KUNGL
TEKNISKA
HÖGSKOLAN

TRITA-NA-E01102 ● CID-184 ● ISSN 1403-0721 ● Department of Numerical Analysis and Computer Science

Multimodal Interface for mobile clients

Niklas Becker

CID, CENTRE FOR USER ORIENTED IT DESIGN

Niklas Becker

Multimodal Interface for mobile clients

Report number: TRITA-NA-E01102, CID-184
ISSN number: ISSN 1403 - 0721 (print) 1403 - 073 X (Web/PDF)
Publication date: Dec 2001

Reports can be ordered from:

CID, Centre for User Oriented IT Design
NADA, Deptartment of Numerical Analysis and Computer Science
KTH (Royal Institute of Technology)
SE-100 44 Stockhom, Sweden
Telephone: + 46 (0) 8 790 91 00
Fax: + 46 (0) 8 790 90 99
E-mail: cid@nada.kth.se
URL: http://cid.nada.kth.se

SchlumbergerSema

Multimodal Interface

For mobile clients

Multimodalt gränssnitt

För mobil klient

Master Thesis by

Niklas Becker

2001-12-05

Människa dator interaction (MDI)

Handledare KTH: Anders Hedman (CID)

Examinator: Yngve Sundblad

Uppdragivare: SchlumbergerSema / infodata, Göran Derefeldt, Mikael Fredriksson

Abstract

Multimodal interfaces – for mobile clients

With the development of wireless gadgets like the personal digital assistants (PDA) and the cellular phones, the issues regarding whether there is enough computer power become more distant. What killer application can be implemented on top of the hardware? An application that might be useful (and of course worth the effort to develop) is an application with support for switching between different modalities in different environments/scenarios. To meet the requirements of such an application, the underlying application and the user interface need to be carefully designed for multimodal use. There are many levels of multimodal interaction. Designing for how and where the user can switch modality is complicated (multimodal applications). In general, the user should be able to switch modality at any time (but there are important exceptions). This might be easy to implement with the release of General Packet Radio Service (GPRS) technology, since multiple network sessions are allowed. But, if several sessions can be held simultaneously, does the mobile device support simultaneous sessions? How can text and speech work together to construct more usable and flexible user interfaces? The report gives an insight in multimodal interaction and how switching of modalities should be implemented.

Referat

Multimodala gränssnitt – för mobila klienter

Med utvecklingen av mobila klienter som PDA eller mobiltelefoner, blir det mindre intressant om det finns tillräckligt med processorkraft i klienterna. Vilken typ av applikation kan implementeras på mobiltelefonsystemet? En typ av applikation som kan vara användbar är ett multimodalt gränssnitt med stöd för byte av modalitet i olika miljöer/scenarios. För att implementera en multimodal tjänst måste programmeraren ta hänsyn till underliggande applikation och användargränssnitt. Det finns flera olika aspekter på hur interaktion mellan modaliteterna ska vara och när användaren ska kunna byta modalitet. En generell definition är att användaren ska kunna byta modalitet närsomhelst (med vissa undantag). General Packet Radio Service (GPRS) kan underlätta implementationen av detta, eftersom multipla nätverk sessioner kan öppnas. Även om nätverket stödjer simultana sessioner, stödjer den mobila klienten detta? Hur kan text och röst komplettera varandra för att öka användbarheten av gränssnittet och flexibiliteten för användaren? Den här rapporten ger en inblick i multimodal interaktion och hur byte av modalitet ska implementeras.

Executive summary

Multimodal interfaces enable interaction through different communication channels. Modalities are channels humans can use to interact with people or machines. Examples of such channels are: visual, audio, tactile (feel), smell, taste and proprioception (orientation of body). This paper concerns in the integration of two modalities (bimodal interface), visual and audio. The interface is the window to the software application and this interface differs depending on modality.

Integrating several modalities into one single interface increase the complexity of the application. It is important for the designer to understand how to implement and design the interface. Moreover, human-computer interaction tends to become complex as new modalities are added. Each modality has its own interface and specific human behavior. How should the modalities be designed to work together in a product, which is easy to use?

Multimodal interfaces is not a new feature for home computers or mobile devices, multi channel communication is an “old” technology. The use of keyboard, mouse, and audio feedback in the home computer is a primitive form of multimodal system. The new concept is that all tactile interactions are excluded, audio and visual channels are the main modalities.

Mobile devices do not allow simultaneous modalities to be used. However, sequential use of modalities may be used. This implies that the user can only use one modality at a time, for example using voice and at the same time push buttons is *not* allowed. Further development of mobile devices will change this limitation. Without a GPRS - enabled mobile device there are also some network constraints. Especially using WAP over GSM, the user has to choose whether using WAP or make a voice call. This has to do with that ordinary GSM technology does not allow simultaneous data and voice call.

There are two constraints limiting the multimodal interaction. GPRS eliminate the network limit, but does not allow multiple communication channels. Identifying the limits of an implementation can be done, a user study can be conducted and a multimodal interface can be developed.

The results from the user study pointed to a content problem with mobile interaction. There are too few services, especially on top of WAP, to attract new users. Whether this is hardware or content problem cannot easily be defined, although content is missing not the hardware. Infodata has content and why not extend it to mobile devices and multimodal interfaces? The development cost might be higher, but the user rating will be higher (for further information see chapter 7, conclusions). If the multimodal interface is not satisfactory or to complex, it can still be used as an unimodal interface!

Acknowledgements

The Master Thesis was completed November 16, 2001 at Infodata / SchlumbergerSema in collaboration with CID (Centre for User Oriented IT – Design, Nada, A. Hedman). I would like to thank especially the people at Infodata for giving me the opportunity to learn Speech technology and new technologies as an aid for user interaction and user experience. I realized that if new technology is introduced, further development is a necessity. Not only does new technology demand more learning from the user, but also the developer must realize that systems integrating different technologies is complex. I also learned and experienced that software and hardware developer's problems are intimately related.

The hardware must support the software and the software support the hardware, all too often this lead to a deadlock, catch22.

SchlumbergerSema, 2001

Niklas Becker

1 Background

With development of new technology in communication systems, voice enabled applications (IVR – applications) are being developed. At the present time most IVR applications are based on traditional telephony (GSM systems for mobile devices). With GPRS technology, IP based communication can be used allowing constant access to the Internet. Several IVR applications test using SIP on the Internet with GPRS and this technology can be expanded to mobile communications systems. SIP is not used in this project. Voice over IP is used with GPRS and 3G – enabled IVR applications when mobile devices get support for Java2ME and high level program languages. Most devices does not fully support software development, this will change with further development with devices such as Compac iPAQ, HP Jornada and Nokia Communicator.

Although new technology has been introduced, what benefits can be made? Is the new technology usable? If it is, study the new technology and draw conclusion whether it is worth money and effort to develop. Not only does a human - computer interaction (HCI) designer have to focus on user interaction with the system, but assess if the technology is worth research and development. Two different aspects of HCI can be stated, usability for user interaction and usability as a technology. WAP is a good example of technology that has not yet been successful. Even though, WAP is the only current standard for presenting content on mobile clients, it is rarely used. Would WAP have suffered the same setback with more modalities[1]? With multimodality, WAP might get a new chance.

Multimodal interaction is different technologies (modalities) merged together to form an easier to use and understand user interface[2]. Combining these modalities, user interaction can be faster and easier, but also slower and more complex. It depends on the designer and his / her grasp of human – computer interaction involved.

1.1 Usability and HCI

Voice is the perfect modality for users who cannot use keypads or similar devices, it is perfect! Users with muscle and eye impairment will gain better access to mobile application and services. Using two modalities or more, a larger audience will be targeted. Preferably even more modalities should be used, why is this not the case?

Instead of learning how to use the interface, the user can try to have a natural language dialogue and the application will be easier to use.

[1]Modalities are different types of input/output media between humans, machines or other actors. Different modalities can include gestures, voice, keyboards, mice and much more.

[2] The interface is the window towards the application, the application's representation.

If the user is an expert and does not want to use natural dialogue the user can switch modality. With gestures and tactile[3] communication users who are both blind and mute can still use the interface.

However, the user does not have to be physically disabled to enjoy the benefits of multimodal interfaces. Some situations it could be much easier to use voice interaction as a complement to a dominating modality or vice versa.

Examples of these situations could be when the user moves between different places, where background noise (S/N ratio is bad) is too loud or when driving a car etc. When does the user want to switch modality? How should the switching between modalities be handled in the interface? Which modality should be used to present the context? The user should decide on which modality to use.

1.2 Using current technologies

A problem with analogue telephony is its inability to use TCP/IP. To implement a multimodal application, the infrastructure has to support voice and data transfer. This is important since the WAP protocol does not support multi - channel communication in non - IP mediums neither does the mobile devices. Next generation mobile devices allow opening of several sessions simultaneously. With GPRS the PDA/Mobile phone can maintain a TCP/IP session open, without having to reconnect. With GPRS the WAP browser does not have to drop the WAP session, when initiating call functions. When not using GPRS the user has to choose whether to use WAP or make a call (switch modality). If IP communication could be used, SIP session could be initiated.

Figure 1. Ericsson wireless PDA / mobile phone (Courtesy of Ericsson)

GPRS is a network technology based on IP communication. It is with the development of GPRS technologies and 3G, multimodal inter-action becomes available for common users (see figure 1).

[3] Tactile is communication through feel (vibrations, rotation, etc.).

The hardware requirements of multimodal interaction are both on network (Bandwidth, simultaneous sessions) and client (CPU, software compatibility, simultaneous sessions) technology.

PUSH technology enables the server to initiate data streams to the mobile client. Without PUSH, the user has to manually retrieve information from the server. Implementing this technology allows the server to update information, open sessions and stream data. PUSH is not available yet.

The market contains interesting technologies for both client and network, but software and hardware developers have an intimate relationship. Often software developers wait for new hardware technology and hardware developers wait for content. Commonly users feel that content is missing, not the hardware. Content is not the only reason, prize, bandwidth and usability have different impact on the user. Who will be first with a killer application using WAP or equal on GPRS?

The project goal is to implement a prototype with a multimodal interface on a mobile phone. The interface should be on WAP over GPRS and be tested and evaluated in a study.

Apart from the main goal, several sub-goals can be derived. The sub-goals that should be fulfilled during the project are:

1. Study of NUANCE IVR technology, tools and high-level programming languages.
2. Define problems with the implementation and services. The interface is implemented on top of WAP with Java (Beans & Servlets), JSP, VoiceXML and WML.
3. Define HCI aspects of the interface. What are the problems with multi-modal application design? How should voice and text be combined? The modalities used in this project will be voice and text.
4. Test and user studies. It is important to do at least one user study before the end of the project. From the results of the study, several usability faults in the interface can be detected. The interface should be evaluated before the test according to HCI theories and methods.

1.3 Reading instructions

This is a list, which gives a brief overview of the chapters in this report.

1. Chapter 1 gives an introduction to the thesis and multimodal interaction
2. Chapter 2 reviews the requirements on multimodal interaction of HCI. There are several human aspects that should be considered. Multimodality is defined and how a voice and text as modalities should be developed for good user interaction. It is important to understand why and when a user wants to switch modality. The HCI evaluation methods that are used are reviewed.
3. Chapter 3 defines the problems with multimodal interaction and the reader should read this chapter to understand the need for this project and get answers from this report.

