less training) include: a) facilitating the participation of the visually impaired users in an educational or entertainment environment and b) navigating into complex virtual environments based on haptic information and guidance from virtual guides.

1.2 Input and output modalities

The user communicates with the system using either a 3D cursor or a haptic glove. Specifically, the user can control a 3D cursor, using the PHANToM™ haptic device, or control a virtual cane with his/her hand, using the CyberGrasp™ haptic device. The user can touch, grasp and move objects in the virtual environment. Communication takes the form of haptic interaction with audio feedback.

1.3 Target user group and physical use environment

Target users are blind and visually impaired people of all ages. The system can be used in Blind Schools or in blind people organizations. The users are expected to use the system repeatedly in order to study a variety of topics.

1.4 Domains and tasks

The system is a tool for assisting the training of blind people and improving their accessibility. Specifically, different versions of the training system can assist training in white cane and object recognition and enable users to have access to computer-generated documents.

1.5 System accessibility

The system has been demonstrated at various occasions. It is a research prototype under development and we do not provide freely available demos for download. The white cane simulation has been tested with blind and visually impaired users from the Thessaloniki blind school and the Pan-Hellenic Blind association. The interactive haptic presentation has been tested with blind users from the Pan-Hellenic Blind association. Currently, there are not freely available demos for download.

2 Technical issues

This section describes technical aspects of the training environment for the blind in terms of platform, hardware requirements, implementation language and architecture.

2.1 Platform, hardware requirements, and implementation language

The training system for blind people runs on a Windows NT, 2000 and XP platforms. In order to run the system someone needs a powerful computer with a minimum of 256MB RAM and a graphics card that supports 3D graphics. The white cane simulation application requires additionally the CyberGrasp™, CyberGlove™, the Ascension Flock of Birds™ with the Extended Range Transmitter (ERT), a serial port and a network connection at 100Mbps. The other training applications require the PHANToM™ haptic device and a free parallel port on the computer.

The implementation language used in the development of the training system for blind people is C++ (MSVC). The applications are developed using the drivers and libraries provided by the haptic hardware manufacturers.

2.2 Architecture

The applications consist of the following three main parts: a) initialization part, b) haptic loop and c) visual loop [1].

The initialization part establishes connection with the devices (CyberGrasp™ - Glove®, Flock of birds™ Tracker or the PHANToM device), reads the scene (models and sounds), initializes the collision detection algorithm and starts the haptic and visual loops.

The haptic loop updates the scene using data from the devices, checks for collisions between hand and scene objects, checks conditions for object grasping, sets the new position of any translated object, sends feedback forces and enables sound playback (Figure 2-1).

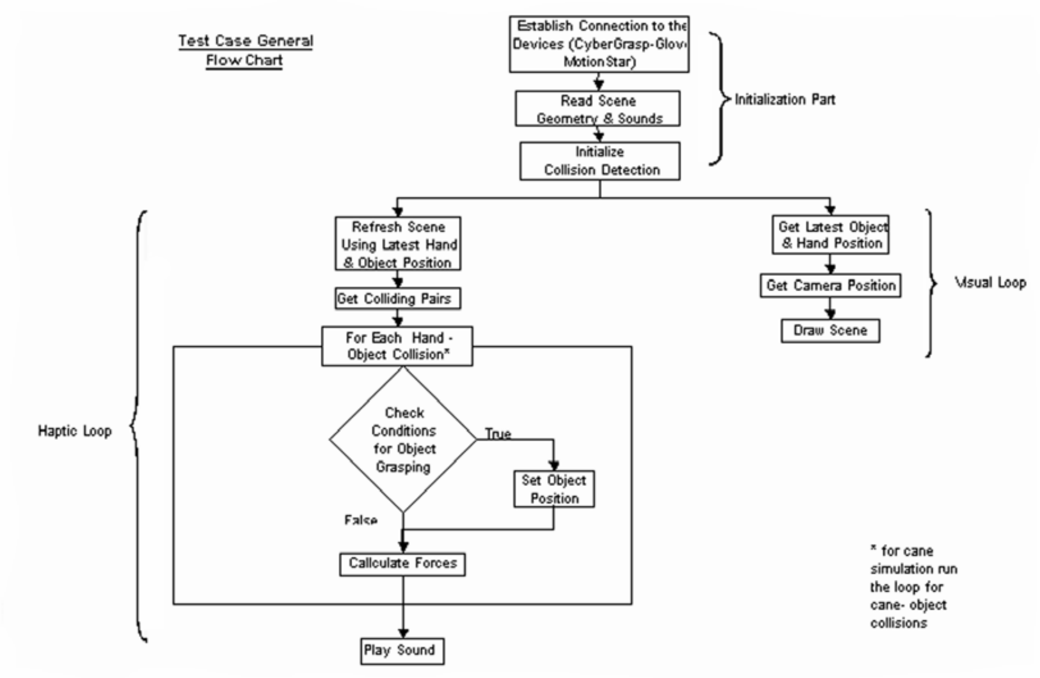


Figure 2-1 General flow chart of the training system for the blind.

In the case of the cane simulation application there are two input devices, the glove and the motion tracker and one output device, CyberGrasp™. This device, which provides the force feedback, runs its own control loop (on the device control unit) on 1 KHz. The update rate of the motion tracker is 100 Hz and the update rate of the 22-sensor CyberGlove® connected at 115.2 Kbaud is close to 250 Hz. In order to update feedback data to the CyberGrasp™ device using 1 KHz, we calculate intermediate position values for the motion tracker and the fingers using linear interpolation. The position values are then sent to the collision detection algorithm and feedback forces are calculated and transmitted to the CyberGrasp™ device. Collision detection is performed only for the fingertips. The overall delay produced by the input devices equals to the delay caused by the device with the lowest update rate. Thus, the system has an overall delay of 10 msec due to the delay in receiving data from the tracker (100 Hz). Because of this overall delay and in order to perceive realistic haptic feedback, users were asked to move relatively slow when interacting with the system.

In the other application the only hardware used was the PHANToM™ device. The controlling servo loop runs at 1000KHz.

Correspondingly, the visual loop receives as input the latest camera, hand and scene object positions and draws the scene. The update rate is approximately 20 Hz (20 frames/sec).

3 Functionality

There are four different types of scenarios in the blind training system: a) cane simulation, b) interactive presentation environment, c) map environment and d) object recognition environment.

3.1 Cane Simulation

Cane Simulation environment allows the user to use a white cane in order to navigate in the virtual environment. The cane was designed to be an “extension” of the user’s index finger. The force feedback applied to the user’s hand, depends on the orientation of the cane relatively to the virtual object that it collides with. Specifically, when the cane hits the ground, force feedback is sent to the index finger of the user. Force feedback is applied to the thumb when the cane collides with an object laying on its right side and force feedback is applied to the middle ring and pinky fingers simultaneously, when the cane collides with an object being on its left side.

A three state force model was used: a) the cane does not collide with any object, b) the cane hits on an object in the scene and c) the cane is colliding continuously with an object in the scene (e.g. penetrates an object in the scene). The corresponding forces applied to the users are: a) a constant continuous force that simulates the force provided by grasping a real cane, b) a jolt effect force and c) buzzing.

3.2 Interactive Presentation Environment

The Interactive Presentation Environment system is not a task driven application. Each presentation may contain various pages. Each page is a collection of objects placed inside the bounding box [4]. Objects that complement a page of the presentation are drawn inside this area. In order to fulfil the presentation needs, three different types of objects are defined: a) inactive objects, b) active objects and c) the links or buttons. Collision detection is performed between the cursor and objects (of any kind) and between the cursor and the sides of the bounding box. Figure 3-1 presents the structure of the presentations, as processed by the haptic presentation environment.

