Building Usable Multimodal and Natural Interactive Systems

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

September 2004

Contents

1 Introduction

This chapter describes the approach of the SIMILAR Special Interest Group (SIG) on usability to research on usability evaluation of multimodal and natural interactive systems.

Before presenting our plan for the chapter, it seems useful to take a closer look at the key terms of the preceding paragraph. *Usability evaluation* forms part of systems and component evaluation more generally. However, usability evaluation is not evaluation of yet another system component. Rather, system usability is affected by the performance of all or most system components as well as of their integration into the system, which is why usability evaluation cannot ignore the results of more technology-oriented evaluation. Usability evaluation also forms part of the remit of the field of human-computer interaction (HCI) which has been an active research area for more than 30 years. However, until much more recently, say, the mid-1990s, HCI research (i) has been dominated by research on GUIs (graphical user interfaces) and (ii) has suffered from the fact that many HCI researchers did not form part of software development teams. Research on interaction and usability of other, non-GUI technologies, such as interactive speech systems, has been carried out more or less independently of the HCI community by developers of interactive speech systems. Research on the usability of natural and multimodal interaction had to await the emergence of the corresponding technologies during the past 5-10 years.

A *natural interactive system* interacts with users by means of one or several of the modalities which humans use to communicate with one another, such as speech, gesture, facial expression, body posture, object manipulation as part of the communication, hand-writing, possibly typing, etc. A *multimodal system* is a system which uses more than a single modality for the exchange of information with the users. Thus, multimodal systems employ several individual, or *unimodal*, modalities, either as input modalities, as output modalities, or both. It follows that natural interactive systems can be made *more* naturally interactive through relevant multimodality and that multimodal systems are not necessarily naturally interactive.

It is worth keeping in mind that the GUI system paradigm is itself multimodal, taking several kinds of haptic input provided by *input devices,* such as keyboard and mouse, and outputting graphics in many different modalities. Yet, arguably, in the past ten years or so, the world of multimodal systems has been augmented with, literally, scores of new input/output modality combinations, reducing past results on the usability of GUI interfaces to a minor fraction of the complexity facing us today. To some extent, the usability of individual systems representing many of those modality combinations has been studied already. However, the sheer complexity of the task demands more systematic approaches than single-system usability evaluation, if, indeed, any more systematic approaches are possible. This question forms the point of departure for the SIMILAR usability SIG which, furthermore, aims to address the question, for a start, at least, on the specific premises of the SIMILAR network, cf. Section 3.

In the following, we briefly outline the state-of-the-art in usability evaluation in multimodal and natural interactive systems (Section 2). Section 3 presents the SIMILAR Usability SIG objectives and the approach taken to address these issues. The approach includes as one of its first steps a common application description structure which is described and exemplified in Section 4. Section 5 discusses next steps in the Usability SIG's work.

2 State of the Art in Brief

Apart from GUI applications which are inherently multimodal, research has for many years concentrated on unimodal systems. In recent years, research systems have been moving towards combining several input and/or output modalities, as in talking heads or embodied conversational agents, in-car applications using spoken dialogue and a small display, games using computer vision input and graphics output, and many others.

This recent trend has generated a need for knowledge of how to evaluate the usability of multimodal systems. In many respects this remains an open research issue. We are not necessarily starting from scratch, though, since it would seem obvious to draw on methods and criteria from usability evaluation of unimodal systems to the extent that they are transferable into the multimodal context.

However, even as regards unimodal systems there is a major gap in our usability evaluation knowledge. This gap concerns what usability actually is and what exactly makes a user like, or accept, a system. We know that there are several factors contributing to user satisfaction but we hardly know them all nor the extent to which each of them contributes. Moreover, the importance of each factor may differ across users and user groups.

When addressing multimodal systems a new main challenge is to find criteria for evaluating the combinatorial contribution to usability and user satisfaction of the input/output modalities involved. For the moment, exploration of how well different modalities work in combination and of their effects on users is often carried out via comparative studies of users interacting with different systems. However, results from such studies are not generalisable to any larger extent which means that for new applications new studies must usually be made. An alternative, theory-based approach is to continue to develop heuristics based on Modality Theory [Bernsen 2002].