4. Chapter 4 shows how a multimodal system could be implemented and how the prototype has been implemented. Some examples on how the interaction could be done are also shown. Mostly this chapter shows technical and programming implementations done at Infodata.
5. Chapter 5 defined how the user study has been conducted and problems that the users might encounter. This is a small chapter and the reader should se chapter 6 and 7 for further information regarding the user study.
6. Chapter 6 gives information on if the user study was successful and shows the results from the questionnaires. Users comments are also posted.
7. Chapter 7 is a conclusion considering the results from the user study and what was said during the tests. Future work that needs to be done to a next version of the prototype has been summarized. Guidelines for multimodal design for future use has also been summarized in a 10 - point list.
8. Chapter 8 reviews the references.
9. Appendix A is a glossary with abbreviations.
10. Appendix B shows screen dumps from the prototype on a emulator for a Ericsson 380 mobile phone. All different scenarios are included in this appendix.
11. Appendix C gives some information regarding companies that are currently researching in the area of multimodal interaction. There are several more companies, but the most important are stated in this appendix.
12. Attachments A and B show the questionnaires from the user study.

2 Theories and methods of HCI

When designing user interfaces in HCI a specific approach solving usability problems should be used. HCI method of usable interface design mostly involves guidelines regarding user behavior and training. As these guidelines often are of general approach it is important to complete the design process with a study of each specific implementation.

By deploying guidelines tested by HCI researchers, designed interfaces can exhibit good usability from start. These guidelines show problems with the design and can be used for future design. There are guidelines on how to conduct qualitative user studies and how to test what is intended by the experimenter to test. It is not always that the study is successful in measuring what was intended.

Common guidelines used by HCI researchers that will be used are heuristic evaluation lists, GOMS, user study. A Heuristic evaluation list can be used to test whether the application fulfils basic usability according to a 10-point list (J. Nielsen, ´93). There are several other methods to describe human - computer interaction (for further information see keystroke - level model, layered model, 3 - level model, cognitive walkthrough).

The methods give the designer information on user behavior and how to analyze interfaces. The GOMS (Goal, Operators, Method and Selection Rules proposed by Card, Moran and Newell ´83) method is a well-used method in HCI design. The analyst performing the test of the system, does a walkthrough of the system with the user (or monitor the user). Every single action can be described in simple task. Goal is what the user should retrieve from the system, operators are modalities, method is how and selection rules are a set of rules, which the user can use to perform the task (and retrieve the goal). Problems with GOMS are that often the goal of the user is difficult to analyze and the GOMS method can tend to be detailed. GOMS will not be used, but the method will be kept in mind for the user study. So what is multimodality and how does it relate to the user?

2.1 Multimodality and intentionality

The problem of human computer interaction is the lack of intention - and goal oriented concept on the computer side. The user has an intention and a goal with the interaction whereas the computer is focused on the present task and does not understand what the intention or goal of the task is. This is a common problem when designing usable interfaces. In the end it is the user and designer who communicate by the interface.

Multimodality is when user and computer are physically separated, but are able to exchange information through a number of information channels.

According to L.Shomaker et. al. ´95, H. J. Charwat, ´92, the term multi-modality can be defined as:

“ *Perception* via one of the three *perception - channels*. You can distinguish the three modalities: *visual*, *auditive*, and *tactile* (physiology of senses).”

However, only three types of modalities are presented in the quotation. Other types of modalities like smell, taste, and balance might be included (see figure 2). When two or more modalities are in use, we talk of a multimodal system. Using two modalities the system is defined as a bimodal system. The interface is how the output is presented to the user. It can be text on a screen, displayed graphics, voice prompts, etc.

Sensory perception	Sense organ	Modality
Sight	Eyes	Visual
Hearing	Ears	Auditive
Touch	Skin	Tactile
Smell	Nose	Olfactory
Taste	Tongue	Gustatory
Balance	Organ of equilibrium	Vestibular
Body orientation	Joints, nerve system	Proprioception

Figure 2. Type of modalities for humans. (Courtsey of Silbernagel ´79, extended by Becker ´01)

Each modality has its own interface, therefore a more multimodal interface is far more complex. Combining all modalities is therefore an integration of several interfaces. Multimodal interface can involve a combination and synchronization of interfaces in a single interface! Synchronization is required since the user moves through different states and all interfaces need to know which state, otherwise mismatch occurs (also known as state errors). This project will involve a bimodal interface using voice and text.

How does perception for a modality change when introducing multimodality? Apart from the unimodal constraints, an interesting phenomenon the level is the improved perception for a given modality under multimodal conditions (L. Shomaker et. al., ´95). This is true, since development of multimodal interfaces develop each modality in a Physical/Physiological observation. The user develops a synergy between the modalities (see figure 3, next page). The user interacts with the application from different kind of views. This would be valuable for Interactive Voice Recognition (IVR) interfaces and speech dialogues.

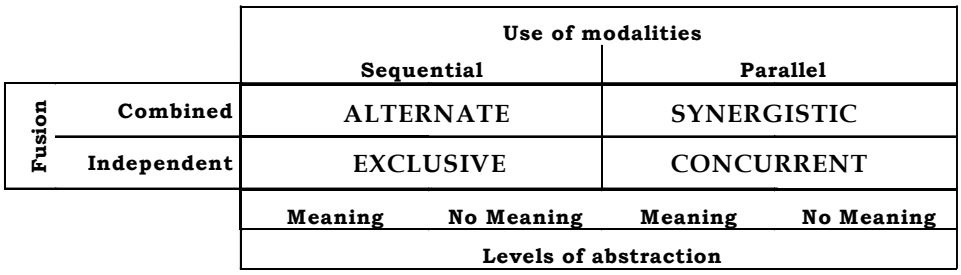

Figure 3. Different types of multimodal interfaces. (Courtesy of Nigay and Coutaz, ´93)

2.2 Speech dialogue

Interactive Voice Recognition (IVR) design an HCI oriented task since its main focus is on exactly human - computer interaction. Speech applications rely on ability to adopt, since the computer can never communicate with the user in "full" natural spoken language. The ideal case would be where the user can speak with the computer without any recognition errors. Speech dialogue was originally designed for human – human interaction, which is much more complex than human computer interaction at the moment (if interested see Bruce Balantine et. al., ´99, How to design a speech recognition application).

With computer development and mobile devices, a variety of communication channels will be used. IVR applications can at present engage the user in an almost natural language dialogue (Bruce Balantine et. al., ´99).

The interface has to be designed so that the user does what is expected, the limiting factor here is the designers imagination. A user study can minimize unexpected behavior from the user as stated earlier. Such a study can be integrated when examining the multimodal interface, but the study will become significantly larger.

2.2.1 State error and turn taking

There are two different approaches (Bruce Balantine et. al., ´99) of IVR interface design, either reveal the states to the user or hide the states. Both design works well, but depends on the user group defined for the specific interface. Novice users should be more engaged in a natural spoken language dialogue and expert user might get information on present state of the application. Presenting state information for novice users might confuse and mislead (note, this is true for an IVR interface).

If the interface and user is not synchronized a state error occurs and all user input mismatches the application's expectation. With a graphical user interface, the user can almost always follow the current state of the application. Text and graphics modal interfaces therefore display the state unintentionally. Whether state information should be displayed using voice may be optional. Expert users might want to see the state information, while novice users maybe do not want to see the state information.

State error, which most likely could happen in complex dialogue applications, can be amplified if the user is not aware of the application change of state. The user continues to interact with the interface and after several switches of state, an unrecoverable error occurs (breakdown).

A specific form of state errors is when the user thinks it is his turn, but the computer claims it is its turn. Therefore the user speaks too soon or too late. This form of interaction is called turn taking. If barge in (see below) is allowed, the computer might change state before prompting to do so. This will lead to state errors, which are difficult to detect and prevent. The worst-case state error would result in a breakdown, where the user and computer have to start over because an undoable error has occurred.

However turn taking does not have to be faulty either, the user might be an expert user and wants to change state when the computer claims its turn.

2.2.2 Barge in and memory

An IVR application (Bruce Balantine et. al., ´99) can support “Barge in”. This is when the user can interrupt the prompts read by the computer and force the computer to interpret what was said. Barge in is especially important when the user has already used the application and does not need to listen before knowing what to say (expert users).

It is important to have a high threshold for barge in recognition. The common approach would be to reject rather than accept. If a faulty recognition occurs it is more likely to confuse the user. Depending on the confidence score of the recognition, the application should determine whether to reject or accept what was said. The confidence score should be high, otherwise grunts, coughs, mumbles or background noise could be accepted by the recognition system. This leads to state errors.

An IVR application lacks the difference in memory the user has to use while interacting with the application (Bruce Balantine et. al., ´99). The IVR application has to be simple and layers should be kept to a minimum. Lists and prompts should also be short and concise so that the user can remember what was said and in which state the application is. Proposed and studied by G. A. Miller the short-term memory is limited to 7 +- conceptual tasks. A too complicated IVR – application where the user has more than 7 tasks in mind would not be usable.

Preferably less than 7 tasks / commands would be best. This is for short-term memory only, for long term memory, which can be identified as expert users. Good user interface design result in that the user learns the states of the application and does not have to hold information in the short-term memory.

2.2.3 Feedback and learning

Proposed by J. Nielsen ´92, system status should always be shown to the user, this helps minimize the state errors and amplification. This is most likely depending on the modality in use.

The application should give a fast feedback to the user so he understands that the computer is processing the input. If the user does not receive feedback in reasonable time, he probably thinks that something has gone wrong. This will often lead to turn-taking problems, state error and breakdown. The feedback can be on the form of a status bar or a melody playing. If feedback has not been received between two and ten seconds, the user will probably suspect that an error has occurred. A guideline formed by M. C. Maguire, ´99 for feedback should not take longer than 10 seconds, if longer feedback a status bar should be displayed

The state switches and interface should also be a cognitive learning process. It is important that the user understands how the application should be used and acts. A successful interface makes all users feel like expert users. Depending on the service this can be done to different extents.

2.2.4 Summary for speech dialogue design

An application using only text and voice lacks different important properties a graphical user interface has. Simple problems in visual interaction tend to grow to be a much bigger problem in an IVR application. Some important factors to consider:

1. Memory, difference between visual and voice.
2. Dialogues, barge in, turntaking, breakdown, state errors (Bruce Balantine et. al., ´99). The psychological interaction between human and machine.
3. Switching between modalities.

2.3 Identify users for the project

The user group interested in using mobile services and IVR interfaces in the future should be large. In the near future the development of Internet and computer companies will be based on wireless devices, services and applications.

At the present moment the telecommunication industry has suffered a setback in development of new IT-technology. Although content is often missing for new technologies, support for switching between different modalities is limited in the hardware. A diplomatic response would be that both hardware and software needs to be improved.

2.4 Scenarios

When and where can it be useful to switch modalities? Switching modalities can depend on where the user is physically. If the user moves between different environments different modalities are needed. The different scenarios can also depend on which type of application i.e. implemented. It may be easier for the user to use voice instead for text.

2.4.1 Moving between environments

When the user moves in the physical world, modalities have to be switched. There are unlimited situations where one modality has to replace another. Text and graphic interaction are the most common modalities used. The more modalities introduced the larger the user group will be. When the user moves, switching between modalities becomes an important issue.

Moving between different environments or scenarios is one of the most common reasons to introduce multimodal interfaces. Examples of scenarios where the user have to switch modality are summarized in sections 2.4.2 – 2.4.5.

2.4.2 S/N ratio

Depending on the environment the interface is being used in, the S/N ratio changes. At some time the Noise is too overwhelming for the IVR application to do successful voice recognition (see figure 4). Whenever the voice recognition fails, the user has two options to continue the interaction, either by redo the voice recognition or switch to another modality.