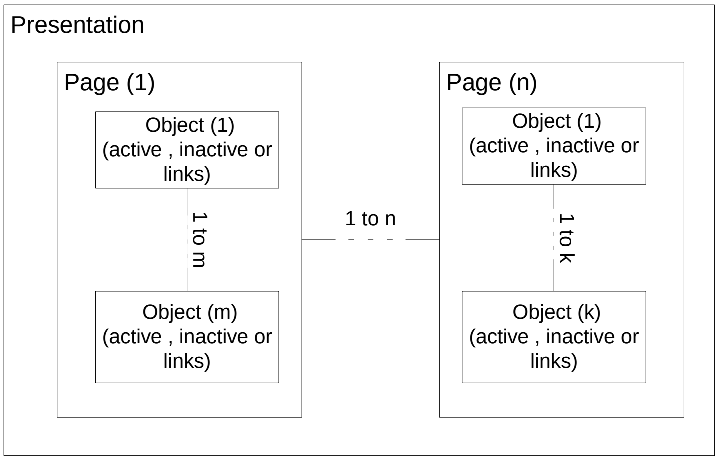
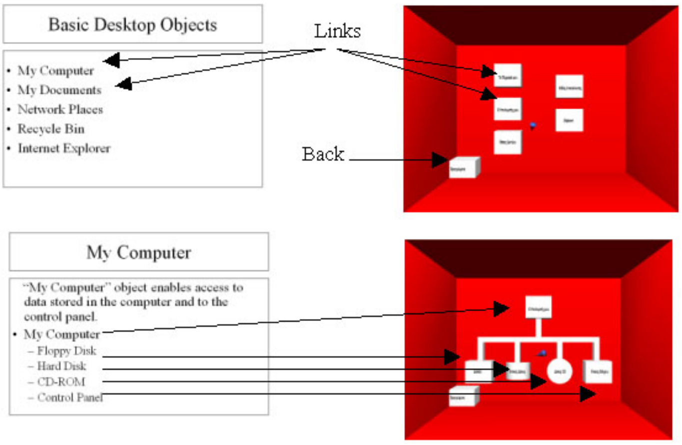


Figure 3-1 Presentation structure

When the cursor collides to an object or a part of the bounding box, force feedback is returned to the user. Thus, users are not allowed to penetrate any of the objects in the scene, or move the cursor outside the bounding box that defines the working area. In order to make the interface more comfortable for the user, all the objects are placed on the side of the bounding box that faces the user. That way, the user has to search using the cursor only one side of the box and not the whole 3D workspace.

Active Objects have two main characteristics: i) when the cursor contacts an active object it declaims a short description of what it represents and ii) when the user presses the stylus button, while the 3D cursor resides on the object, it declaims a detailed description of itself. The links have also the characteristic (i) of active objects. Another characteristic of link objects is that when the user presses the button on the stylus while the cursor resides on a link object, it declaims a detailed description of itself, a short description of changes that will happen in the presentation environment and finally changes the objects that exist in the current page with new objects necessary to present the selected page.

Currently, an initial version of a tool that converts HTML pages to haptic presentations has been created [4]. The HTML pages are initially processed and filtered in order to provide easy access to the main document, lists and links of each page. Figure 3-2 presents an original HTML page and the created haptic presentation.


Basic Desktop Objects

- My Computer
- My Documents
- Network Places
- Recycle Bin
- Internet Explorer

My Computer

"My Computer" object enables access to data stored in the computer and to the control panel.

- My Computer
  - Floppy Disk
  - Hard Disk
  - CD-ROM
  - Control Panel

Figure 3-2 Original HTML and haptic presentation.

In order to allow the user to study the presentation, bi-directionally, every page contains a link object called “back”. The user can return to the previous page of the presentation by clicking on that object. The “back” object is always placed on the left bottom corner of the bounding box. When the user is on the initial page of a presentation the object is deactivated.

3.3 Map Navigation

Map Navigation application allows the user to explore an area using an embossed map of a city area. The user can hear the names of the streets and explore the desired area. Specific places are also tagged on the map and associated with audio information. The user can explore the map using the 3D cursor.

3.4 Object Recognition

Object Recognition environment is an environment where the user can examine 3D models using the PHANToM device and understand their shape. The objects are imported from VRML files. The application allows the user to change physical characteristics of the objects like stiffness and damping, in order to make it easier to examine the 3Dshape of the object and to provide more realistic feedback [5].

4 Interface and usability

In order to provide a better description of the interfaces and the usability of the applications the cane application, which utilizes different hardware, setup will be described separately from the others.

4.1 Cane Simulation Interfaces

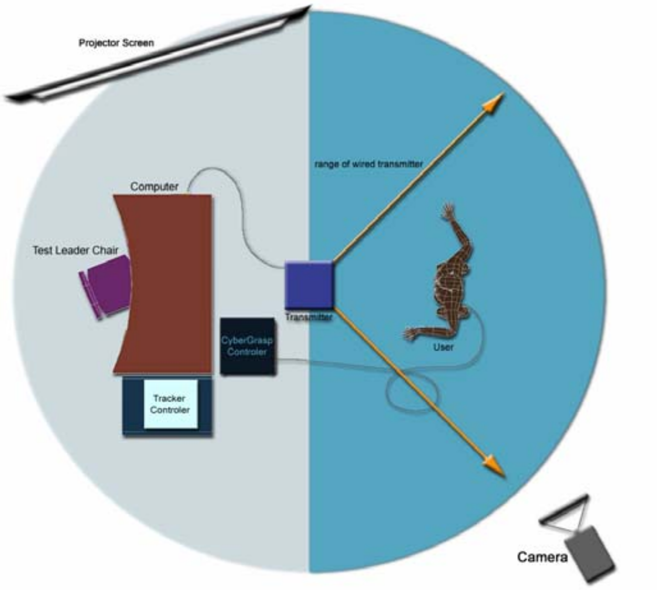


Figure 4-1 Cane simulation room setup

Figure 4-1 presents schematically the test setup for the cane simulation test case. The user wears the CyberGrasp and a waistcoat for carrying the Force Control Unit (FCU) for the CyberGrasp (Figure 4-2). Sound and haptic feedback are provided by the system upon collision of the cane with the virtual objects. The parameters of the virtual cane (size, grasping forces, collision forces) are adjusted so that the user feels that it is similar to the real one. Environmental sounds can be assigned to objects in the scene (e.g. realistic traffic lights sound is assigned to traffic lights in the virtual scene) [2].

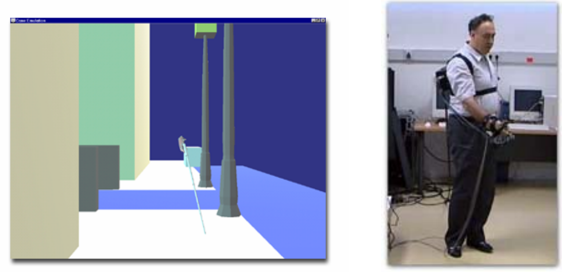


Figure 4-2 Cane simulation – Outdoors test – a) setup, b) a user performing the test.

There are two test cases implemented an outdoor test case and an indoors test case. In the outdoors test case the user is informed that he is standing in the corner of a pavement (Figure 4-2). There are two perpendicular streets, one on his/her left side and the other in his/her front. Then he/she is asked to cross the street in front of him/her.

The user should walk ahead and find the traffic light located at about one meter on his left side. Realistic sound is attached to the traffic light informing the user about the condition of the light. The user should then wait close to it until the sound informs him/her to cross the street passage (green traffic light for pedestrians). When the traffic lights turn to green the user must cross the two meters wide passage until he/she finds the pavement at the other side of the street. It is also desirable that the user finds the traffic light at the other side of the street.

Figure 4-3 Cane simulation – Indoors test : setup.

In the indoor test case the user is asked to navigate into an indoor environment using the virtual cane. When the test starts, the user is asked to grasp the virtual cane. The goal for the user is to find the second door on his/her left side and enter the room (Figure 4-3). There he/she should find a chair. During his/her walk the user should find successively the wall on the left side, the first door where he/she is not supposed to enter, the wall of the second room and the door where he/she is supposed to enter. After entering the room he/she should find the chair located in his right side.

4.2 Haptic Interfaces using the PHANToM™ device

Figure 4-4 Desktop setup for applications with PHANToM

Figure 4-4 presents the hardware setup for the desktop applications with PHANToM. The user controls a 3D cursor using the PHANToM™ Desktop device. Collision detection is performed between the cursor and objects in the scene and force feedback is send to the user. The user is allowed to move the cursor in the PHANToM workspace (approximately dimensions 12cm x 12 cm x 16cm).