In addition to the modality combination problem, there is also the continued proliferation of new system types and the increasing sophistication of systems, whichever their modalities, both of which factors continue to demand new usability evaluation metrics. For example, systems may be operated in mobile environments and not only in a static environment. Other recent system type innovations include systems for education, edutainment, and entertainment. As regards increased sophistication, there are now systems which explore the inclusion of on-line user modelling to provide more flexible and adaptive interaction behaviour. Some systems aim to recognise the user's emotional state and/or to exhibit emotional states of their own, in both cases in order to provide more appropriate and natural system reactions. User preferences and priorities raise new issues in such systems.

A number of handheld and other mobile devices have become available which allow multimodal applications. One example is mobile phones which allow spoken as well as key-pad input, and PDAs which allow pen-based input in addition to spoken and keypad input. Another example is in-car applications. Mobile systems raise several evaluation issues which have not been fully solved, including how (not) to use, and when (not) to use which modalities and which input/output devices, and for which purposes (not) to use location awareness and situation awareness.

Usability evaluation often includes application of the three ISO-recommended (International Standardization Organisation, www.iso.org) usability parameters, i.e., effectiveness, often measured as task success rate, efficiency, often measured as time to task completion, and user satisfaction, often evaluated based on a questionnaire, cf. ISO 9241-11. Even as regards these basic approaches, however, it may be noted that some of the new system types are not task-oriented at all, such as entertainment systems. For such systems, arguably, considerations of effectiveness and efficiency – at least in the traditional sense - are simply irrelevant whereas new usability criteria, such as some new form of interaction success, entertainment quality, and interaction naturalness, push themselves to the forefront when evaluating usability.

On-line user modelling is receiving increasing attention for several reasons. Users of mobile devices which are usually personal belongings, may benefit from functionality which builds knowledge of the individual user. Generic user modelling may also be useful. For instance, novice users could receive more extensive interaction guidance and users who repeatedly make particular types of error could be helped by explicit advice or by adaptation of the interaction structure. Some key evaluation questions regarding on-line user modelling concern: (i) if the user modelling functionality is technically feasible in the first place and (ii) whether it will be of benefit rather than a nuisance to the majority of users of the application. For instance, even if the system has enough information on an individual user, adaptation may fail because of too primitive update algorithms or insufficient information to the user about when the user model has been applied.

Not only recognition of users' emotional states but also systems' expression of emotion is an active research area. Usability evaluation must consider which positive and negative impacts emotion modelling has on users.

User preferences can make life hard for the developer as they may contradict what is empirically the most efficient solution. Some users may, e.g., prefer pen-based input to spoken input or keypad-based input to spoken input, simply because they feel more familiar with GUI-style interfaces. Depending on the target user group(s), alternative modalities may be needed because it is likely that each of them will be preferred by some significant fraction of the user population. This is just one reason why user involvement from early on in the development process is recommended and why on-line user modelling appears attractive. Some preferences we can design for, such as modality preferences. Others, however, are hard to cope with. Thus, some users may prioritise speed or economical benefit, while others prioritise human contact. The question is whether we can build systems with a usability profile that will make the latter users change their priorities, and exactly which usability issues must be resolved to do so.

In brief, there seems to be a broad need for usability evaluation that can help us find out how users perceive new kinds of multimodal and natural interactive systems and how well users perform with them, possibly compared to other types of system. There is a strong wish in the field to find ways in which usability and user satisfaction might be correlated with technical aspects in order for the former to be derived from the latter. We do not have methods today that can reliably predict how well users will receive a particular system. We just know that a technically optimal system is not enough to produce user satisfaction. Regarding modality appropriateness which is a central issue in multimodal SDSs, modality theory may be a promising and powerful approach to usability evaluation of modalities at an early stage. However, user tests of the actual design will still be needed, as for unimodal systems. For an overview of usability evaluation of multimodal systems, in particular systems involving speech, see [Dybkjær et al. 2004]