Figure 4. Typical no good S/N ratio (Courtesy of Inspiriogifts, www.inspirogifts.com)

An example of this would be if the user moves between home and work. At home it might be more time saving to use voice, but when moving in traffic text based interaction is more appropriate.

2.4.3 Privacy

Depending on the modality, different levels of secrecy and privacy can be maintained. It is important that the user never be forced to use a modality with less privacy. Certain applications may force the user to confirm transactions or if pushed information should be accepted. But with secret information, the user should decide which modality to use. There may also be some personal matters on why not use a modality.

The application should be context – aware, some context in the application should not be revealed on a speaker. It is simple to implement an user interface with context awareness, but to make a scenario - aware application is much more complex (technology like GPS, GSM positioning).

The user should be able to switch modality whenever (expert). The interface can begin the interaction in a more secure modality (see figure 5) and then switch.

Figure 5. Privacy should be nice to address (Courtesy of www.links.net)

Examples of interfaces where voice input could be questionable are banking interfaces, logins, dictations, personal information, etc.

2.4.4 Time saver / simplify

The idea of introducing a modality is to save time and simplify the human computer interaction. The two reasons do not automatically imply each other, but it would be nice. Best would be to introduce a modality that both help save time and enhance the usability. To consider is how the new modality should be used and why.

Examples of this are when a new modality has been added to the application, the complexity of both interface and application increases, but when user has been trained, it might save seconds or even minutes.

2.4.5 Busy in a modality – eyes and hands

In some situations, the user cannot for some reason use a specific modality. This can be when the main modality used for interaction has been disabled. The user should be able to use a different modality (Switching of a modality).

Often the user is busy driving a vehicle or it is too dark to really see the screen (see figure 6). A bad S/N ratio can be included in this category.

Figure 6. Situation where it might be good to change from visual to auditive communication. (Courtesy of GM, www.gm.com)

2.5 Switching modality

2.5.1 Switching of modality for input

The user should have the opportunity to switch modality independent of the environment. Depending on the user, one modality is more likely to be used rather than the other. New technologies such as IVR applications are more likely to be used by people interested in such technologies (World Wide Consortium, verified ´01).

An important research focus therefore emerges in integrating the mobile client into a collaborative system. This focus relates to automatic "information transformation". For example, a graphics-rich stationary computer may transmit a sophisticated image to a less-capable mobile terminal. Voice may be the only functional information modality for the recipient (World Wide Consortium, verified ´01). What does the initiator do? Describe the image? Perhaps a better solution is a sophisticated image analyzer that automatically detects important features and maps these into a text-to-speech synthesis description for audio presentation.

Alternatively, if the mobile receiver is operating from a mobile client with a small screen, the image analysis could transmit features for a regeneration of the original image, to supplement the voice description. The mobile client may give a voice description of important changes to be made in the image and displays the suggested modifications.

Similar information translation for the mobile devices can be applied to tactile and gesture information, as obtained from force feedback and from handwriting pads. Depending on the hardware located at the client side, different system architecture can be implemented.

Multimodal support (World Wide Consortium, at http://www.w3.org /TR/multimodal-reqs, verified ´01) can be categorized into three categories:

1. There is no requirement that the input modalities are simultaneously active. In a particular dialogue state, there is only one input mode available but in the whole interaction more than one modality is used. Inputs from different modalities are interpreted separately. For example, a browser can interpret speech input in one dialogue state and keyboard input in another.
2. There is no requirement that interpretation of the input modalities is coordinated. In a particular dialogue state, there is more than one input modality available but only input from one of the modalities is interpreted. For example, a voice browser in a desktop environment could accept either keyboard input or spoken input in same dialogue state.
3. In a particular dialogue state, there is more than one modality available and input from multiple modalities is interpreted. When the user takes some action it can be composed of inputs from several modalities – e.g. a voice browser in a desktop environment could accept keyboard input and spoken input together in same dialogue state.

2.5.2 Switching of modality for output

The same as above can be stated for output when switching modality (World Wide Consortium, at www.w3c.org/voice, verified ´01):

1. There is no requirement that the output media are rendered simultaneously. For example, a browser can output speech in one dialogue state and graphics in another.
2. There is no requirement that the rendering of output media is coordinated any further.
3. Coordinated, simultaneous multi-media Output

2.5.3 User groups novice / experts

Before designing the application, a target user group has to be defined. Users can range from novice - to expert users. The design of the user interface for novice users differs from expert users, depending of the complexity of the interface.

Expert users use the application heavily and have excellent knowledge of how the application works. Novice users do not use the service often and need an interface i.e. simple and does stimulate fast learning of how the application should be used. It is important for the designer to make a good impression on the user and implement an intuitive interface design.

2.6 Multimodal system architecture

Consider the different modalities as different clients in a client-server model. Since mobile clients often lack computing power, the application should keep processing at the server-side and not rely on the clients. Keeping the client interface simple helps the application in its capabilities and makes integration of new modalities easy (as long as the new modality rely on the same communication protocol as the other modalities).

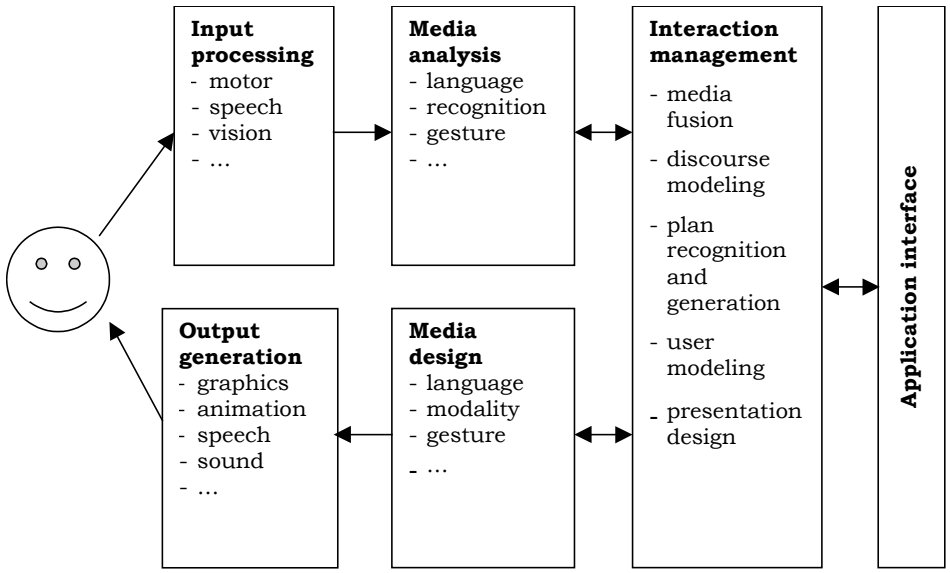

Figure 7. This is what a multimodal infrastructure would look like. (Courtesy Maybury and Wahlster, ´98)

The modalities should be synchronized with each other so state error does not occur between the modalities. This is similar to when the user gives two conflicting inputs in different modalities. The application cannot know which input is correct.

As we can recollect, the modalities should only mirror the same application to improve design and usability (see figure 7). This has to do with the limitations of mobile devices.

2.6.1 Guidelines

Literature regarding guidelines for general multimodal interface design are easy to find (Chris Johnson et. al., ´98). These documents involve movement of mouse/pen and voice as a bimodal interface. It is harder to find any guidelines for multimodal interaction combining text and voice. However, with mobile devices beginning to emerge, different multimodal interfaces will become more and more common.

The perfect multimodal interface is when the user always can decide what modality to use. The user can switch between the modalities without having to be active in switching. In other words, the user can be passive and the application responds in the same modality as the user uses. This could be somewhat annoying, since the input modality might not be what the user wants as output modality.

2.6.2 Limitations of mobile devices

The communications environment for the mobile user is characterized by limited bandwidth and interference (fading or shadowing that may contribute to packet loss).

Consequently, low bit-rate, robust coding of transmitted information is more of a central issue than it is in broadband wire networks. Additionally, wireless transmission carries increased vulnerability to interception, so interest in economical techniques for encryption and privacy is large (Ericsson, verified ´01).

Power for mobile devices is a major concern. Major vehicles, on the other hand, are usually able to supply enough power from the main power supply (see figure 8).

Figure 8. The mobile device used for the implementation (Courtesy of Ericsson)

With these constraints, we can view global networking as having the following architecture (Steve Jefferson, ´00):

The underlying broadband wire/fibre network will be as implemented as economics and geographical limits disappear. Intelligent routers and switches will fulfil quality-of-service (QoS) and bandwidth requirements. With further development, dynamic allocation of resources (communication bandwidth, distributed storage, and computation power) will appear. Stationary computers will have the obvious advantages of the broadband, low error and stabile environment. Servers can be available for intense computation, archiving, digital libraries and network management.

Stationary computers may also enjoy the full range of sophisticated modalities for interacting with the information system and other users. In particular, user interfaces using sight, sound and touch dimensions can be implemented. Visual gesture, speech recognition, text-to-speech synthesis (TTS) and tactile feedback are all technologies now evolving.

This range of capabilities has not yet been demonstrated for mobile devices (which may be significant amounts of time in transit – time that might otherwise be used for productive work). Complex graphics, video, and large database visualization are difficult to manage. Conversational control and interaction will consequently be a killer application in the mobile environment.

Hardware limitations of mobile clients set the limits for multimodal interaction. Most of the technology today is suited for stationary computers and designers have to revoke to an earlier state of development of technology, a much more constrained computing environment (WAP forum, verified ´01).

Mobile devices are limited in:

1. Less powerful CPUs
2. Less ROM / RAM
3. Restricted power consumption
4. Smaller displays
5. Different input devices

And with limits of mobile devices comes limits in mobile networks:

1. Less bandwidth
2. More Latency
3. Less connection stability
4. Less predicted availability

These facts results in the following requirements for mobile applications (and even more constraint on multimodal interface design):

1. Compatibility in services
2. Secure
3. Robust and Reliable
4. QoS

With these limitations in mind, a first interface for WAP can be made supporting fast and reliable applications. But whether the user interface supports good usability, heuristic list can be a useful tool.

2.7 Heuristic evaluation

Heuristic evaluation helps the designer to rethink the interface. The list below states 10 of the most important steps in a heuristic evaluation (J. Nielsen et. al., ´94).

The output of the evaluation consists of a list of errors produced in reference to the heuristic evaluation list.

1. Visibility of system status. The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.
2. Match between system and the real world. The system should speak the user's language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.
3. User control and freedom. Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undoes and redoes.
4. Consistency and standards. Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.
5. Error prevention. Even better than good error messages is a careful design, which prevents a problem from occurring in the first place.
6. Recognition rather than recall. Make objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.
7. Flexibility and efficiency of use. Accelerators, unseen by the novice user, may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.
8. Aesthetic and minimal design. Dialogues should not contain information i.e. irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.
9. Help user recognize, diagnose and recover from errors. Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.

10. Help and documentation. Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

2.8 User study and tests

User studies can reveal fatal flaws with the application. This happens because the diversity between developers and users of the application. The developer is an expert in contrast to the users. The user study should be taken seriously and problems found should be corrected.

The user study should be an integration of the usergroup, implementation, theories and expected result. The experimenter does have a hypothesis regarding what will be the outcome of the study, but certain details are not defined.

To form the questionnaire and measure what is interesting the experimenter does have to make the study so that it reflects what is true. There are different opinions on how only the presence of an experimenter makes the study faulty. If kept in mind (and considered) that the study is under influence of the one conducting it, the study should have significant value.

2.8.1 Prepatory tests

To get some information on what could be interesting to test and on how the user study should be conducted. There are three ways that could be used in this project to gather information regarding what is interesting to monitor.