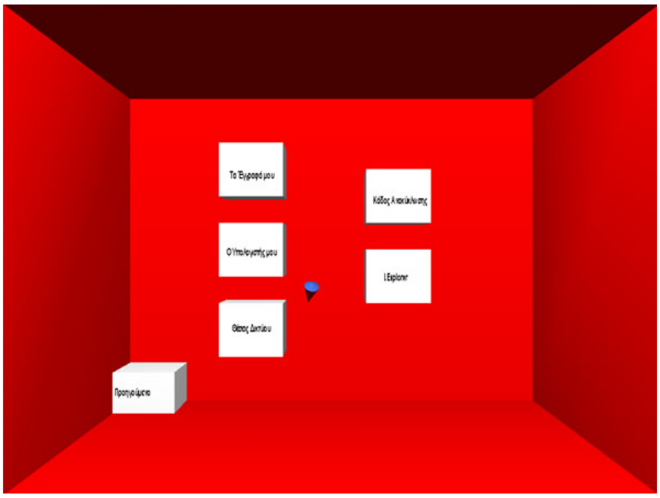


Figure 4-5 Interactive Presentation environment

Three different presentations and a map environment were created in order to evaluate the application. The first presentation is about the default structure of the folders of the Windows® 2000 operating system (Figure 4-5). The second describes the most common local network topologies. The last presentation describes the basic components of a relational database such as tables, relations and provides information about user privilege management. The map environment is a street map of a central area of Athens where the user can find streets that are located near the Pan-Hellenic Blind Association office. VRML 3D objects can be imported to the scene and examined by the users. The users can touch all kind of objects in the scene (such as active and inactive presentation objects, streets, buildings etc.) and listen to descriptions using the button that resides on the PHANToM stylus.

5 Evaluation

Initial versions of the applications have been evaluated with blind and visually impaired users. Specifically the white cane simulation has been tested with blind and visually impaired users from the Thessaloniki Blind School and the Pan-Hellenic Blind association. The interactive haptic presentation has been tested with blind users from the Pan-Hellenic Blind association. Evaluation of the cane simulation and the desktop application have been performed by different groups of users and will be examined separetally.

5.1 Evaluation of the cane simulation application

Twenty-six persons participated in the tests from the Thessaloniki Local Union of the Panhellenic Accosiation for the Blind in Greece. The users were selected so as to represent the following groups: blind from birth, blind at a later age, adults, and children. The evaluation consisted of three phases. In the first phase the users were introduced to the system and were alloud to use it for a while in order to get used to the device and to calculate the most confortable parameters for the cane (i.e. length, force amplitude)[3].

In the second phase the users performed the tasks. The total time to complete the task, alloud comments and success or failure in performing the task were recorded for each user. In the third face the users answered to quastionaire, about the performance and the usability of the system.

According to the comments of the users during the tests and the questionnaires filled by the users after the tests, the following conclusions can be drawn: It was deemed very important to utilize both acoustic and haptic feedback, as they are indispensable for the orientation. It is also important to note that a percentage ranging from 90-100% of the users have characterized the tests as useful or very useful.

The difficulty level of the tests was reconsidered after completion, according to the percentage of the users that needed guidance and the rank that users gave to each test case. The users were asked to rank the challenge of each test using a scale between 1 (easy) and 5 (very difficult). The users that needed guidance to perform the test was 3.8%. The average rate of the challenge of the test, according to the users, was 2.65.

5.2 Evaluation of the interactive presentation, map and object recognition environments

Twelve blind and visually impaired users participated in the evaluation tests of the application. The users ages ranged from 20 to 34 years old. Ten of the users were male and two were female, eight were blind and four visually impaired, eight were students, three unemployed and one employed. Also seven of them were blind from birth and five of them acquired the impairment at a latter age. Six of the users have heard about haptic devices before participating in the evaluation and two of them have used haptic devices in the past. None of them had used the PHANToM™ device in the past. All the participants were members of the Panhellenic Blind Association (the major blind association in Greece), had attended seminars on computer usage and were familiar to conventional computer interfaces for blind users.

The evaluation procedure consisted of three major steps: In the first step, the users were introduced to the system. The introductory training took approximately ten minutes. The majority of the users had no particular problems when interacting with the system. The motivation for this pre-test was to introduce the users to a technology completely unknown to them, while ensuring that they feel comfortable with the environment of the laboratory.

In the second step, the users were asked to use the application. The evaluation tests took approximately one hour per user, including pauses between the presentations. The purpose of the evaluation was not to test the reaction of a user to a haptic system. Rather, the idea was to try to obtain information about the use of such a system by a user who is somewhat familiar with the use of haptics. During the test procedure the test leader was monitoring the performance of the users and asked questions in order to find out whether the users understand the presented structures.

In the final step, all the users were interviewed in order to complete an after test questionnaire. The after test questionnaire consisted of two parts: The first part included information about the users like age, previous experience of using haptic interfaces and their expectations from the technology. The second part of the questionnaire was about the ease of use of the specific interface and the interest that the users show for such applications.

6 Conclusions

In terms of usability, we can conclude that the system can be used for educational purposes, mobility, orientation training and exploration / navigation in 3D spaces.

In the case of the cane simulation technical limitations constrain its applicability. Specifically, the system cannot prevent the user from penetrating objects in the virtual environment. The maximum workspace is limited to a 7 m - diameter hemisphere around the tracker transmitter (the 1 m limitation, caused by the CyberGrasp™ device is solved by using a backpack so that the user can carry the CyberGrasp™ actuator enclosure). The maximum force that can be applied is limited to 12N per finger and the feedback update rate is 1KHz.

In the case of the desktop applications the workspace is limited to a square of approximately 12cm x 12cm x 16 cm. The advantage however in this case is that the device is mounted and the user cannot penetrate objects in the virtual environment.

In all cases the following conclusions can be drawn from the evaluation of the Feasibility Study tests in terms of system usability:

- It was deemed very important to utilize both acoustic and haptic feedbacks, as they are indispensable for the orientation. The majority of the participants preferred to have both feedbacks.
- Most of the participants were very positive about beginning with simple objects and then proceeding to more and more complex ones. Some of them would have liked to deal with more complex scenarios.
- All people tested had no problems with the system after an explanation of the technology and some exercises to practice the application.
- The participants needed little or no guidance at all, i.e. the users had no difficulties to handle the software and the devices. On the contrary, they enjoyed completing their tasks, showed a lot of commitment and were very enthusiastic about being able to have this experience.
- No connection was found between the age that blindness occurred and the test results.
- All participants emphasized their demand to use these programs in the future.

- All the users stated that they would like to participate in future tests.

Concluding, the result has unanimously been that the application introduced was considered very promising and useful, whereas it still leaves a lot of room for improvement and supplement. Provided that further development is carried out, the system has the fundamental characteristics and capabilities to incorporate many requests of the users for a very large pool of applications. The approach chosen, fully describes the belief of blind people to facilitate and improve training practices, and to offer access to new employment opportunities. It represents an improvement of life for the blind and the visually impaired people when connected to reality training. These facts are evident from the participant's statements.

Besides the direct benefits of the proposed system, as many of the users mentioned, technology based on virtual environments can eventually provide new training and job opportunities to people with visual disabilities.

7 References

1. D. Tzovaras, G. Nikolakis, G. Fergadis, S. Malasiotis and M. Stavrakis, ``Design and Implementation of Haptic Virtual Environments for the Training of Visually Impaired'', IEEE Trans. on Neural Systems and Rehabilitation Engineering, Vol. 12, No. 2, pp.266-278, June 2004.
2. D. Tzovaras, G. Nikolakis, G. Fergadis, S. Malassiotis and M. Stavrakis, "Design and Implementation of Virtual Environments for Training of the Visually Impaired", International ASSETS 2002 SIGCAPH ACM Conference, Edimburg, UK, July 2002.
3. D. Tzovaras, G. Nikolakis, G. Fergadis, S. Malassiotis and M. Stavrakis, "Feasibility Study of a Human-Computer Interaction System for the Training of Visually Impaired", 2nd Hellenic Conf. on Artificial Intelligence (SETN-2002), Apr. 2002, Thessaloniki, Greece.
4. G. Nikolakis, I. Tsampoulatidis, D. Tzovaras and Michael G. Strintzis, "Haptic Browser: A Haptic Environment to Access HTML Pages", to appear in SPECOM 2004.
5. G. Nikolakis, D. Tzovaras, S. Moustakidis, M. G. Strintzis, "Cybergrasp and PHANTOM integration: Enhanced Haptic Access for Visually Impaired Users" ", to appear in SPECOM 2004.