3 The SIMILAR Usability SIG Approach

The SIMILAR Network has particular properties which should be taken into account by the Usability SIG. Thus, (i) SIMILAR has its focus on natural interaction rather than on multimodal interfaces more generally ("... taskforce creating human-machine interfaces SIMILAR to human-human communication"). Yet, (ii) SIMILAR includes research issues, such as (non-invasive) brain process interpretation, which clearly seem to go beyond human-human communication. To the Usability SIG, the SIMILAR community is already granting access to a selection of innovative multimodal natural interactive technologies, limited only, to some extent, by (iii) the network's focus on three particular application areas, i.e., medical, disabled, and edutainment. Finally, (iv) SIMILAR has a preponderance of signal processing researchers, which poses interesting demands on the Usability SIG because, from a traditional point of view, signal processing does not form part of software engineering at all but, rather, addresses a large class of basic technologies for potential take-up by software engineers and system developers. From a usability point of view, important questions may arise in this context about how to evaluate the usability of new and promising signal processing algorithms which, as is often the case, are "looking for applications" rather than being applications by themselves. This is not usability evaluation in any standard sense of the term.

3.1 General objectives in brief

At the time of writing we are only about half a year into the SIMILAR project. The objectives for the SIMILAR Usability SIG for the first 18 months are to

- establish the group of SIMILAR members who will contribute to the Usability SIG;
- review relevant literature on usability evaluation of natural interactive and multimodal applications;
- establish a pool of accessible natural interactive and multimodal applications (exemplars) developed within SIMILAR, which can be analysed in depth from the point of view of usability evaluation;
- create a template-like analytic structure for obtaining a detailed description of how each application has been evaluated regarding usability;
- use the structure to describe current practice for the available pool of applications, possibly adding, resources permitting, the results of new usability evaluation exercises with respect to selected systems;
- based on the above, develop a first outline of a best practice framework and guidelines for the evaluation of system and component usability in the field of natural interactivity and multimodality.

The best practice outline will necessarily be a very preliminary outline given the effort and time available during the first 18 months as well as the size and complexity of the field. Thus, the main focus during the last 30 months of SIMILAR will be to iteratively consolidate and enlarge the coverage of the best practice framework. Scope consolidation and enlargement will partly be achieved through analysis of the issues arising and partly through analysis of additional exemplars from SIMILAR partners and colleagues outside SIMILAR.

In Section 4 we present the exemplar description structure and exemplify its use. Prior to that, we would like to discuss in more general terms our approach to outlining current practice (Section 3.2).

3.2 A bottom-up approach

The approach we have decided to use in the SIMILAR Usability SIG is inspired from, and partially builds on, the DISC project (Spoken Language Dialogue Systems and Components: Best practice in development and evaluation, www.disc2.dk) in which academic and industrial partners investigated current and best practice in the development and evaluation of spoken dialogue systems and their components. The DISC current practice approach was to (a) analyse a broad range of spoken dialogue systems and components and (b) map out their respective development and evaluation processes. In order to adequately capture current practice and overcome various problems primarily relating to the insufficient and not-easily-comparable information provided for individual systems and components, a common scheme was developed. This scheme was applied in the analysis of about 25 exemplars, i.e. systems and components to which access was provided by the project partners. Each exemplar was analysed independently by two different project partners, yielding 50 internal and confidential reports. For each component level and the system level a synthesis description was made based on the relevant exemplar descriptions. Each synthesis description abstracted from individual component- or system-specific observations and presented the range of practical approaches followed in the development and evaluation of systems or components. Based on normative analysis of the current practice descriptions a draft best practice was then established.

By contrast to the SIMILAR Usability SIG, and apart from a single multimodal system present in the DISC exemplar pool, DISC focused on task-oriented unimodal spoken dialogue systems, and DISC not only considered – technical as well as usability - evaluation but also the entire development process and which features to include when building a spoken dialogue system. In the Usability SIG we will solely look at usability evaluation which is a simplification compared to DISC. However, the range of systems to be included will be much larger, more varied, and more complex than in DISC.