1. Interview different users and ask what can be the benefits of a multimodal interface. It should be interesting to explore ideas of services suitable for multimodal interaction.
2. Ethnographic study on people in different scenarios use technologies like voice and text input / output. An ethnographic study monitor on how people use different modalities and wireless applications in different environments. This kind of a study is not a test of the actual application, but more of a pointer on how wireless applications and clients are being used. However, an ethnic study is very time consuming and restricted to the environments the users are located. The ethnic study in this project will have low significance and involves study of log messages only see point 3.
3. At Sema Group Infodata IVR applications are being developed. Often the user input is being logged for information on how the application can be improved. Looking at this logged input data, makes evident the most apparent problems users encounter with dialogues using natural language.

2.8.2 Setting up the test environment

Several important factors should be considered when setting up a user study for testing the application. The test environment often badly reflects the scenarios of the real world, where the interface is. This applies especially for mobile interfaces since the user often moves between different environments. One way to test the mobile application would be to follow the user around different scenarios and monitor the behavior used (ethnic study). However to conduct this sort of study is time consuming and the test might not be as affective.

Monitoring two test persons at the same time can help to ease up the tension and formal feelings of a test environment. Letting two persons to co-operate could give an interesting dialogue that might give information that would be missed when a single user thinks out aloud (often used in Wizard of Oz tests, E. Bonharme, ´01). The dialogue between the two users towards the application is used when possible.

Therefore it would be best and most convenient to set up a testing environment at a place that suits the test persons. To test two different users at the same time can be effective to retrieve information.

2.8.3 Questionnaire for user study

Forming questions with a general approach leaves the user to freely associate what is good / bad with the application.

Too specific questions increase the influence of the experimenter and an invalid study could be conducted. Too general questions will lead to scattered results, but invites the test person to use more imagination. In this project, the questions have a relative general approach to leave the user associate freely (see attachment A and B, Questionnaire 1 and 2).

Both before and after a user study has been conducted it is important to get user information and thoughts about the application. A questionnaire before the test to get user information regarding computer knowledge, age and experience with both IVR – and mobile applications was used in the study.

After the test person has tested the interface, another questionnaire should be answered. The first questionnaire gives background information about the user and the second questionnaire gives information regarding the service. It is important that the questions reflect what is being studied. The questionnaire should be formed so that the question does not trick the user to answer in certain way, but answer questions that are really important. As stated earlier, the experimenter should realize that the study is under his /hers influence.

2.8.4 Selecting users

People with different backgrounds, cultural status, age and technology experiences should be selected. This is often a difficult task to do and the test persons in this test all have some computer skills and have all been in touch with computers.

It is also better the more users that will be tested. Due to the time frame and the scope of the project, 15 different persons will be tested.

2.8.5 Analyzing results

The results from the test show different statistics on how the application could be improved and what problems the users experienced. All results are integrated and graphs should be displayed.

2.9 Using the results from the study

A complete redesign of the application is generally not necessary. The user study gives developer a hint of what should be corrected in the implementation. After the modifications are done a new user study should be conducted to see whether the new design was better or worse. It is essential that the user study mirror real life situations. It can be hard to create a real world environment and scenarios in a laboratory. There are several factors that have to be considered to do a successful user study and therefore a successful user interface.

3 Problems with multimodal interaction

With a user study a simple prototype can be tested and evaluated, but whether a technology will have some future use only speculation and hypotheses can be made. If an application will not have any future use, it might be a waste of time and effort to develop the new technology. If the technology is useful and might have further use, an implementation might be worth the time and effort.

The markup language used must support the detection of conflicting input from several modalities. In a speech and graphical user interface, there may be simultaneous but conflicting speech and mouse inputs, the markup language should allow the conflict to be detected so that an appropriate action can be taken (World Wide Consortium, verified ´01).

Consider a multimodal number search, the user might say “Anders Andersson”, while typing “Bengt Bengtsson”. How should this be interpreted? The application might resolve this by asking, "Did you mean Anders Andersson or Bengt Bengtsson?”

3.1 Is there a future of multimodal interfaces?

Most speculations and hypothesis can be formed under the above heading. Gathering information and gaining knowledge of a new technology, good assumptions can be made.

Multimodal interfaces is not a new technology, it has in fact been used in several applications in different industries for several years (see references CUBRICON, DARPA). These interfaces are being developed for specific applications or purposes, specific environments and specific work conditions. It is with the development of faster and cheaper home computers that multimodality will gain new grounds. With even further development this will apply to mobile clients. At the present time mobile clients does not meet the hardware requirements of multimodal interface. With iPAQ and similar devices computer power for interfaces will increase. However it is not likely to have program logic in the client, but at the server – side. Another problem with multimodal interface development is that when implementing an interface with client – server approach, network communication is an essential factor. It is preferred to use IP based network technology, but it requires at least GPRS technology (preferably 3G).

WAP has not been very successful in mobile information technology, why? Since WAP is the only existing working protocol for mobile application development, is it a good idea to develop multimodal interfaces based on WAP?

Make a hypothesis that multimodal interfaces will be enhance the WAP user interface. What kind of services should be developed?

3.2 If there is a future, what should be considered in multimodal interfaces?

Suppose there is a future market for multimodal interfaces not only for stationary computers but also for mobile clients. What can the user expect from a multimodal interface?

Introducing mobile services, work process will change to be more distributed. Users can communicate in new ways and questions regarding collaboration in work become important. How should the services be used in work? And how will these distributed system look like?

4 Implementation

The implementation will include some functions of the Elektroniska TelefonKatalogen (ETK) service at Infotorg. The prototype will show how different modalities can be integrated into a single interface. A useful mobile multimodal application is the ETK service. It is a small service, it has a low layer of interaction and it is needed. Elektroniska TelefonKatalogen is the number search implemented at Infodata. Searching through the database for numbers is being done by using methods (functions) in Java. These methods are derived from a Service Manager (SM) holding different kinds of services. So in order to make a search, call the SM and supply it with the name of the service and search criteria.

To implement bigger interface such as an multimodal interface towards Statliga PersonAdressRegister (SPAR) would be a much more time consuming task and would not be used as much as ETK searches (at least for the persons with no professional interests). ETK has a broader user group, since the common man have access to the service.

4.1 Overview

The system for supporting these relatively simple tasks is a basic telecom system based on Java. A multimodal architecture should be designed on a client – server basis, where the server contains most program logic and client serves as a window towards the server. In multimodal programming this is important, since different modalities have different requirements. It is easy to introduce new modalities, since the core application can be reused. The server provides the application and the clients the modalities. When a new modality is going to be used it can easily be integrated with the server (see figure 9). Since the application will always be the same, all that has to be done is to design the modality interface.

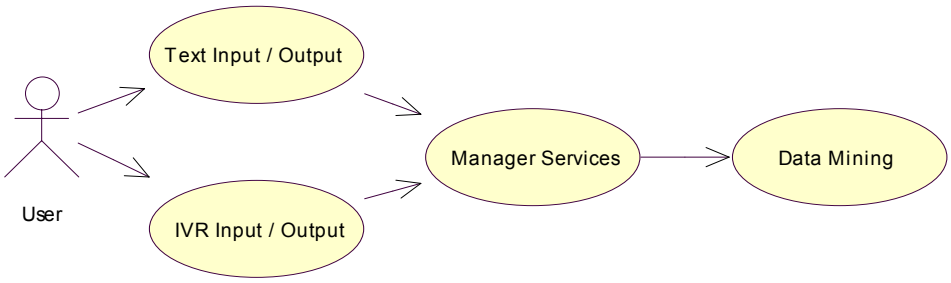

Figure 9. Implementation user case, the functions that can be performed.

The data mining can be done by different modalities, but only from the same service manager.

This is consistent with the view of client – server programming. This architecture is not only multimodality specific, but also how a multi – device application would look like. Because of the requirements, the system preferably has to be developed on a program language that supports good network communication, Java.

A client – server approach is a preferred implementation in multimodal interaction, but with development of mobile clients, peer to peer technology can be used. This will take some time though, since already "ordinary" stationary computers are "slow" acting as servers. Peer to peer (p2p) technology will take time to adapt to mobile clients, since computational power is not the same as with the stationary computers.

In the WAP interface the user can decide whether to use voice recognition or text input. At least for input this is true.

4.1.1 IVR Application

At Sema Group Infodata a toolkit for building voice recognition application is used. The toolkit is developed by Nuance and contains different tools based on Java. The developer using the Nuance system implement reusable components called Speechobjects (SO). These Speechobjects is a Java API made by Nuance. In other words the whole recognition process, playing prompts and call control functions can be done in a Java program using the Nuance API.

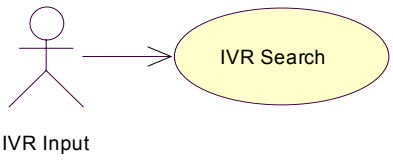

Figure 10. IVR input use case, Functions that can be made by the user.

A typical use of the IVR application is to say the given name and last name of the person i.e. being searched for (see figure 10). If the application does not recognize the spoken name, then the user will be prompted to do another try. If the application does recognize the name a list of results matching the name will be read with a TTS agent (also known as Ingmar). When the list of person is being read the user can make a transfer call by saying "transfer call".

The search in the database takes on average 10 seconds to perform, this might make the user to feel that a fault has occurred. Often the system does not take more than 10 seconds before the first feedback, but a prompt saying, "one moment please" exist. This gives the user information that the input is being processed.

Example (English translation):

IVR - *"Say the person first - and last name"*

User – *"Anders Andersson"*

IVR – *"A list will be read, if you want to make a transfer call to the person, say transfer call. One moment please"*

TIME: 0-15sec

IVR – *"Number of results [n]"*

IVR – *"1, Anders Andersson, Stockholm, 08-123456"*

IVR – *"2, Maj Lindström o Anders Andersson, Nacka, 08-654321"*

IVR – *"3, Anders Andersson, Sollentuna, 08-678910"*

User – *"Transfer call"*

IVR – *"One moment please"*

TIME: 1-2 sec

If the IVR application does not recognize the said "transfer call", the TTS agent will continue to read the result list. If the list is at the end, the IVR will say, "This is the end of the list, the call will be terminated". This is a problem since this is not a browsable list and the user has only one chance to get a successful recognition of "transfer call" (see figure 11).

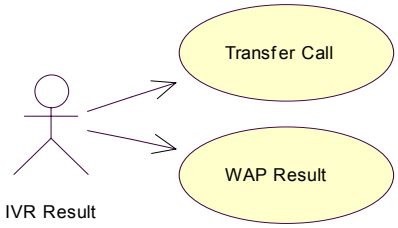

Figure 11. IVR result use case, Functions that can be made with the IVR result.

It is important to design the speech dialogue so the user feels comfortable using the IVR application. It is important that the user understand what needs to be done to accomplish the task. The application has to direct the user to a new state and the user should understand what is needed to be done. Also the prompts should not be too complex or long, since then the IVR application will be much slower than the WAP interface.