Description of the SINERGIA laparoscopy virtual simulator

Enrique Gómez, Pablo Lamata and Samuel Rodríguez
GBT, Polytechnic University of Madrid

July 2004

1 Introduction

This document describes the laparoscopic surgical simulator from the UPM node following the SIMILAR Usability SIG (WP07) system description structure. This simulator is being developed by a Spanish cooperative network called SINERGIA, funded by the Spanish Ministry of Health.

We describe in this report the first prototype (PT1) of the system, which is being developed. This research departs from the work of MedICLab, where an early prototype was developed [Monserrat et al. 2003].

1.1 Purpose of the application

The main goal of the laparoscopic simulator is to enhance the surgical training process. This system offers a virtual environment similar to the operating theatre that enables a surgeon to practice different surgical procedures as many times as required.

1.2 Input and output modalities

The surgeon interacts with the simulator using two sensorial channels:

- Visual: a monitor displays the virtual surgical scene, as it may be captured with a laparoscope.
- Haptic: a haptic feedback device that simulates the haptic interaction of the surgeon. This device provides force feedback to the user, who can feel the consistency of virtual organs.

1.3 Target user group and physical use environment

Target users are mainly surgeons with very little laparoscopic experience. The simulator will be integrated in the training curriculum of the workshops in laparoscopic surgery course organized by the MISC (Minimally Invasive Surgery Centre) of Cáceres (Spain).

1.4 Domains and tasks

The simulator is focused on simulating different basic steps of Nissen founduplicature. This surgical procedure aims to solve the Gastroesophageal reflux disease, a failure of the antireflux barrier that allows abnormal reflux of gastric contents into the esophagus. The first prototype of the simulator will have four basic steps: grasping and pulling, cutting, dissection and suture.

1.5 System accessibility

The system has been demonstrated to our surgical partners, but it is a research prototype under development and we do not provide freely available demos for download.

2 Technical issues

This section describes technical aspects of the laparoscopic surgical simulator in terms of platform, hardware requirements, implementation language, and architecture.

2.1 Platform

The simulator makes use of two personal computers (PCs) in a client-server architecture. One PC runs the simulation software (collision detection, biomechanical model and visual rendering) and the other controls the haptic device, sensing the position of the virtual tools and rendering the interaction forces to the user.

2.2 Hardware requirements

One special hardware is needed for the simulator: the haptic device. Our choice has been the Laparoscopic Surgical Workstation (Immersion Corp., San Jose, U.S.A.). The two PCs needed do not need special requirements, but it is better to have as much computational power as possible in order to increase the real time performance. It have been observed that running in a PC Pentium III, 450 MHz, with 256 RAM memory, visual and haptic update rates of 15 and 500 Hz can be reached.

2.3 Implementation language

The simulator can run on a Windows NT/2000/XP platform, and is written in C++. For its developing has been used Microsoft Visual C++ environment.

2.4 Architecture

The simulator is composed by four main modules controlled by the main programme, the architecture can be observed in Fig 1:

- Biomechanical model: calculates de deformation and the behaviour of the organ in the virtual scene.
- Collision module: calculates the interaction between the virtual models
- Visual motor: represents the geometry in the visual device (screen)
- Haptic motor: reads the positions of the haptic device and returns the haptic forces to the user.

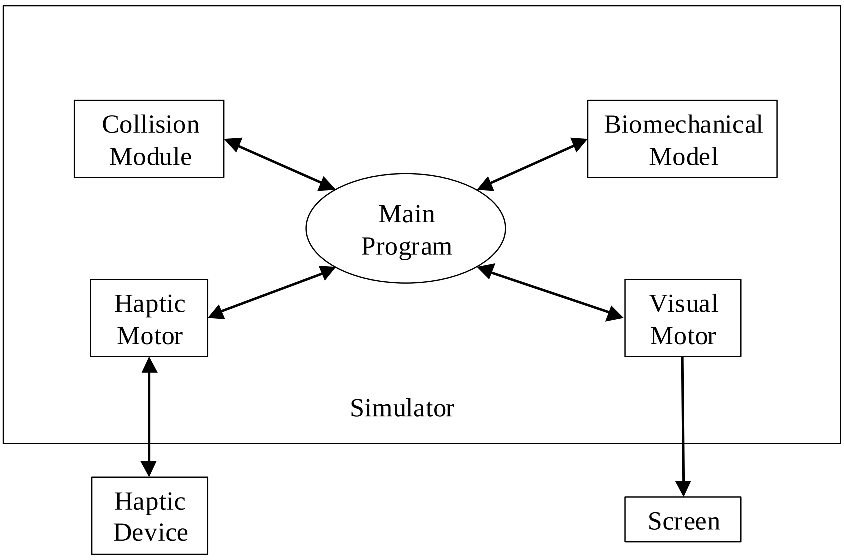


Fig.1 Sinergia Simulator Arquitecture

As mentioned before, all these components are controlled by the simulator main program that executes the following steps is its main simulation loop:

1. Read the positions of the tools represented by the haptic device.
2. Calculate the collisions between the elements in the scene and the response of these collisions.
3. Calculate de deformation of the deformable models that represents the organs.
4. Display the new geometry and the reaction forces resulted from the deformation process.

3 Functionality

The simulator is task-oriented, i.e. there are different particular tasks which the user is meant to perform with the system. In our PT1 four will be these surgical tasks:

- **Grasping and pulling**: the user has to orientate the surgical tools to grasp different objects and pull them until tearing. With this basic task users familiarise with different tissue consistencies and tearing thresholds, learning how to interpret visual and haptic information.
- **Cutting**: the user has to cut a surface through a drawn line on it. Metrics of precision, time and economy of movements are recorded, which allows the user to make an auto-evaluation of the task, providing constructive feedback.
- **Dissection**: the user has to dissect two virtual layers of a virtual stomach. Metrics of errors, time and economy of movements are recorder.
- **Suture**: the user has to perform a virtual suture on a flat surface.

4 Interface and usability

As explained before, the user interacts through two sensorial channels at the same time: visual and haptic. When he moves a virtual tool with the haptic device, this tool is moved in the monitor and interacts with the virtual organs, and he feels the interaction forces. And all this is performed several times in a second (visual update rate: 15Hz, haptic update rate: 500Hz). The simulator, with its visual and haptic interface, is shown in Fig. 2.

Fig. 2: Surgical simulator

There are no skills assumed for the user. Any case, users are supposed to have some knowledge of surgical tools, terminology and procedures.

The application starts with a menu in which the user selects the training task he wants to practise. And when the task is finished, a report is provided to the user to provide constructive feedback about his performance.

5 Evaluation

The PT1 is still on development, but some functionality is already finished. The system has been shown to some expert physicians who have assessed its face validity. They have described the training tool as very promising.

In our network we are paying special attention to the requirements of the simulator to be an effective educational tool. Some research has been done conducted to the study of the sensory interaction in the laparoscopic theatre [Lamata et al. 2004]. The aim of this research is to find which components of the human-machine interaction are essential to the training purpose, and which are not.

Our evaluation studies will be performed during the workshops in laparoscopic surgery course organized by the MISC (Minimally Invasive Surgery Centre) of Cáceres (Spain). Different expert and novice users will use the system in a construct validity study. We are also planning to conduct studies about the training transfer from the virtual reality environment to the operating room.

6 Conclusion

The surgical simulator of the SINERGIA network is being developed for the laparoscopic training of surgeons. The simulator will be inserted in the training curriculum of the MISC of Cáceres (Spain) where the evaluation studies will be performed.