Like in DISC we have established a first pool of systems gracefully made available for usability evaluation research by the SIG participants. The exemplar pool includes a total of seven applications, including two entertainment/edutainment systems, three surgery/operation room systems, a museum application, and a training system for the blind. An in-car application may be added later. Individually, the systems are very different and, collectively, their properties go far beyond those addressed in the DISC scheme. To solve this problem, we are drawing on the approach used in the MATE (Multilevel Annotation Tools Engineering, mate.nis.sdu.dk) and ISLE (International Standards for Language Engineering, isle.nis.sdu.dk) projects to collect information about, and subsequently describe, a wide variety of different natural interactivity data resources, annotation schemes, and annotation tools. In both projects, common description structures were developed and applied to the collected information about each data resource, annotation scheme, and annotation tool. This turned out to work rather well. We have therefore adopted a similar approach in the Usability SIG, establishing a common description structure which will be described in more detail in Section 4. Once we have collected the corresponding information about all SIMILAR exemplars, we will look into the problem of establishing and refining a scheme for capturing current practice descriptions of usability issues in the systems.

4 Application Description Structure

In the following, we introduce (Section 4.1) and exemplify (Section 4.2) the common structure used for exemplar descriptions in the SIMILAR Usability SIG.

4.1 Introduction

The purposes of having a common application description structure are to ensure (i) presentation, at a common level of detail, of all applications to be analysed and evaluated in the SIMILAR Usability SIG, and (ii) a minimum of information on each application, subject to additional information gathering when required for usability evaluation purposes.

The precise entries of the common application description structure are shown in Section 4.2. It is quite possible that some of the entries can be filled in only tentatively, if at all, for a particular exemplar, such as when the system is still under development and usability evaluation is still ongoing.

Reports may be classified confidential if a contributor so wishes. This means that their contents will only become public at the higher, more abstract level at which all references to the properties of particular systems will be removed. Report contributors will of course have the opportunity to check, prior to publication, that confidential information has been removed at that higher level of presentation.

4.2 Application description

In the following we exemplify the entries in our application description structure. The examples draw on Usability SIG work, i.e. [Bernsen and Dybkjær 2004] for the NICE Hans Christian Andersen (HCA) example, [Trevisan et al. 2004] for the Image Guided Neurosurgery example, [Nikolakis et al. 2004] for the system for blind people, [Hernandez and Marichal 2004] for the AlterStation game system, and [Berti et al. 2004] for the Portable Cicero museum application.

4.2.1 Purpose of the application

Insert the goals of the development project. These will typically be to demonstrate, or deliver to customers, certain functionalities, often adding usability descriptors and aims regarding the users appreciation of the application. Other goal parameters may include price and quality descriptors, target users, use settings, etc., informally described and further detailed below.

Example: 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.

Example: 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.

4.2.2 Input modalities

Describe the way(s) in which the user inputs information to the system, either using the terminology of Modality Theory [Bernsen 2002] or using informal descriptions, such as "GUI-style input augmented with ...", possibly referring to input *devices* rather than modalities, or referring to both modalities and devices. For instance, we all understand what standard mouse input is even if we do not realise that the mouse is a simple haptic input code device.

Example: spontaneous English speech and 2D gesture via mouse or touch screen (HCA system).

Example: pen and infrared (museum application).

4.2.3 Output modalities

Describe the way(s) in which the system outputs information to the user, cf. Section 4.2.2.

Example: 3D animated, life-like embodied HCA communicates with the user through English conversational speech, gesture, facial expression, body movement, and action (HCA system).

Example: speech, video, map, text (museum application).

4.2.4 Target user group(s)

Arguably, no interactive application can be meaningfully targeted at all users, just think of different user properties, such as the language(s) they speak, their culture, their age, their educational and professional background, their interests, their mastery of the human senses, their mastery of their body, etc. It follows that any application must be targeted at (a) specific user group(s).

Example: target users are 10-18 years old children and teenagers (HCA system). Basically, the system requires no training in order to start using it.

Example: surgeons (Image Guided Neurosurgery).

4.2.5 Physical use environment

Despite application classifiers like "ubiquitous computing" and "ambient intelligence", most applications are aimed to be used in particular environments, or across them.

Example: the primary use setting of the HCA system is in museums and other public locations.

Example: Image Guided Neurosurgery is meant for use in operation rooms.