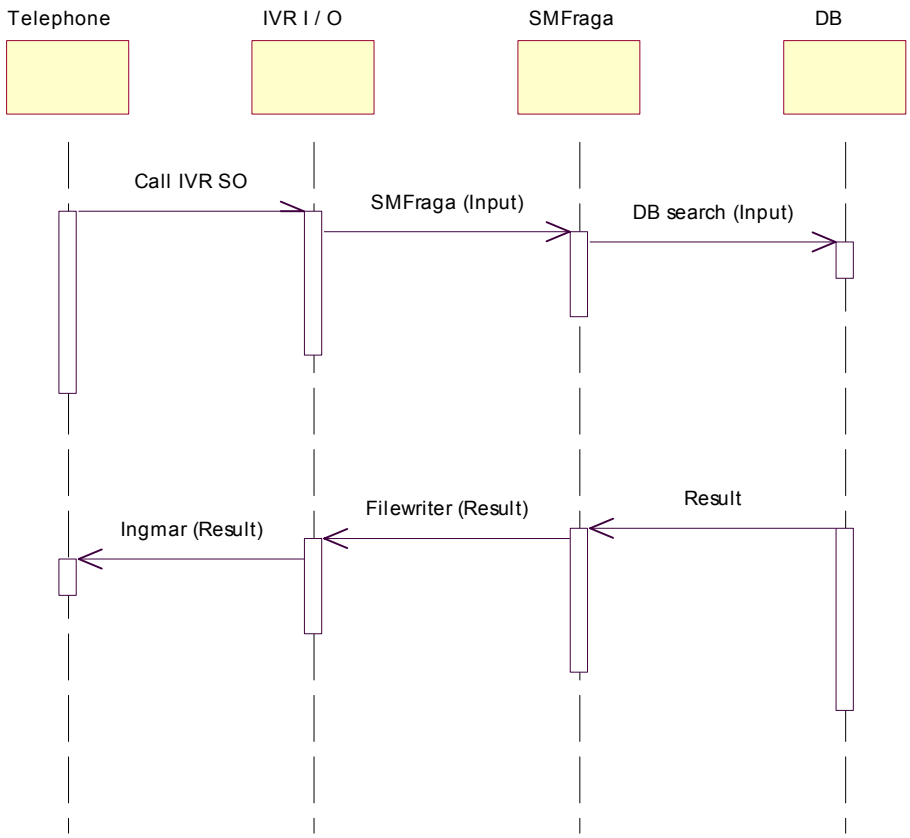

Figure 12. Sequence diagram of IVR interaction.

The result is written to the files jspres.jsp (see figure 12) with 10 persons / page. If an error occurs, the file jspres.jsp is written containing the error message. Typical error messages are too many results, the person does not exist etc. The error message is read by the TTS agent (Ingmar).

4.1.2 Text interface

The other modality to be used, text, is implemented on top of WAP. To keep the WAP interface simple and relative basic is important, because of the limited display and keyboard on most mobile devices. As the limits of mobile devices are known, the design of the text interface should be approached by using many different WAP pages instead of using large text strings. If the bandwidth is relative high, this is true. This also depends on the cache size of the client.

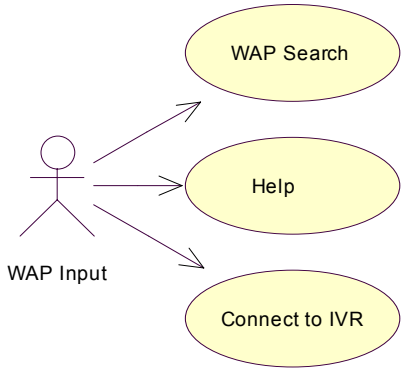

Figure 13. WAP input use case, text functions.

The text interface consists of several input fields where the user can enter: family name, given name, city, street name and street number. To perform the search in the database, the user has to click the “go” hyperlink (see figure 13). With the hyperlink, the input data is posted to a servlet (SokEtk.java). The servlet extracts results (see figure 14) from the database fulfilling the input data and writes the result to files (jspres “x”.jsp). Depending on how many results extracted from the database, a number “x” of files are written with 10 objects in each file (this is due to the limited cache in most mobile clients, the r520 can handle roughly 1.3 kbytes in one character stream).

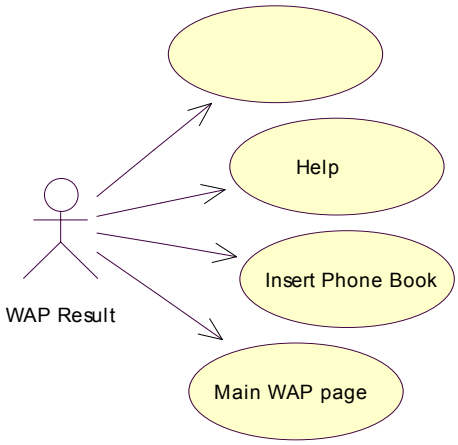

Figure 14. WAP result use case, functions of the text result.

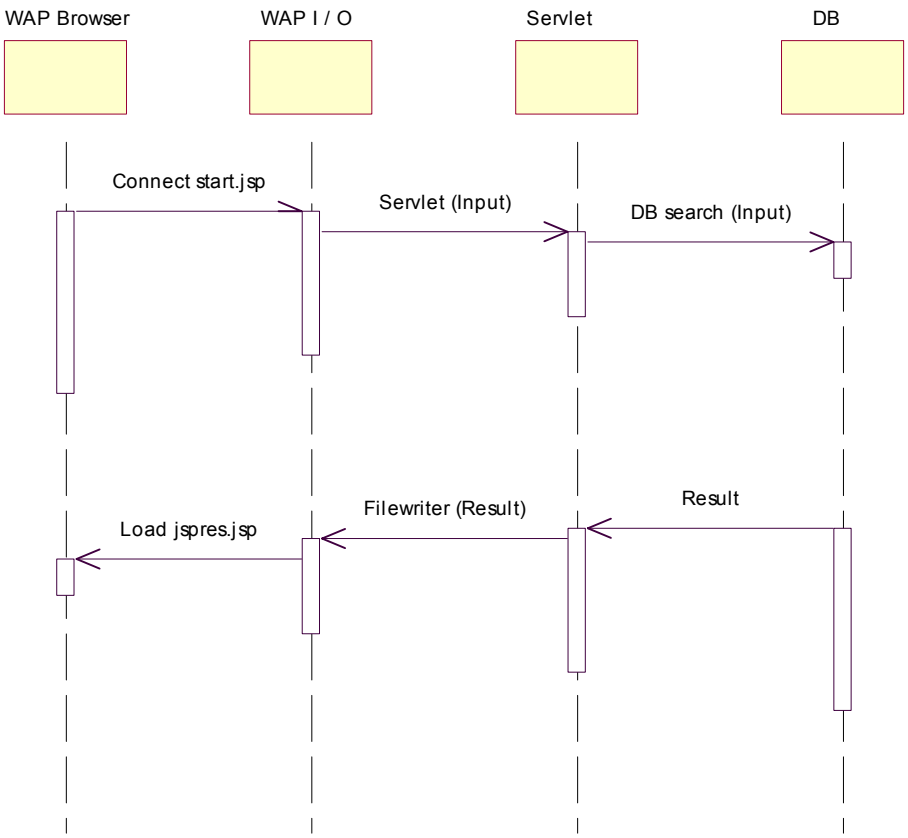

Figure 15. Sequence diagram of the WAP interaction.

The limited screen on most mobile devices makes the presentation of data user hostile. The result shown on the WAP page cannot consist of unnecessary characters as even the WML tags are counted. Because of the limited screen it is difficult to get a good overview of the interface and what the services does.

4.1.3 IVR – WAP synchronization

To implement a multi – modal application the IVR - and WAP interfaces have to be synchronized. What architecture the human – computer interaction is not complex, which makes synchronization between the two interfaces/modalities simple. In more advanced applications the different interfaces have to switch states dependent on each other. As one can see from figure 7, the system is a distributed system. The system can be integrated into one server and one client. This approach has been done for technical reasons and to decrease the load on the server.

This implementation uses one modality at a time and the interaction will be sequential.

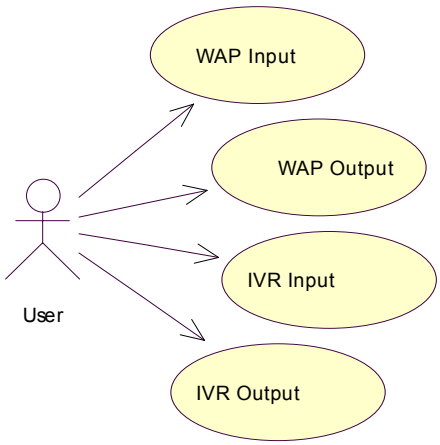

Figure 16. IVR – WAP Synchronization use case, the functions that can be made through the multimodal WAP interface.

In a more advanced application, the different modalities have to exchange data for the different modalities to be in the same state. If we recall the first architectural view of multimodal system (chapter 2.4), most program logic of the application resident on the server (see figure 17). If this is true in the application, the modalities do not have to exchange data to know what state the application is in. This architecture is better, but more complex for the programmer to implement.

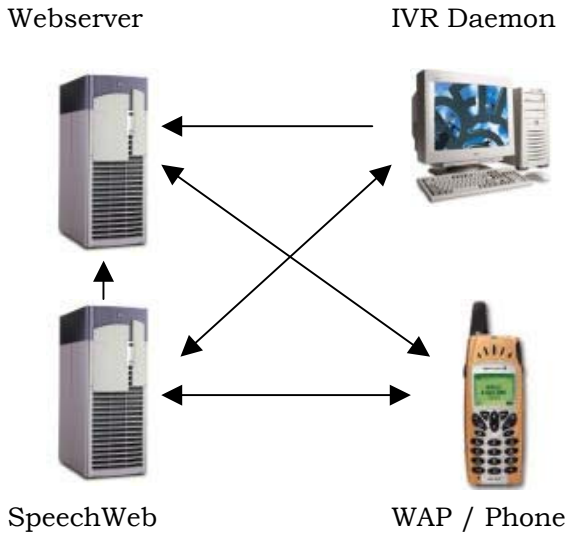

Figure 17. This is the architecture of the distributed system. (Images from www.ericsson.se)

4.2 Java

This system is developed in Java, with input / output presented in XML, WAP, Servlets and voice. WAP and voice will be the presentation of data towards the user. XML, JSP, Servlets and Java are what the underlying system is based upon. However Java and XML is the foundation of the system.

Sun has released a new API, where the developer can access several call-control functions in mobile clients. In the future, application development tools supplied from NUANCE will not be necessary to develop IVR and call – control applications. Although a tool makes implementation easier, it is possible to make a whole IVR application by using different Java classes.

4.3 WAP

Presenting data, WML is used. WML has limited functionality and JSP is a good complement. WML is the HTML for WAP. WML and HTML are in fact two related markup languages, with WML being stripped to ensure faster interaction and lower network load. WAP having problems in standardization does not increase the use of mobile services. Unfortunately WAP does not function properly and there are few content providers. The user study points to a content problem with WAP and new standard should be made.

Generally WAP communicates through two gateways, one gateway at the ISP provider and the other at the content provider. Communication between these gateways has failed, especially over GPRS.

4.4 WTAI

WTAI or Wireless Telephony Application Interface help designers implement call-control functions. The WTAI functions can be called as URI:s in WML or as functions in WMLScript. WTAI functions are few and limited in functionality

Sun Java has released a Call Control API, where Call Control functions can be implemented directly in the Java code.

4.5 VoiceXML

VoiceXML is a markup language similar to HTML and WML. VoiceXML is a markup language for speech dialogues and built with XML. SpeechObjects (SO) by NUANCE can be incorporated into the VoiceXML code by using the object tag (<object></object>). The SO is built by methods included in the NUANCE Java API. SpeechObjects can be tested individually by the tool V-Builder (NUANCE). Since SO only contain Java functions, SO can be compiled with the ordinary java compiler, and executed with java commands.

VoiceXML contains several tags for building a speech dialogue (e.g. <audio>, <record>, etc.). VoiceXML is an XML standard held by W3C and the specification can be read at www.w3c.org/voice. V – builder itself is built on a C++ core.