7 References

[1] C.Monserrat, O.López, U.Meier, M.Alcañiz, C.Juan, y V.Grau, "GeRTiSS: A Generic Multi-model Surgery Simulation," Proc. Surgery Simulation and Soft Tissue Modeling Interntaional Symposium, Juan-Les-Pinnes (Fr), Lecture Notes in Computer Science 2673, pp. 59-66, 2003.

[2] P.Lamata, E.J.Gómez, F.J.Sánchez-Margallo, F.Lamata, F.Gayá, J.B.Pagador, J.Usón, y F.del Pozo, "A new methodology to characterize sensory interaction for use in laparoscopic surgery simulation," Proc. International Symposium on Medical Simulation (ISMS), Cambridge, MA, USA, Lecture Notes in Computer Science 3078, pp. 177-184, 2004.

Description of an Image Guided Application Using Contextual Focus Driven Interaction

Daniela G.Trevisan, Monica Gemo, Vincent Nicolas,
Jean Vanderdonckt and Benoit Macq
TELE and BCHI – Université Catholique de Louvain

September 2004

1 Introduction

This document describes an Image guided system, following the SIMILAR Usability SIG system description structure in order to ensure similarly structured system descriptions at similar levels of detail and to provide sufficient information for undertaking usability evaluation comparisons of the systems in question, and for documenting current practice.

The related system is being developed as part of the MERCATOR (Méthodes de co-Registration et Constitution d'ATlas pour Opération chirurgicale et Radiothérapie) project under contract WALEO 21/5129.

1.1 Purpose of the application

Image guided surgery is a type of computer assisted surgery which uses advanced three dimensional visualization techniques to provide the surgeon with a wealth of valuable information not normally available in the operating room. By essence, a complex surgical procedure can be navigated visually with great precision by overlaying on an image of the patient a colour coded preoperative plan specifying details such as the locations of incisions, areas to be avoided and the diseased tissue. The added information can be anything from text to icons or colour-coding to three-dimensional surfaces.

1.2 Input and output modalities

The user should first select the information what s/he wants to have in the augmented scene. The user interacts with a mouse and keyboard for instance to entry the text information, segment and reconstruct organs and define audio information for output.

As output modalities the composite scene can contain text, 3D objects, live video and audio data. The focus-driven user interaction with the augmented is done based on the user's position.

1.3 Target user group and physical use environment

Target users are the surgeons in the operating room. The physical environment is an operating room described in Figure 1.

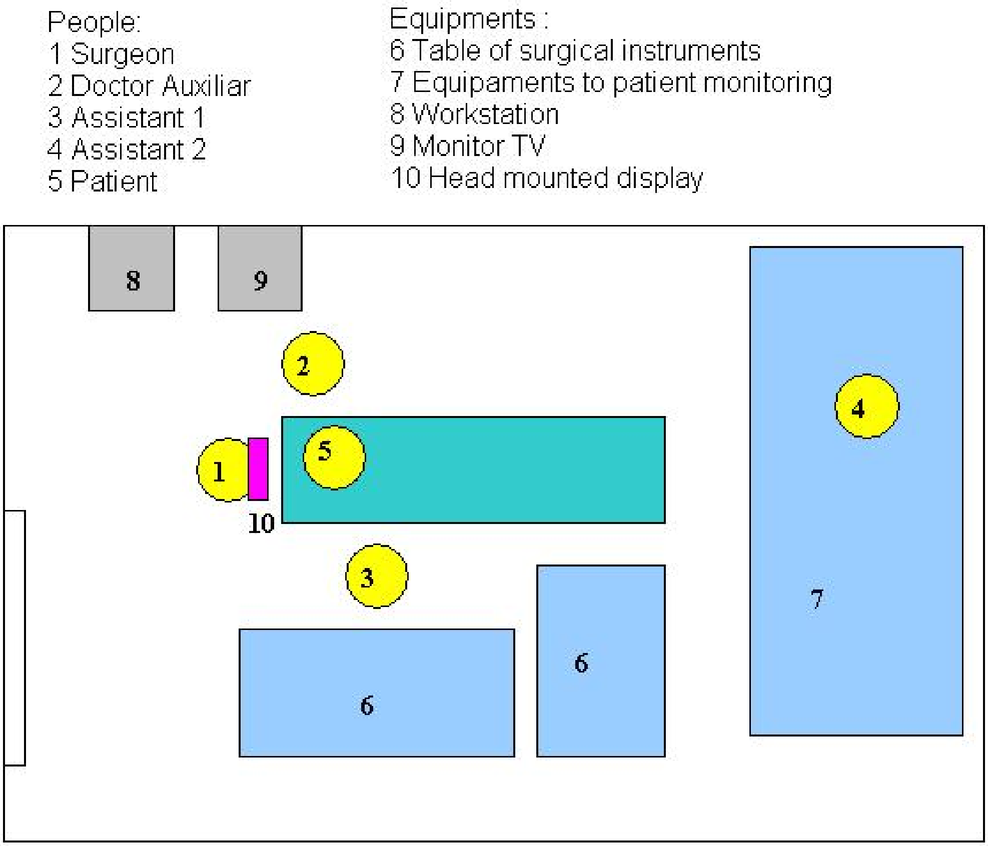


Figure 1. Operating room scenario.

1.4 Domains and tasks

We have started to model the craniotomy task involved in a neurosurgery. Usually this task consists of drawing over the patient's head the path to open the head. Three technical stages are necessary in order to achieve guidance for the surgeon during this task (Figure 2):

1. In the pre-operative planning, the surgeon selects the structures of interest that must be segmented from the pre-operative images and may define additional information, such as annotations relevant to task execution as well as the order for auditory output. These sub tasks are represented in the CTTE model showed in Figure 3.
2. .At the pre-operative calibration, the segmented structures must be registered with the physical scene, and any movement of the patient or surgeon compensated for. The camera parameters are calibrated. As well as the spatial and colours filters parameters.
3. Finally, the virtual structures with additional annotations must be visualized and/or auditory rendered during the surgical intervention, i.e. intra-operative phase. These sub tasks are represented in the CTTE model showed in Figure 4.

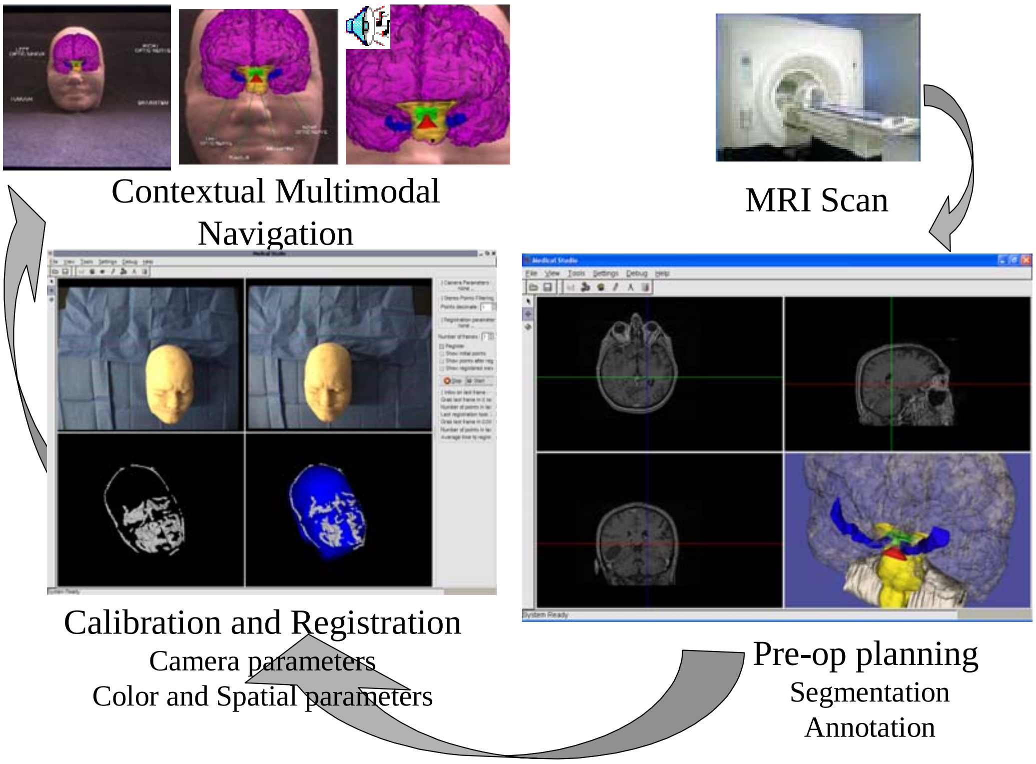


Figure 2. Image guided system phases.