4.2.6 Which domain does the application cover

Enter what the application "is about", i.e. the general domains of information, action, or otherwise, which the application addresses.

Example: the general domain of the HCA system is a combination of education and entertainment. More specifically, the system allows users to have conversation with HCA about his life and fairytales, himself and his study, as well as about the user, games and technical inventions.

Example: Image Guided Neurosurgery is within the medical domain.

4.2.7 Which tasks (if any) does the application solve

A task is a far more specific entity than a domain. Most applications are aimed at enabling the user to do a, or some, more or less specific task(s). Some applications, however, are not task-oriented at all in any clear sense of this term. The HCA system is an example of the latter.

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

Example: 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 (application for the blind).

4.2.8 Is the application free or what is the price

This question is obviously important to anyone interested in the application. If the application is "free", the user might want to know if this means that it is open source or if a free executable is available. Moreover, the user will want to know how to get hold of the system.

Example: this research prototype application is not free nor does it have a price. If someone wishes us to, e.g., port the application to a different language or even replace HCA with a different character, please contact us and we will estimate the cost (HCA system).

Example: Image Guided Neurosurgery is a proprietary system. No price information is available.

4.2.9 If not free, is a demo available?

Among other things, demos include short-time test licenses, reduced-capability system versions, simple demos, such as sound or video recordings of human-system interactions, etc.

Example: a small demo video is available at http://www.niceproject.com/about/ (HCA system).

Example: the Portable Cicero museum application 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

4.3 Technical issues

4.3.1 Platform(s) (operating system(s))

This is crucial information on any system. In many cases, the information should be supplemented by information on specific non-standard software needed to run the application, including APIs. For research prototypes in particular, it is important to describe which platform compatibilities have been actually tested. This also applies to "platform-independent" software. Any other "exotic" information needed to run the software should be listed here as well, such as specific platform settings required which may not be intuitively obvious.

Example: the 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.

Example: The training system for blind people runs on a Windows NT, 2000 and XP platforms.

4.3.2 Hardware requirements

In principle, this information should be provided for any application, however small and limited in its requirements. Potential users of the application should receive sufficient help rather than having to do under-informed guesswork.

Example: running the system requires a powerful computer with 500-1000 Mb RAM and a good graphics card, such as G-Force 4 (HCA system).

Example: In order to run the training system for the blind one 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.

4.3.3 Implementation language(s)

This information backs up the potentially complex information provided in Sections 4.3.1 and 4.3.2. In the case of highly complex systems, shortcuts are permitted.

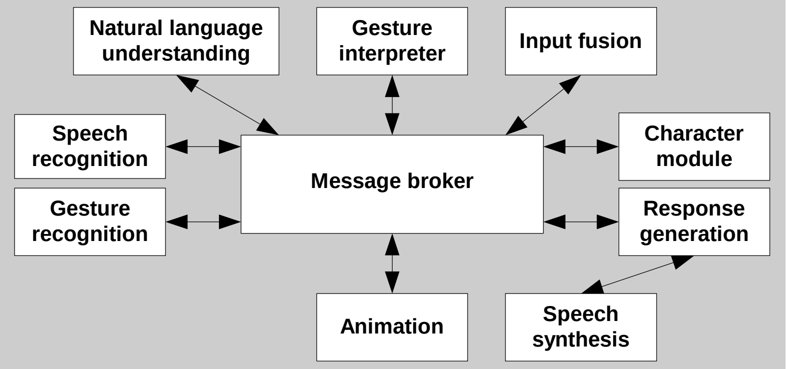
Example: the implementation languages used in the HCA system are mostly Java, C++ and Sicstus Prolog. The HCA system is an example of a complex system for which the need to provide detailed module-per-module programming language information is not obvious.

Example: 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.

4.3.4 Architecture

High-level knowledge of the system architecture is important even in the case of usability evaluation. This knowledge helps understand possible usability shortcomings and supports the asking of additional questions when such shortcomings have been discovered. Preferably, the architecture should be presented in terms of an annotated high-level architecture diagram. The annotation would describe the individual modules, their origin, and the overall information flow.

Example: The diagram below shows the overall architecture for the HCA system. Explanations of components are not included in this brief example.