To start the SO it must be incorporated into VoiceXML code and run by the SpeechWeb server. A specific telephone number associates with the SO and when the user calls the number, the VoiceXML code initiates and consequently the SO.

When the telephone has been associated with the SO, WTAI links can be used to link the number to the WAP interface.

5 User study

5.1 User group

Depending on what context the interface is going to be used in, the user group changes. Mobile services have a large area of use and it follows that a feature such as multimodality, widen these areas. With development of intelligent houses and program language like Jini, an even bigger need for voice-enabled services will emerge. Will multimodal application be the next step of mobile application design? The neat part with multimodality is that mobile applications become user-friendlier (if carefully designed) and more users will be able to use the service. Larger screens, faster communication, cost and more computational power on mobile devices will also increase the use of multimodal services.

One problem with mobile clients and applications is the users are only technology-interested people. Hopefully with development of multimodality, novice users can use this new technology. However, in a worst-case scenario it will take a generation before mobile clients and applications become common.

Interfaces with poor usability will be developed at first and the user should not be afraid to use the new technology but instead be more explorative. Another important factor is that the user should be able to save time and money. This is why the user should switch modality, because the few extra seconds it takes to use the other modality add up to the total time and cost. The interface should be simpler to use than the unimodal user interface. If the multimodal interface is too complex, the interface is not successful as a user interface. At worst the user want to revert back to a unimodal interface. The multimodal interface can of course be used with a single modality.

The user group consists of first time users who are interested in using a multimodal interface. Expect technology - friendly persons with good understanding of the current technologies, probably without WAP experience. Users of mobile devices and wireless Internet, are often already stationary Internet users. This sort of Internet – readiness can lower the threshold of learning to use the Internet with mobile devices. Identifying the users next is to identify the different scenarios.

5.2 Scenario

Scenarios capture how user wants to use a specific modality. The user prefer to use a modality may depend on both external and personal matters. It is important to understand both why a specific modality should be used and when it is appropriate to use it. If the multimodal interface is too complicated to use, the user should always be able to use the interface as a unimodal interface.

The most common scenario would be where the S/N ratio is bad, privacy, “busy hands”, save time and cost. S/N ratio and cost are the two most important reasons depending on service. These two reasons would be most important for the implementation.

A number search does not affect saving time and privacy, there is already a number search “118 118”, where the operator finds the number for you. This is faster than surfing to the bookmark and clicking on a link (although with GPRS this is done fast). It is a service that does not need any specific privacy either.

Using the service “118 118” is affected by high costs (about 11 Swedish Krona /minute), high background noise and “busy hands”. Identifying the scenarios where a multimodal interface can be useful, when implementing a prototype.

6 Results

The results presented will be from the heuristic evaluation and the user study. The heuristic evaluation was done before the user study was conducted (of course). Results and comments from the users were considered and analysed prior further development of the application.

Problems and difficulties using the prototype are presented, but the implementation and the system architecture have already been explained (see chapter 5 implementation). The study consisted of 15 persons with different backgrounds. All users had different types of jobs, interests and experience with mobile devices and Internet. The study was conducted in the users own work environment. This was comfortable for all users and a more positive and attentive approach was done towards the application and interface.

6.1 Heuristic evaluation

The heuristic list used to evaluate the usability of the interface is the one proposed by J. Nielsen, ´92. Some violations against the list are stated below:

1. Condition 3, User control and freedom. The IVR application does not support any browsing through the list and does not support any “hotwords” (help, quit, next, previous, pause). The WAP interface does fully support undo and redo.
2. Condition 5, Error prevention. Although the multimodal application does display error messages, there is nothing like a fool - proof application. The application itself does not cast any exceptions, but the database might (in terms of, user does not exist, to many results, wrong password etc.).
3. Condition 7, Flexibility and efficiency of use. The IVR application is stationary and does not support any acceleration for expert users. There is no possibility to do a faster interaction. Expert users will have to use the multimodal application the same as the novice users.

6.2 User study

The test began with filling out the first questionnaire, browsing through the interface and afterwards filling out the second questionnaire. The users were asked to search for a number in Stockholm using the modality, which seemed best. Some of the questions had to be neglected, since they did not serve the intended purpose see (2.8.3 Questionnaire for user study). Some of the users were recorded on video in order to review the user interaction later and extract information. The results from the questionnaires have been summarized (same kind of answers will be stated once) in the following sections.

6.2.1 Results from questionnaire 1

Why are you (not) using mobile services?

+ Interesting to test the new services and the new technology (GPRS).

- No need for WAP or mobile services, have not connected yet to the service.

- Cannot with “pay card” (kontantkort) for mobile phone.

- Can find the services on stationary computers or other places.

- There are no usable services/applications.

- Slow interaction and too many buttons have to be pressed.

- No mobile device.

- Slow connection (at least with non GPRS).

- Too high costs for private users.

Do you think multimodal interfaces are useful?

+ Improves usability of a service.

+ More persons can be helped as a contrast to manual service. Those who really need help from an operator can get help without queues. Multimodal interface is more efficient of allocating resources.

+ Does not have to disturb other persons (meetings) and if hands are occupied with something.

+ Flexibility. Best from a usability aspect to get information the way it is best suited for the situation.

+ The targeted usergroup increases, e.g. physically disabled users.

- As long as it does not mislead the user.

What more services could be usable for multimodal interaction?

+ Ticket sales, tourist information and webportals (Alltomstockholm).

+ Microsoft Office, Email and similar software (stationary computers).

+ Search engines.

+ Personal butler (agenda, calendar, shoppinglists, searches, intelligent houses, identification, verification, etc).

+ Vendor machines, intelligent kiosks.

+ City navigator, Mapquest.

+ Software for physically disabled persons.

6.2.2 Results from questionnaire 2

Will the multimodal interaction be more useful?

+ When it is too dark to see (busy with other things).

+ Exciting new technique.

- Multimodal interfaces involve bigger requirements on the design of the interface.

- Standardization of new technique is needed, this is lacking in WAP.

- More errors than with personal services. Error recovery must be best.

- It is easier to call 118 118, DN 5678 5678. If the OS (EPOC) in the mobile phone could be multimodal, so to make calls and initiate the WAP session could be done multimodal.

- This might be better for businessmen rather than younger people.

Are multimodal interfaces useful on stationary computers?

+ For home use it might be best.

+ Not only for ordinary computers, but all kinds of computers, fridges, ovens, all kinds of machines.

+ Would be much better if one could e.g. open a document by voice.

+ If it would be faster interaction. Tele-surgery, and CAD application which demands usable and fast interaction.

+ Get rid of the complex keyboard.

- Might be disturbing if many uses multimodal interfaces in an open landscape etc.

Was the service well suited for a multimodal interface?

+ Searching for telephone numbers is an important service that should be developed further. The service could be used by most user.

+ Easy to understand and usable service.

+ Relative little in - and output, makes good and fast interaction.

+ Typical service that could be used when being mobile, e.g. when driving a car.

+ Not a ready product, but it is interesting to test new technology.

- One of the problems is the small screen.

- The interface should be developed further, but useful technology.

6.2.3 Statistical results

The persons in the user study were told to set grades on service implemented (see figure 18). The service itself was not the most interesting part of the study, but if multimodality as a technology was useful. Nevertheless, Infodata can be said to be a content provider and many similar services might be implemented. If this was an interesting service to implement a multimodal interface, Infodata has a good start for other multimodal services.

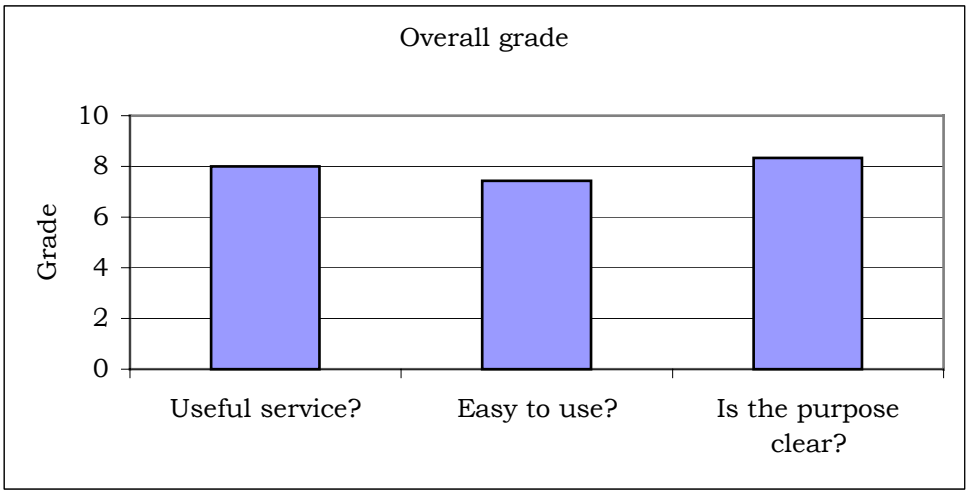

Figure 18. Average grade from the 15 test persons in the user study

It should be interesting to know why mobile services are not used (see figure 19). Formulating an answer to this question can only be a general one.

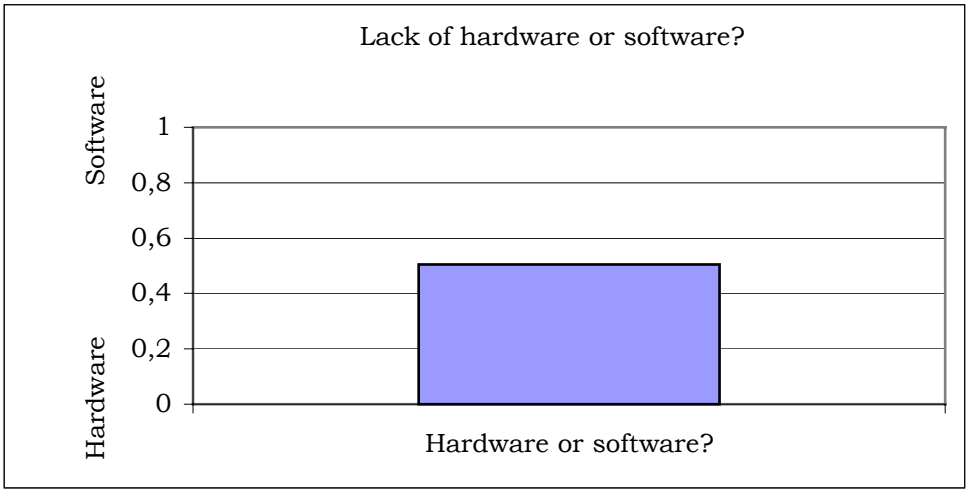

Figure 19. Why do you not use mobile applications?

It is important to know which kind of users you are testing. As I suspected the users will be technology friendly (see figure 20), but this is

not always the case. Maybe this question should be reformulated to “Is new mobile technology an improvement?” A more specific question pointing towards mobile technology might give another answer.