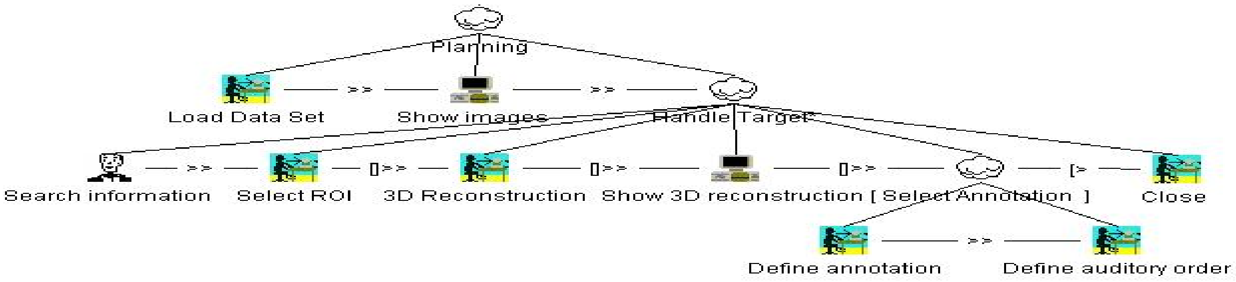


Figure 3. Pre-operative planning with sequential and optional tasks.

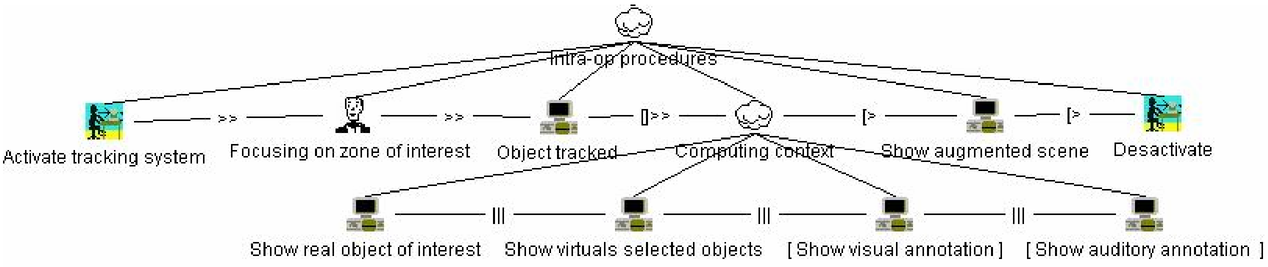


Figure 4. Intra-operative tasks with parallel tasks.

1.5 System accessibility

This application is a research prototype under development and we do not provide freely available demos for download.

2 Technical issues

This section describes technical aspects of the imaged guided application in terms of platform, hardware requirements, implementation language and architecture.

2.1 Platform, operation system, hardware requirements and implementation language

The proposed hardware setup consists of the following components:

- • stereoscopic goggles;
- • stereo video camera (Bumblebee camera from Point Grey Research Inc.).

The augmented scene can be displayed in an AR video-based goggles. Those goggles are opaque, but as we put the images captured by the stereo cameras right in front of them, they become virtually transparent. In the optimal case the distance between the camera lenses must correspond to the distance between the screens of the AR goggles and the distance between the operator’s eyes (Figure 5).

Figure 5. Head mounted display using stereovision cameras.

This application is written in C++. The language is commonly used by the signal and image processing community. Publicly available libraries are used to provide well known functionalities and avoid re-implementation of already validated methods. These libraries have been chosen in function of their specifications, their development language, their cross-compilation possibilities and the community working with them. VTK [1] is used for all visualization tasks. Gtk [2] is used for all graphical user interfaces. ITK [3] is used for signal and image processing. Dcmtk [4] is used for Dicom 3.0 standard compliance. This list of libraries is not fixed. Thanks to the modular architecture any other library can be linked into a new component providing high extensibility to Medical Studio. We define a plug-in as a group of components compiled together into one shared library. These plug-ins can group components by any criteria but by convention it is preferable to group them either by type or procedure context. An xml file is created for each plug-in describing its content. With this, the kernel is able to create a repository of all available components without loading them into memory, plug-ins will be loaded lately, only when required letting resources available for data or processing. This facilitates the distribution process as only two files will be needed to add this procedure to the basic Medical Studio platform. This allows easy customization of final applications. For example all the components for a specific neurosurgery procedure will be grouped into one plug-in. For applications that do not need the neurosurgery plug-in (e.g. maxillo-facial surgery), removing the two files is enough to eliminate the unwanted functionalities.

The stereo pictures provided by the stereo camera are processed through stereovision algorithms (Digiclops and Triclops libraries provided by Point Grey Research [5] in order to obtain a cloud of points of the real scene.

2.2 Architecture

A framework allowing the centralization of all tasks of assisted surgery must have a consistent and evolvable architecture. For that reason he have developed such application as a plugin into the Medical Studio Framework[1] which uses a component-based architecture. The software architecture implemented in Medical Studio is a modified version of the ARCH model [6] (Figure 6(a)) to add a specialization of the input and output modalities.

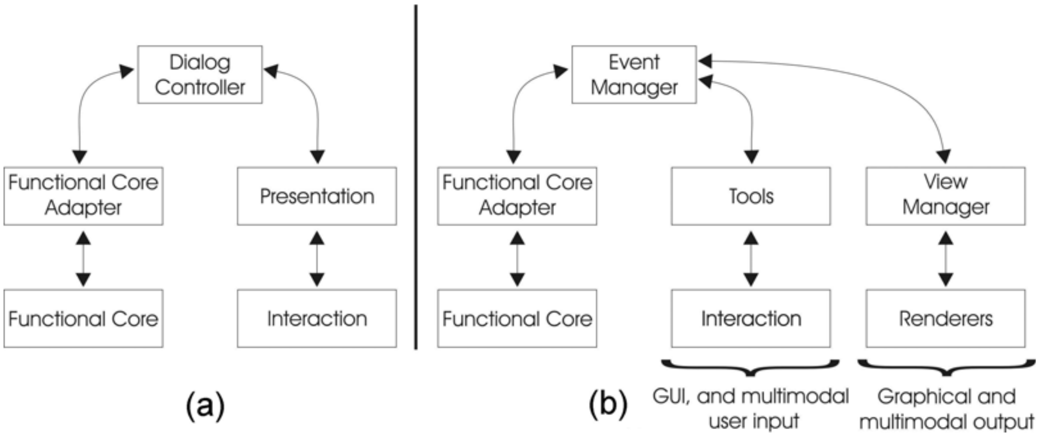


Figure 6. (a) The ARCH architecture model. (b) Medical Studio architecture based on the ARCH model.

The principal components of the architecture illustrated in Figure 6 (b) are detailed as follows:

Functional Core: contains all data processing components such as registration algorithms, I/O filters, etc.

Functional Core Adapter: is an abstraction layer that allows the event manager to communicate with the functional core.

Event Manager: is the dialog controller between interactions and functional core. When interactions occur, the event manager will propagate them to the functional core, and if needed interpret them before.

Interaction: represents user input components such as mouse, keyboard, magnetic pen, vocal recognition, etc.

Tools: are components that map interactions into data processing or visualization modification. For example, a mouse click will be translated into a rotation (view manager) by

---

[1] Available at www.medicalstudio.org

the navigation tool, but it will be translated into a data modification (functional core) by a segmentation tool.

View Manager: manages all views and knows how data can be visualized on which type of view.

Renderers: are components which render specific data type onto specific views. An example is the rendering of an image on a 3D view with the raycasting algorithm.

In order to clarify the development process and to fix where components should be in the visualization pipeline, Medical Studio classified them into operators. The architecture groups the data operators into data components, the visualization operators into output components, and adds input components (Figure 7). There are plans to affine this subdivision to add more control on the visualization pipeline.