4.4 Functionality

4.4.1 Which functionality does the application offer

System functionality is closely related to system purpose and system task(s). System evaluators are likely to address the described system functionality by asking if, e.g., the functionality is adequate to the system's purpose and to the tasks supported by the system. If, as is often the case in research prototypes, the existing system functionality is deemed to fall short of the desired functionality, it is important to describe the functionality which is deemed missing, saving the usability evaluators the effort of pointing this out.

The properties of system functionality ("what the system can do for you") and usability ("how usable the system's functionality is") are analytically distinct but closely related. A functionally adequate system may be partly or wholly impossible to use due to its inadequate user interface. Conversely, a functionally inadequate system may become popular among users due to its intuitive user interface. Listing the system's functionality is prerequisite to enabling the usability evaluator to judge if the functionality is adequate for the system's task(s), or otherwise, and whether the functionality which is there is actually usable. An inadequate user interface can effectively hide much useful functionality, excepting only the users who study the manual carefully or who receive substantial training.

Example: 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.

Example: The Portable Cicero museum application 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.

4.4.2 Description of each main functionality

For complex systems, it is sufficient, at this point, to list the major system functionalities, referring to any supporting documentation for more detail.

Example: The 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.

Example: 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.

4.5 Interface and usability

4.5.1 Description of interface design and possible design for usability

Given the system functionality descriptions in Sections 4.4.1-2, this section describes the usability aims, heuristics, guidelines, standards, underlying theory, or otherwise, or the lack of them, which were adopted in designing the user interface. The tentative or consolidated nature of the user interface should be described as well.

Example: (from the AlterStation system) 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..

Example: 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 (museum application, abbreviated description).

4.5.2 Which user skills (if any) are assumed

This entry should mention any in-context, non-trivial skill requirements which the user is assumed to satisfy. At this point in the application description, we already have plenty of context. Thus, e.g., if the application is a statistics package, the task is to use the package for statistics purposes and the standard assumption will be that the user is familiar with principles of statistics. However, if the statistics package does not assume user knowledge of statistics, this would be contextually non-trivial information. If no particular skills are required, this should be stated as well.

Example: very good knowledge of the system is required (Image Guided Neurosurgery). Moreover, the user must be educated as a surgeon.

Example: The AlterStation 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.

4.5.3 Is it walk-up-and-use? Is training foreseen? Is there a manual?

A walk-up-and-use system is a system designed for ordinary users who should be able to use the system without any training or manual consultation. It is important to describe any non-trivial learning (or training) requirements imposed by the application. If there is a user manual, information on how to access the manual should be provided.

Example: the first HCA system prototype only requires that the user knows how to change camera angles using a function key, how to control HCA's locomotion using the arrow keys, and how to use a touch screen, if available (otherwise, the mouse may be used). There is no manual for the system.

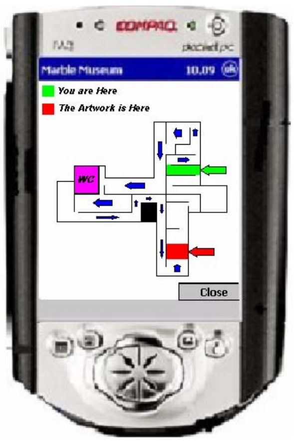
Example: no particular skills or experience are required. There is no manual (museum application).

4.5.4 Illustrative examples of use and interface

User interface illustrations are important for early inspection of the look-and-feel of the system and prior to trying out the system first-hand. Illustrations especially concern static or dynamic graphical user interfaces, haptic input or output devices, and the like. For spoken/acoustic interfaces or interface parts, transcriptions of example interactions, or sound files, are helpful.

Example: use of the HCA system in the HCA museum in Odense Denmark.

Example: PDA showing museum map (museum application).



4.5.5 Advantages and disadvantages of functionality and interface

This entry should provide a global assessment of the system's functionality and interface, possibly referring to the functionality descriptions in Sections 4.4.1-2. The assessment may only reflect the developers' anticipations but may also reflect general lessons learned from user tests performed with the system.