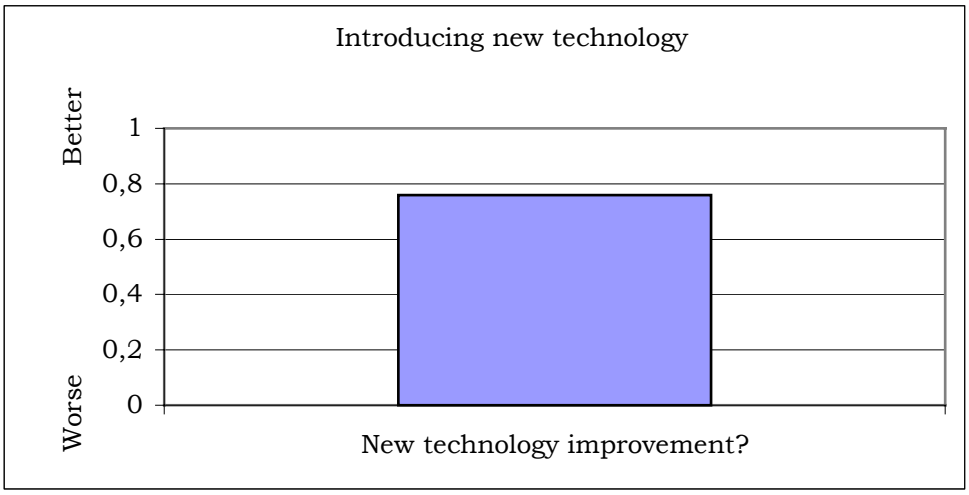

Figure 20. Does new technology tend to make more complex systems?

As said earlier, the problem with users that does not understand what multimodal technology could be used for should be accounted for. The results stated in sections 6.2.1 – 6.2.2 showed that there are areas where mobile applications might be used, but whether this is multimodal systems or not should have been asked. People that worked with technology and had tested IVR applications before were more positive towards new technology.

7 Conclusion

Multimodal interface is not a new technology. It has existed for decades in specific industrial environments (R., Bolt, ´80 and M. Heigel, ´55). It is not a new feature for home computers, multi-channel communication is an “old” technology. The use of keyboard, mouse, audio feedback in the home computer is a primitive form of multimodal system. The new concept is that all tactile interactions are excluded and audio and visual channels are the main modalities.

An interesting observation from the user study is the existence of an obvious misunderstanding of what a multimodal interface is and what it can do. Most users in the study did not understand what a multimodal system was and how it should be used. The users did not understand the fundamental concept (as I understood it). Multimodal or multimedia interfaces are obviously something not seen before by the common man. One can think that this is strange, since multimodal systems surround us everywhere. Stationary computers are multimodal systems with both keyboard and mouse as input and sound, graphics, force feedback and text as an output. Obviously these kinds of multimodal systems are integrated to that extent that no one notices them. Voice as a modality is a new feature for most persons and consequently more obvious than other modalities. Force feedback has been incorporated successfully in many computer games will voice do the same? Voice is usful as stated in different scenarios, the question is not if, but when multimodal interfaces will emerge.

Why multimodal interface? Pros and cons can be discussed on whether to develop multimodal interfaces or not. The negative aspects are the extra work load on the designer, license costs, design time and (hopefully not) decreased usability for the user. The positive aspects are the widen user group, increased usability, flexibility and faster interaction. If the user is not satisfied by the multimodal interaction, he can always use the interface as a unimodal one. In some sense a multimodal interface is an extension of a unimodal interface! This is true if the modalities can be used separately.

Current mobile technology is a bit restrained. Interaction is tedious having to browse through several menus before connecting to the service. With PDA and increased support for Java technology, the device itself could be implemented as a multimodal device. This could make the user connect to the service easier. There are many limitations at the present time, but in the near future with PDA and 3G supports multimodal interaction will be easy. This might be the case, but the results from the user study pointed to a content problem with mobile interaction. There exist in fact few services especially on top of WAP. Almost no one uses WAP and the fact that telephony technology companies (Nokia, Ericsson, Motorola, etc) have different standards makes WAP even more users hostile. Whether this is hardware or content problem cannot be determined, although content is missing not the hardware.

The most significant difference between stationary computers and mobile devices is obviously the mobility. With mobile interaction the user does encounter more scenarios than with stationary computers. With limitations of mobile devices, multimodality becomes a more interesting issue.

The services proposed by the users in the study show that information services, where the user needs important information are needed and usable services. This has to do with either lack of vision or the fact that multimodality suits certain services better. Both are right, it is hard to see what future use of a new technology can be. It is not suitable to have a bank service or dictation service using voice as an input or output. Of course this depends on the surrounding environment, but some applications would be better of with multimodality?

Multimodal interfaces have a dynamic allocation of resources at the content provider. Hopefully more users can use the service without having to talk to an operator. The availability increase with these interface, since people can perform tasks without being physically present (see section 7.3 Multimodal Future).

PUSH technology enhances multimodal interaction further. From the IVR – application result and updated information cannot be sent to the mobile client. With PUSH enabled, the IVR application can initiate the WAP session. Without PUSH the user has to begin the interaction with the WAP session and then switch to the IVR application.

The technology exists to implement advanced multimodal systems (at least for stationary computers), but the lack of vision and perspective slows the development of new applications.

7.1 Future work

Multimodal interaction will be a great enhancement not only for standard interfaces, more likely will have a greater impact on mobile applications. Combine different modalities is not a new feature in mobile interfaces (industrial specific applications), but it is now with the development of computers the new technology have become available for more users.

Human - computer interaction knowledge will be more important as modalities increase.

Problems with VoIP for Wireless devices are that the overhead information has to be reduced. This is not always easy due to security and QoS reasons. QoS may be more difficult to obtain on wireless networks.

With integration of data communication into mobile devices, new market areas emerge. Several examples of these new markets are OS, PDA, Portals, transaction specific services, e-commerce, and multimedia applications.

After the prototype has been implemented further work could be done to make a newer and better application. As of now the prototype has some limitations and even if I can identify the limitations they might be difficult to solve.

Future work can be done to make a more complex application where these functions should be supported:

1. Search the whole database of entries. For now to minimize the result list, the search is only done within Stockholm.

   Solution: It is easy to expand the search, it is only to switch a parameter. The problem is not in the software, but in the hardware. The mobile clients has limits in cache, with further development of mobile clients and increased bandwidth this will not be a problem. Remember that, long lists are not ideal to use in a IVR application.

2. Search the database with more criteria. In the IVR application the user can only search by given name and last name. On the other hand the WAP application more criteria can be given (street, street number), but not in the IVR application

   Solution: Letting the user say all these inputs after each other may be easy to implement, as it is only to expand the IVR application to take more criteria.

3. Browsable list. If using the barge in feature the user can browse the TTS result list by using commands like, “next”, “previous”, “pause” and more.

   Solution: The feature in the prototype where the user can say “call” and transfer to the actual number is an example of this kind of a feature.

4. Read results from WAP. The WAP application should initiate a call when the results are returned.

   Solution: This can be done by sending the information to the IVR application from the WAP servlet. The IVR application initiates the call procedure to the user and reads the result. Another implementation of this could be to add a “read” link at the WAP result page, where the user self initiate a call to get the result read by the TTS engine. This should be fairly easy to implement if the NUANCE API supports initiation of telephony calls.

5. PUSH enabled service. The user should have the possibility to begin using the multimodal interaction with the IVR application. As for now the user has to begin the interaction with WAP.

   Solution: Wait for PUSH technology development. When PUSH is enabled this is an easy feature to implement. Other ways to implement PUSH is to let the user do all the PUSH features. A manual PUSH service.

However as we can recollect from (Bruce Balantine et. al., ´98) it is important to decrease the cognitive load on the user and the more features implemented the more complex the application will be. It is important for the designer of multimodal application to fully understand the differences in design in different modalities.

What might seem obvious in a WAP application does not automatically imply the same in an IVR application.

7.2 Guidelines for multimodal design

There are several points that should be considered when integrating several modalities into one single interface. Mobile devices do indeed have many limitations, but also many possibilities. With small devices and limitations of mobile devices, HCI guidelines for implementing multimodal interface are important. Design for multimodal interface for mobile devices should be:

1. Simple. The interface should be low in hierarchy.
2. Low network and device load. No complex images or unnecessary information should be included. Both network bandwidth and client computer power is poor.
3. Each modality has it own user interface. It is important to understand the HCI aspects of each modality.
4. HCI aspects of the integrated interface. In accordance with (3), the resulting interface has to be considered. The designer must remember that the multimodal interface may not be usable if only the modalities are considered individually.
5. Synchronization where? The different modalities have to be synchronized somewhere. All modalities have to now the current status of the interaction.
6. Consistency. The different modalities should mirror the same interface. If different modalities mirror different functions, the user will be confused.
7. Switching of a modality. How and when should the user be able to switch modality?
8. Scenario. Where should a specific modality be used? This is associated with point 7, but the designer should consider the environments the service will use.
9. Feedback. As in section 2.2.3, it is important for the user to know what the current system status is. It can however also decrease the usability in the interface.
10. For real multimodal design, I encourage the reader to read this thesis thoroughly and browse the references.

7.3 Multimodal future

A future use of multimodal systems is at least a tempting area to discuss. Some of these applications or systems does exist and when exploited will gain new grounds for multimodal technology. Multimodal systems are especially interesting in autonomous systems (see figure 21), VR applications (see figure 22), surgery and complex developing systems (CAD, Game industry).

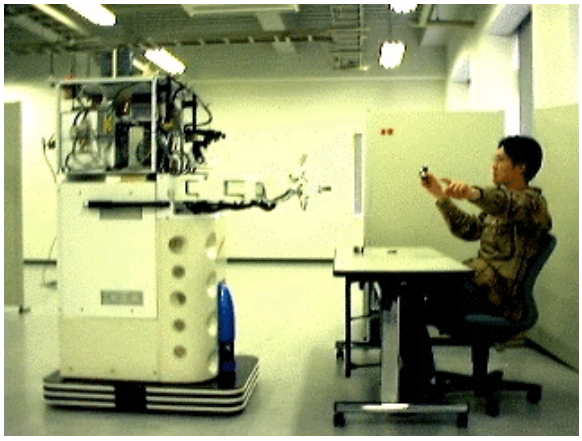

Figure 21. Autonomous system control, by voice and gesture. (Picture taken from www.transit-port.net)

The definition of multimodality could be widened to include more than human computer interaction (HCI), but also to human - machine interaction (HMI), human – human interaction (HHI) and machine-machine interaction (MMI). The modalities can thus be expanded and the limitation of the five human sensatory organs can be abandoned. This is not included in this thesis, but the area of multimodality is large, we should keep that in mind.

Figure 22. VR – Cube, visual and gesture interaction. (Picture taken from of www.tan.de)

8 References

8.1 Human - computer interaction

BALANTINE, BRUCE AND MORGAN, DAVID P. (1999) *How to build a speech recognition application.* (Published enterprise integration group)

BELL, L., BOYE, J, GUSTAFSON, J AND WIRÉN, M. (2000) *Modality convergence in a multimodal dialogue system.* Proceedings of Götalog 2000, Fourth Workshop on the Semantics and Pragmatics of Dialogue, pages 29-34.

BELL, L., EKLUND, R. AND GUSTAFSON J. (2000) *A comparison of disfluency distribution in a unimodal and a multimodal speech interface,* In Proceedings of ICSLP 2000.

BOLT, RICHARD A. *Put-that-there*. (1980) SIGGRAPH ´80 Conference Proceedings. (http://www.acm.org)

Bonharme, Eric. *Usability evaluation techniques* (2001), (http://www.dcs.napier.ac.uk/marble/Usability/Evaluation.html)

COHEN, P. R. & OVIATT, S. L. (1995). *The role of voice input for human-machine communication*, Proceedings of the National Academy of Sciences, 92 (22) 9921-9927. (http://www.cse.ogi.edu/CHCC/Publications/text.html)

GRASSO, MICHAEL, EBERT, DAVID S. AND FININ, TIMOTHY. (1997) *The Integrality of speech in mulimodal interfaces.* (http://www.csee.umbc.edu/~mikeg/papers/report03.html)

HEILIG, MORTON L. (1955) *El cine del futuro,* Espacios. p 23-24.