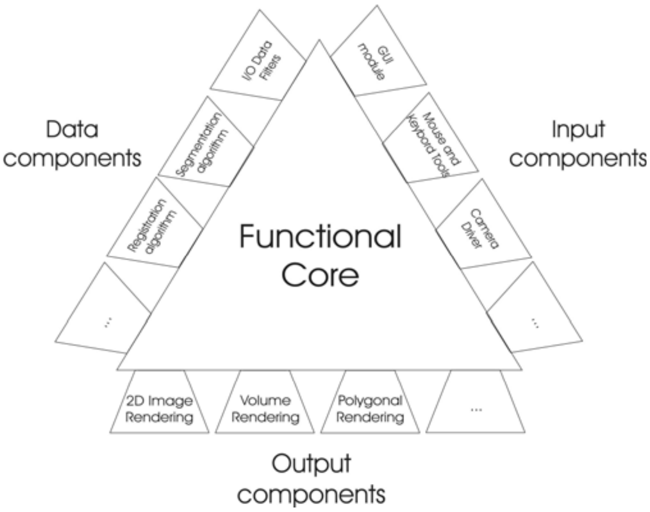


Figure 7. Medical Studio component-based architecture.

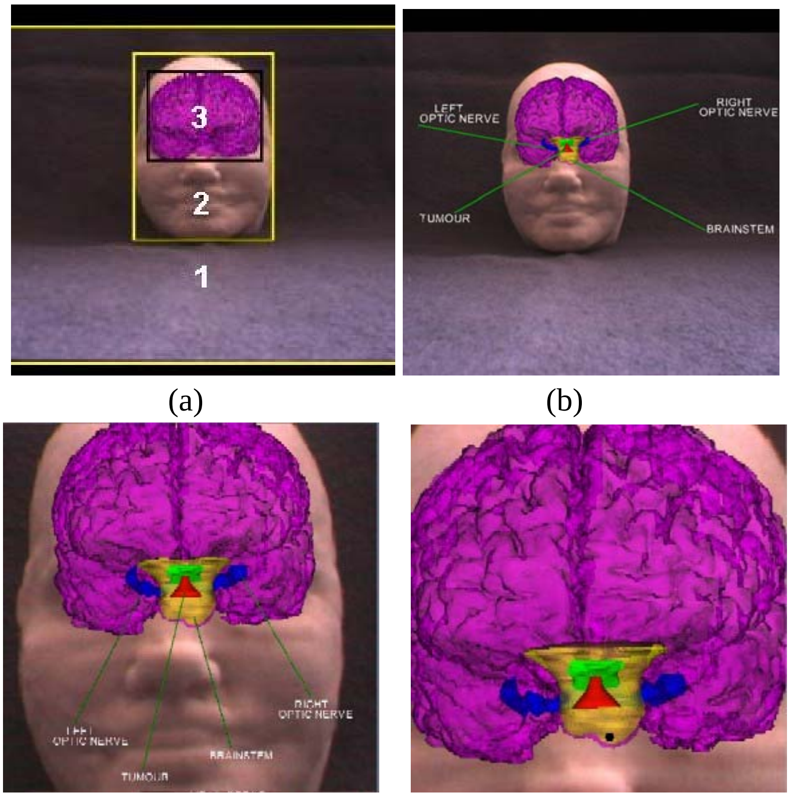
3 Functionality

The elements composing the multimodal scene for our case scenario are the patient's head, registered virtual objects, such as the patient's anatomical organs or segmented functional zones, and annotations pointing to specific sub-areas that need to be preserved during surgery. While planning the surgical paths the user selects the anatomical organ that is of primary concern for his task (the registered virtual object). This is the object on which the surgeon will fix his attention during the intra-operative procedures. Besides, as the user wears a stereo-video head mounted display with a camera on top of helmet the physical environment is configured in such a way that the registered virtual object is always displayed in the central zone of the view and its position, though being still influenced by the viewing specification is constrained as it is registered with the patient's head. This does not hold for annotations: their position is flexible to ensure that they are overlaid close to related objects and to prevent undesired occlusions. The user can make the following assumptions to define the layout strategy and place the annotations in the scene:

1. First of all annotations cannot overlap one another for readability (so we could say that annotations too have the highest priority).
2. The registered virtual object has the highest level of priority and annotations can never be overlaid on it.

3. The tracked real object has medium priority, which means that, depending on context, annotations may be placed onto it.
4. Others objects are considered as part of the background and have the lowest priority, thus annotations can always be superimposed on the background.

Another rule is that annotations are always placed from the extremes to the centre of the view to occupy peripheral space at best. To implement the requirements for annotation placement the user specifies 4 types of bounding boxes (Background, Real object, Virtual object, Annotations) in the scene defining three different levels of priority. The latter are respectively shown in Figure 8: the region labelled as 1 has the lowest priority, the region labelled as 2 has medium priority and the region labelled as 3 has the highest priority. The algorithm first tries to place annotations only in the level 1 area avoiding intersection with the level 2 zone. If the available area is not sufficient, the algorithm tries to place annotations by spanning also over the level 2 zone, but avoiding to overlay annotations onto the level 3 zone. If this is still not possible, the modality for annotation rendering is automatically switched to audio output. Annotation localization is preserved by visualizing the corresponding anchor point onto region 3. The audio visual output follows the order defined by the user in the pre-operative phase. Then the condition that forces the modality switching for annotation rendering can be related to the overlayable area of regions 1 and 2: when it drops below a certain value, the switching is performed. The annotation size is fixed and constant. Even if the scale factor of the scene changes (i.e., as the user's head moves closer or farther), the size of annotation remains the same. In addition, the virtual objects being added to the real scene depend on the side of the real object that the user is viewing. For instance, if the user is focusing the right side of the patient's head then only the anatomical structures and the visible annotations corresponding to the right side of the brain should be visualized.



(a) (b)

(c) (d)

Figure 8. Context-aware annotation rendering in the composite scene: (a) Priority regions in the composite scene (b) Background Context (c) Background/Real Object Context (d) Virtual Object Context.

4 Interface and usability

To avoid inconsistent behaviour while supporting the different contexts of use previously highlighted, we consider the framework of plastic user interfaces [7]. A plastic user interface is capable of adapting to variations of the context of use while preserving predetermined usability properties, which in our case is continuity of interaction as it is crucial for surgical systems. Plasticity impacts information being presented to the user and tasks to execute. For simplicity we consider variations in user focus when only the user's distance from the real scene changes and not orientation. This is because variations in user's position account for the most predominant contribution in adaptation. As we have already discussed in the previous subsection, our actual composite scene can adapt to three classes of context of use and operate in two modalities. The first modality corresponds to the visual presentation model for annotations, while the second combines the visual and audio output. Figure 9. shows indicative threshold values for user's distance from the scene and orientation. These influence and determine user's focus of attention and thus the context of use. Supposing the system is operating within context 1, as user focus increases (that is the user gets closer to the scene), the area decreases and at a time the system performs the switch to context 2 (see Figure 8 and Figure 9). The same behaviour holds for the system operating in context 2 as user focus increases. When the area where annotations can be superimposed is not sufficient anymore, the system switches to context 3 (see Fig.8 and Fig.9). The presentation model for the annotations now couples visual and audio representations. The visual representation highlights the location of interest, while the audio renders the same information (organ functionality, volume, etc.) that was represented visually in the other two contexts. The graphics highlights contexts at the boundary, as plasticity thresholds. These require particular care in interface design to ensure usability is preserved during transitions and are essential system parameters to be determined through empirical usability tests establishing the values of physical entities that control them.