Example: The first HCA system prototype of January 2004 is still incomplete and primitive in several basic respects: speech recognition is not integrated yet, graphics rendering is primitive wrt. number and timing of non-verbal behaviour primitives, the spoken conversation is too inflexible, the speech synthesis must be improved. The quality of graphics rendering is good.

4.6 Evaluation

4.6.1 Who/how many have used the application so far

For products, this question may be unanswerable, of course. For research prototypes, on the other hand, the information provided is likely to be important. For instance, if only the developers have used the application, this is important information. Also, for research prototypes, the test users who have worked with a more or less simulated system version should be described together with the nature and setup of the simulation.

Example: Eighteen kids and teenagers have used the first HCA prototype in a controlled lab test.

Example: Thirty five museum visitors have used the system and subsequently filled in a questionnaire (museum application).

4.6.2 How has the application been usability tested (detailed description of methods and criteria)

If one or several user tests have been made with the system, it is important to describe the test protocol and evaluation criteria applied, possibly referring to additional accessible information.

Example: Museum visitors were given a PDA with the museum application installed and after their visit to the museum they filled in a questionnaire. 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, and capacity to help users orient themselves in the museum.

4.6.3 Which evaluation results are available so far (including references to where they are documented)

This section should describe any evaluation results obtained. If the results are complex and detailed, a global description is sufficient together with references to additional accessible information. If future user tests are being planned, this should be explained.

Example: 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 2004]

4.7 Conclusion

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

This entry should provide general lessons learned so far, if any, on the system's usability. If additional usability evaluation of the system is required, this should be stated, preferably with details on the aspects in need of further investigation. It is useful to also mention ongoing work on analysing user test data which has been gathered already.

Example: Basically we are interested in measuring 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/augmented system (with alternative interactions) is being developed (Image Guided Neurosurgery).

Example: 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.

4.8 References

4.8.1 Any references which provide more information about the application

It is useful to not only provide references but also explain, for each reference, the information is contributes. For confidential references, please state what may be disclosed, and possibly in which way.

Example:

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

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

5 Next Steps

We have almost finished writing the system descriptions following the common structure presented in Section 4. The next step will concentrate on the evaluation description given per application to see how far we can get in extracting and synthesizing a current practice scheme for usability evaluation. This is likely to be an iterative process requiring more detailed information about evaluation criteria and methods used than what has been made available in the descriptions. It may also turn out that we need to include additional applications to consolidate the scheme. When we have a reasonable current practice description we will try on that basis to distil a draft best practice framework.

6 References

Bernsen, N. O.: Multimodality in Language and Speech Systems - from Theory to Design Support Tool. In Granström, B., House, D., and Karlsson, I. (Eds.): Multimodality in Language and Speech Systems, Kluwer Academic Publishers, Dordrecht, 2002, 93-148.

Bernsen, N. O. and Dybkjær, L.: Description of the NICE Hans Christian Andersen System. NISLab, University of Southern Denmark, SIMILAR Report D17, 2004.

Berti, S., Paternò, F. and Santoro, C.: Description of the Portable Cicero Application. HIIS Laboratory, ISTI – CNR, SIMILAR Report D17, 2004.

Dybkjær, L., Bernsen, N. O. and Minker, W.: Evaluation and Usability of Multimodal Spoken Language Dialogue. Speech Communication, Vol. 43/1-2, Elsevier, 2004, 33-54.

Hernandez, P. C. and Marichal, X.: Description of the AlterStationTM System. Tele, Université Catholique de Louvain, Alterface SA, SIMILAR Report D17, 2004.

Nikolakis, G., Tzovaras D. and Strintzis, M. G.: Description of a Training System for Blind People. ITI-CERTH, SIMILAR Report D17, 2004.

Trevisan, D., Gemo, M., Nicolas, V., Vanderdonckt, J. and Macq, B: Description of an Image Guided Application Using Contextual Focus Driven Interaction. UCL-TELE and BCHI, Université Catholique de Louvain, SIMILAR Report D17, 2004.

DISC project: www.disc2.dk

ISLE project: isle.nis.sdu.dk

ISO: www.iso.org

MATE project: mate.nis.sdu.dk