JEFFERSON, STEVE. (2000). *Mobile computing advances on reality*. (http://www.infoworld.com/articles/eu/xml/00/09/25/000925eumobile.xml)

JOHNSON, CHRIS. (1998) *First workshop on human - computer interaction with mobile devices.* GIST Technical Report G98-1. (http://www.dcs.gla.ac.uk/~johnson/papers/mobile/HCIMD1.html)

KLEINLÜTZUM, JAN, MERSCH, HENNING (2000/2001). *Multimodale mensch-maschine kommunikation seminar (Hauptstudium) im wintersemester.* (http://www.techfak.uni-bielefeld.de/ags/wbski/lehre/digiSA/MMK-Seminar/)

MIAMI, SCHOMAKER, L., NIJTMANS (NICI), J., CAMURRI, A., LAVAGETTO, F., MORASSO (DIST), P., BENOÎT, C., GUIARD-MARIGNY, T., LE GOFF,,B., ROBERT-RIBES, J., ADJOUDANI (ICP), A., DEFÉE (RIIT), I., MÜNCH (UKA), S., HARTUNG, K., BLAUERT (RUB), J. (1995) *A taxonomy of multimodal interaction in the human information processing system.* (http://hwr.nici.kun.nl/~miami/taxonomy/taxonomy.html)

NIELSEN, JAKOB AND MACK, ROBERT L., (1994). *How to conduct a heuristic evaluation.* (http://www.useit.com/papers/heuristic/heuristic_evaluation.html)

OVIATT, S. L. & OLSEN, E. (1994). *Integration themes in multimodal human-computer interaction.* In Shirai, K., Furui, S. & Kakehi (Eds.) Proceedings of the International Conference on Spoken Language Processing, 2 551-554. (http://www.cse.ogi.edu/CHCC/Publications/text.html)

OVIATT, S. L. (1996). *User - centered modeling for spoken language and multimodal interfaces.* In *IEEE Multimedia, 3* (4) 26-35. (To be reprinted in Morgan-Kaufmann *Readings on Intelligent User Interfaces*, ed. by M. Maybury & W. Wahlster). (http://www.cse.ogi.edu/CHCC/Publications/text.html)

SHNEIDERMAN, BEN. (2000) *Designing the user interface*, 3rd edition, pages 328-333.

SIEGEL, JANE, KRAUT, ROBERT E., JOHN, BONNIE E. AND CARLEY, CATHELENE. (1995) *An empirical study of collaborative wearable computer systems.* (http://www.acm.org/sigchi/chi95/proceedings/shortppr/js_bdy.htm)

WHALSTER, WOLFGANG (1991). *User and discourse models for multimodal communication, Intelligent User Interfaces.* , pages 45-67. (http://www.acm.org)

WORLD WIDE CONSORTIUM (2000), Workshop on multimodal requirements for mobile devices. (http://www.w3.org /TR/multimodal-reqs)

For further reading, www.acm.org has several articles and publications regarding HCI and multimodal interaction.

8.2 Technical / programming

ERICSSON, (http://www.ericsson.se)

8.2.1 Java SUN

Java API specification, (http://www.java.sun.com)

Java 2 micro edition Specification (http://java.sun.com/j2me/?frontpage-javaplatform)

8.2.2 SIM toolkit

Toolkit API Specification (http://www.cellular.co.za/sim_toolkit.htm)

8.2.3 SIN/RFC[4], WAP/WML and PUSH

THORZIDE, *PUSH technology* (http://www.thorzide.de)

WAP-195, *Wireless application environment overview* WAP-195-WAEOverview-20000329-a.pdf

WAP-195_101, *Wireless application environment overview SIN* WAP-195_101-WAEOverview-20000329-a.pdf

[4] *All SIN/RFC can be downloaded at http://www.wapforum.org.*

WAP-190, *Wireless application environment specification* WAP-190-WAESpec-20000329-a.pdf

WAP-191, *Wireless markup language specification* WAP-191-WML-20000219-a.pdf

WAP-170, *Wireless telephony application interface specification* WAP-170-WTAI-20000707-a.pdf

WAP-200, *Wireless datagram protocol*, WAP-200-WDP-20000219-a.pdf

WAPFORUM, (http://www.wapforum.org)

8.2.4 VoiceXML

WORLD WIDE CONSORTIUM, *VoiceXML specification 1.0*, (www.w3c.org/voice)

NUANCE VOICEXML SYSTEM, *Introduction to the Nuance system v7.0,* (http://www.nuance.com)

Appendix A, Glossary

3G – 3rd Generation mobile communication systems

API - Application Programmer Interface

CSD - Circuit Switched Data

EKT - Elektroniska TelefonKatalogen

ETSI - European Telecommunications Standard Institute

GPRS – General Packet Radio Service

HCI – Human Computer Interaction

IVR – Interactive Voice Recognition

JSP – Java Server Pages

OTA – Over The Air

PDA – Personal Digital Assistants

PIN – Personal Identification Number

QoS – Quality of Service

S/N – Signal to Noise

SDK – Software Development Kit

SIM - Subscriber Identity Module

SIP – Session Initiation Protocol

SPAR - Statliga PersonAdressRegistret

SMS – Short Message Service

SMSC – Short Message Service Center

SO - SpeechObject

TTS – Text To Speech

USSD - Unstructured Supplementary Services Data

VoIP – Voice over IP

VR – Virtual Reality

WAE - Wireless Application Environment

WAP – Wireless Application Protocol

WML - Wireless Markup Language

WTA - Wireless Telephony Application

WTAI - Wireless Telephony Application Interface

XML – eXtensible Markup Language

Appendix B, Screendumps from teleplus, multimodal version

Unfortunately to make screendumps of the WAP interface, an Ericsson 380 emulator have been used. There are no Ericsson r520 emulators available.

Figure 1. This is the splash screen, which welcomes the user and presents the service. It will be show for 10 seconds, depending on the connection speed.

Figure 2. The user can use the browser to see the main menu i.e. consistent throughout the service. Here the link “voice” can be pressed to start the IVR application. To see the results from the IVR application as a WAP page, the user has to press the link “Resultat”.

Figure 3. This is how text input is made, could be either a soft keyboard or written recognition.

Figure 4. This is what typical search criteria would look like. In a future application this could be expanded to include several more criteria. All the input fields do not have to be filled to perform a search.

Figure 5. The search returned an error message, that no one in Stockholm called Niklas Becker at Sveavägen 132 is registered.

Figure 6. After the result is returned, the help link is pressed to receive additional help. The message says that you should use either the keyboard or click the voice link to input data.

Figure 7. A new input is made to test the application.

Figure 8. The result (5) returned by the new search. It begins with number telling the order, last name, given name, area and a link that could be pressed to initiate a call to the specific person.

Appendix C, Corporate profiles

Pipebeach

Pipebeach is an IVR application design company. Built on VoiceXML technology, several applications are developed. Pipebeach have implemented the product SpeechWeb, which is a platform compatible with several different speech recognition technologies. Founder of Pipebeach is Scott McGlashan a W3C member of voice technology. Pipebeach does have some ideas on how multimodal systems architecture would look like but nothing implemented (what I know) and for further information read about Pipebeach at:

(http://www.pipebeach.com)

Catch2004

CATCH 2004 is a research project funded by the European Commision in the scope of the IST programme. The goal of this research activity is to develop a multilingual, conversational system with a novel unifying architecture across devices and services. The system will provide pervasive access to multiple applications and sources of information available to citizens from public and private service providers by supporting multiple client devices, and by using multiple input modalities. Client devices are kiosks, telephones (standard and wireless) and smart wireless devices. Applications include access to information over the Internet, travel and city information/services, phone-directories and completion of transactions. Catch2004 includes companies as IBM, Nokia, ELISA. Further information at:

(http://www.catch2004.org)

IBM

The HCI research area at IBM is large, which is apart of their research area. IBM has developed a multimodal prototype for network management at Lockheed Martin Advanced Technology Center in Sunnyvale, California. The prototype includes speech I/O and 3D visualisation. IBM research are trying to combine modalities in user interfaces include modalities as visual, auditative and tactile. Further information at:

(http://www.research.ibm.com/compsci/hci/)

AT&T

AT&T Labs have ambition to develop new multimodal services including voice, image and video processing. AT&T has developed multimodal telephone systems (VoiceTone) for more flexibility and TTS agents. The vision is to deliver these systems in “not-too-distant future” (this remarkable statement was made 1999 and has not been updated). Further information at:

(http://www.att.com)

Philips

Philips research is developing multimodal systems to gain competitive edge and user benefits. A Prototype exists with user interfaces that integrate speech, 3D graphics and touch as modalities. Further information regarding several prototypes see:

(http://www.research.philips.com/pressmedia/highlights/index.html)

Microsoft / SALT

Cisco, Comverse, Intel, Microsoft, Philips and SpeechWorks founded Speech Application Language Tags (SALT) Forum to develop a new standard for multimodal and telephony-enabled applications and services. The SALT forum develop speech tags for mobile and stationary computers, which will make it easier to develop multimodal interfaces. SALT will be a new markup language specified in XML.

(http://www.saltforum.org/)

Attachment A, Questionnaire 1

Frågeformulär 1

En multimodal applikation är ett program som man kan kommunicera på olika sätt med, Till exempel genom röst, text, rörelse, accelometrar, direktmanipulation (peka på skärmen). Programmet tolkar dessa former av kommunikation och utför det den uppfattat. Vänligen markera det svarsalternativ som stämmer bäst genom att ringa in eller kryssa för.
Ställ gärna frågor om något är oklart!

Ålder

<20 21 – 30 31 – 40 41 – 50 51 – 60 >60

Kön

Man Kvinna

Använder dator

Inte alls <1ggr/vecka 3 – 4 ggr/vecka 1 - 2 tim/dag 3 - 4 tim/dag 5 - 6 tim/dag > 7 tim/dag

Använder WAP

inte alls <1ggr/vecka 3 - 4 ggr/vecka 1 – 2 tim/dag 3 - 4 tim/dag >5tim/dag

Varför?

Brukar ny teknik innebära en förbättring eller försämring?

(Sätt ett kryss på linjen som stämmer överens med din uppfattning)

Försämring | _____ | Förbättring

Är multimodalitet bra?

_____

_____

Vad kan det finnas för användningområden för multimodala applikationer?

_____

_____

_____

_____

_____

Vad skulle öka din användning av mobila tjänster?

(Sätt ett kryss på linjen som stämmer överens med din uppfattning)

Bättre, Hårdvara | _____ | Tjänster

Intresse, lite | _____ | Mycket

Användarvänligt, dålig | _____ | Bra

Beror din användning av mobila tjänster av något helt annat?

_____

_____

Attachment B, Questionnaire 2

Frågeformulär 2

Tror du på en mer användarvänlig interaktion med multimodala applikationer?

Är multimodalitet bra (Nya synpunkter)?

Vad kan det finnas för användningområden för multimodala applikationer (Nya synpunkter)?

Skulle det vara bra att ha multimodala gränssnitt på fasta datorer? För- nackdelar?

Var tjänsten ett bra val att implementera en prototyp? Varför?

Tjänsten (inte själva gränssnittet) ger jag ett betyg

Dåligt 1 2 3 4 5 6 7 8 9 10 Bra

Var det lätt att använda tjänsten?

Svårt 1 2 3 4 5 6 7 8 9 10 Lätt

Framgår det vad man kan göra, dvs får man en överblick av tjänsten?

Dåligt 1 2 3 4 5 6 7 8 9 10 Bra

Skulle fler tjänster som denna öka din användning av WAP (mobilt Internet)?

_____

Övrigt?

_____

_____

_____

_____