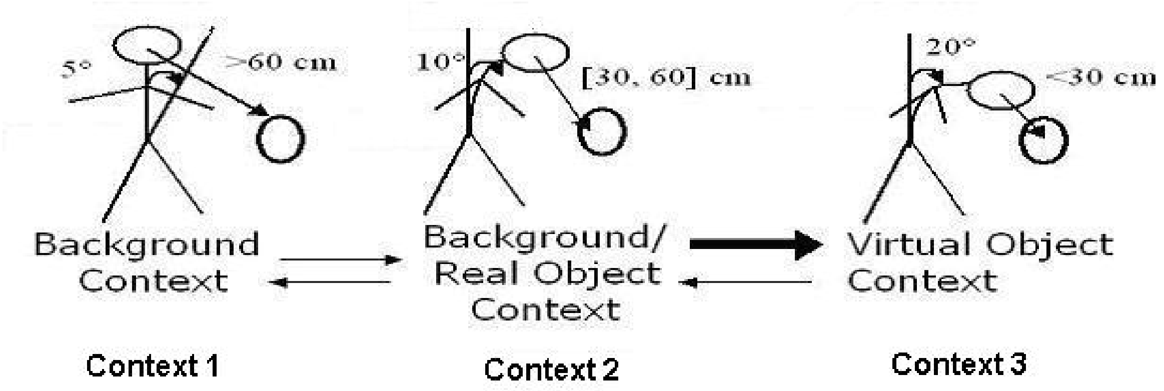


Figure 9. Contextual values to set changes in the scene composite.

The predetermined metric chosen in the design of the composite scene to evaluate the usability of our system is continuity [8]. This is the capability of the system to promote a smooth interaction scheme during task accomplishment considering perceptual, cognitive and functional properties. The perceptual property is defined as an ability of the system to make all data involved in the user's task available in one perceptual environment in order to avoid changes in the user's focus. The cognitive property is defined as an ability of the system to ensure that the user will correctly interpret the perceived information and that is correct with regards to the internal state of the system. The functional property is related to the effort of the user in experiencing a new interaction mode. A preliminary analysis of the adaptation support in our composite scene shows that there are no perceptual discontinuities when the system operates in context 1 and 2. This is because all annotations are represented in the same visual environment, whereas the presentation for context 3 is split in two separate perceptual (i.e. visual and auditory) environments. Presentation based on contexts 1 and 2 do not present any cognitive discontinuity, the user easily and immediately interprets annotation text and localization. Whereas context 3 requires a higher cognitive workload as information must be combined from two perceptual environments to interpret the concept. Regarding functional continuity, the presentation does not involve discontinuity in any contexts. The user interacts with the system by using always the same mode: hand-free user focus driven control. More accurate usability evaluation is required to validate the design, parameterize the system and practically assess the user experience. Continuity can be also used to estimate the migration cost for physical, cognitive and conative efforts [9] that users have to pay when switching between contexts in different plasticity sub-domains (the transition arrows in Fig.9). To better differentiate the two metrics, we propose to name intra-context continuity, the property referring to interaction schemes within a context, and inter-context continuity, the property addressing smooth interaction when switching contexts at plasticity thresholds. Our design does not present any perceptive, cognitive or functional inter-context discontinuity during the transition between contexts 1 and 2, whereas the transition between 2 and 3 shows perceptive and cognitive inter-context discontinuities. Fig.9 assesses the inter-context continuity of our composite scene, where arrow thickness depicts human costs.

5 Evaluation

Taking account the craniotomy task three types of interaction measurements will be considered: perceptive, cognitive and functional. For the perceptive and cognitive workload each design property will receive a weight and after the experiment, participants complete the post-test questionnaires, including the NASA TLX rating. For the functional property the task performance will be taken account and the surgeons should validate it. By the time of this writing, the experimental phase is in progress and as result of these experiments we intend to provide a set of guidelines based on perceptive, cognitive and functional properties to develop smooth interactions for Image guided systems.

6 Conclusion

Basically we are interested in measure the user’s interaction with the system during the surgical intervention. Then the continuity of task and interaction are very important points for the success of the application. A new multimodal system based on contextual focus driven interaction is being developed.

7 References

1. Schroeder, W., Martin, K. and Loresen, B.: The visualization toolkit. Kitware, 2002.
2. Blandford, J., Clasen, M., Janik, T. et al.: GTK software. [Online]. Available: http://www.gtk.org/
3. Schroeder, W. and Ibanez, L.: The itk software guide. The Insight Consortium.
4. DICOM OFFISComputer Science Institute. DCMT software. [Online]. Available: http://www.dicom.offis.de/dcmtk.php.en/

5. (2003) The Point Grey Research website. [Online]. Available: http://www.ptgrey.com/
6. Bouchet, J. and Nigay, L.: Icare: A component-based approach for the design and development of multimodal interfaces. Proceedings of 11th International Conference on Human-Computer Interaction HCI International. Vienna: Lawrence Erlbaum Associates, 2004.
7. Calvary, G., Coutaz, J. and Thevenin, D.: A unifying reference framework for the development of plastic user interfaces. Proceedings of the 8th IFIP International Conference on Engineering for Human-Computer Interaction. London, UK: Springer-Verlag, 2001, 173–192.
8. Trevisan, D., Vanderdonckt, J., and Macq, B.: Continuity as usability property. Proceedings of 10th International Conference on Human- Computer Interaction HCI International. Mahwah: Lawrence Erlbaum Associates, 2003, 1268–1272.
9. Dowell, J. and Long, J.: Toward a conception for an engineering discipline of human factors. Ergonomics, vol. 32, 1989, 1513–1535.

Medical Studio

Vincent Nicolas
TELE - Université Catholique de Louvain

August 2004

1 Introduction

- Purpose of the application

  Help surgeon during the whole surgery, from planification to intra-operative guidance. The application allows surgeon to visualize acquired medical images and interact with it. Examples of interactions are 3D real time navigation, manual or automatic segmentation, computation of distances or volumes. At this point we are terminating a wizard for maxillo-facial planification procedure. It concerns the positioning of well defined points and planes, which are used to compute distances, angles and create final planification report.

  I propose to constraint this evaluation procedure on this well defined procedure.
- Input modalities

  Medical Images acquired (CT-SCAN)

  User manual input (mouse, keyboard)
- Output modalities

  Screen

  (Possibility to use stereovision glasses)
- Target user group(s)

  Surgeons
- Physical use environment

  On desktop, with keyboard and mouse.
- Which domain does the application cover

  Medical Imaging, planification
- Which tasks (if any) does the application solve

  Surgical planification in maxillo-facial surgery
- Is the application free or what is the price

  Should be free … but under NDA
- If not free, is a demo available?

2 Technical issues

- Platform(s) (operation system(s))

  Linux, Windows
- Hardware requirements

  Pentium IV 2.5 or equivalent

  3D Graphic accelerator (NVIDIA GeForce 4 ti 4600 or equivalent)

  512Mb Ram (at least, 1Gb is preferable)
- Implementation language(s)

  C++ (using VTK, ITK and Gtk libraries)
- Architecture

  Intel x86 compatible

3 Functionality

- Which functionality does the application offer

  - Medical Image visualisation (CT, MRI, PET, …Dicom, Ecat, Analyze, …)

- 3D reconstruction
- Computation of distances, volumes, angles
- Easy creation of wizards

- Description of each main functionality

4 Interface and usability

- Description of interface design and possible design for usability

  No usability design have been created.
- Which user skills (if any) are assumed

  No particular, base of computer skills
- Is it walk-up-and-use? Is training foreseen? Is there a manual?

  It's working but procedure and advanced functionalities are not yet completely implemented.

  Manual must still be terminated.
- Illustrative examples of use and interface

  Screenshots available on the website http://www.medicalstudio.org
- Advantages and disadvantages of functionality and interface

  Not yet evaluated … some functionalities could be enhanced.

5 Evaluation

- Who/how many have used the application so far

  3 developers and 2 surgeons
- How has the application been usability tested (detailed description of methods and criteria)

  Still in development so not yet been tested
- Which evaluation results are available so far (including references to where they are documented)

6 Conclusion

- General assessment of usability and functionality of the application. Please make clear what is the basis for the assessment.

  - It could be interesting to measure the usability of the graphical user interface or the wizard and semi-automatic tasks like positioning a plane. In fact all functionalities are very small and it could be a good point to know if people that are not used to computer software could learn it rapidly.

  - Precision inter-user could also be a good way to determine how the software if a good help or not.

  - Precision between normal screen and virtual-reality google during planification procedure could be a more interesting point to evaluate.

7 References

Website : http://www.medicalstudio.org