Mobile Wireless Interfaces: In Search for the Limits

Jari Veijalainen* and Tom Gross+

* Dept. of CS&IS, FIN-40014 Univ. of Jyvaskyla, Finland
+Fraunhofer FIT, Schloss Birlinghoven, D-53754 Sankt Augustin, Germany
veijalainen@cs.jyu.fi
tom.gross@fit.fraunhofer.de

Abstract. With the proliferation of 2G and 3G Telecom and other wireless networks hundred of millions of users will be able to access wireless services with their terminals in only a few years. In this paper we discuss the possibilities and limitations of existing and emerging mobile technologies and methodologies for porting information and functionality from traditional PCs to mobile devices and vice versa. We argue that in order to be able to properly port things between devices an equivalence concept for user interfaces is required.

1 Introduction

Mobile wireless technologies have developed rapidly during the last ten years through the introduction of digital GSM networks. The so-called 3G networks will also be operational within the next two years all over the world offering faster data transfer (up to 2 Mbits/s) and a larger variety of services than the 2G networks. With the introduction of the Wireless Application Protocol (WAP) [15] and similar technologies in Japan (e.g., NTT DoCoMo's i-mode) [2] as well as infrastructures supporting them, wireless Internet access has become possible and real from Telecom handsets.

The development of the GSM and other digital handsets has indeed been fast. From mere voice terminals at the beginning of 1990s the most advanced models have developed to small wireless Personal Digital Assistants (PDA) that are now also capable of running applications and of communicating asynchronously and synchronously with servers. Both WAP terminals and i-mode terminals are now able to run micro-browsers that interpret markup languages such as WML, WMLScript, or cHMTL and even run Java applications [8]. The newest terminals are more than just telephones and can e.g., be used to conduct e-commerce transactions and authorisation of users in e-government transactions. It is envisioned that these devices are highly personal and carry even the owner's 'cyber identity'. Therefore, the Telecom industry has begun to call them Personal Trusted Devices (PTDs) [9].

The latest news give a strong signal that diverging paths of the markup languages used at WWW sites (HTML), as well as at WAP/i-mode sites (WML, cHTML) are converging towards the common ground, XHTML [14]. For instance, Nokia has already demonstrated its XHTML browser on its handsets and will license the source code to other parties. Nokia plans to include the browser into its handsets [13]. Abstracting from the differences between fixed Internet terminals and wireless ones, one can set up the thesis *that, in the future, the wireless Internet-ready PTDs can be used as user terminals just like the PCs.*

In this paper we will explore the limits of this (somehow provocative) statement. We start by looking at the diverse information processing needs of the users in an advanced computer-supported cooperative work (CSCW) context. A first version of an XML-based interface meta-description is sketched. We believe that in the future a standard user interface meta-description language should be applicable that is supported by the middleware. Based on these ideas we will discuss mappings between full-fledged PC environments and weaker mobile device environments. The application area we use as the example is advanced CSCW with 2D and 3D user interfaces.

2 Information Types in an Advanced CSCW Context

In order to look more closely into the requirements diverse application interfaces can pose, we take the interface facilities under development in the TOWER project [11, 12] at Fraunhofer FIT as a concrete example. The digital information handled and at least partially presented through digital 2D and 3D interfaces of this CSCW application can be coarsely divided into four categories:

- *Ephemeral communication*: Distributed group work often requires face-to-face meetings, audio conferences, and video conferences. The information exchanged is hardly recorded digitally, although in cases excluding face-to-face meetings most information is encoded and transferred in a digital format through IP and/or digital Telecom network. There is no system-supported context that could organise production and usage of this data.
- *Written communication*: Written communication such as Email or on-line chat can be seen as a part of the document base. However, from an implementation point of view it is supported separately and cannot necessarily be accessed and evaluated in the same manner as the actual documents.
- *Shared information*: The work a group of people is co-operatively performing materialises mainly in shared (multimedia) documents such as texts, drawings, graphics, images, video clips, and so forth. These are the core objects in cooperation and the storing of them into the workspace, modification in, retrieval and deletion from the workspace, along the version control, is supported by the core system [3].
- *Up-to-the-moment awareness information*: This information on who is doing what in which workspace, where the person is physically, and so forth. The awareness information here includes only the current information. It is presented using facilities offered by ambient interfaces such as plastic fishbowls, a fan, etc. as well as by 2D or 3D virtual worlds on the PC in which shared artefacts are represented as buildings and the users accessing the artefacts are represented as avatars stopping by at or entering the buildings.
- *Historical awareness information*: This information is closely related to the above awareness information in the sense that evaluating a history can be understood as creating awareness towards history, that is towards events in the past. Daily reports about the recent activities in a shared workspace that the members of a workspace receive via email are an example of this. In TOWER the 3D virtual world offers the users the possibility to replay scenes.

The information is presented through interface facilities that include a PC monitor, and loudspeakers, but also other fancy components for the awareness information (see below). The standard environment consists of desktop computers and a variety of ambient interfaces that use various physical gadgets (like Lego men, Lego vehicles, light sources, and balloons) to provide digital information and that are located in the office of a person. Meeting rooms are equipped with even more powerful interface devices (e.g., large flat screens on the wall, loudspeakers, WebCams, etc.). All these interfaces are stationary and can only be used in the office environment. So, one challenge lies in supporting a person who is roaming and does not have access to the stationary interface facilities in the office building. We are here not primarily interested in the other evident question—namely what part of the interface is reasonable to transfer to a mobile environment from a mobile user's point of view. This is largely a context-dependent issue [6].

3 Equivalence of Interfaces

It is immediately evident that the mobile user with only a portable wireless terminal at her disposal has fewer possibilities to acquire information than the user in the office (e.g., large documents, video clips, etc.) [5]. So, the question is how users with varying technical equipment can cooperate with each other.

3.1 What could equivalence of interfaces be?

How to tackle the problem of porting a stationary interface to mobile environment theoretically? We strive at different, but at the same time in some sense 'equivalent' interfaces. Seen a little bit more mathematically, the question is *when are two interfaces equivalent?* If we look at the concept of interface, it is the border between a human being and a (computer) system. A human being has all the senses—that is, vision, hearing, taste, smell, and, touch sense—at her disposal to receive information over the interface from the computer system (from a system's point of view, this is output). She can feed in information (from a system's point of view, input) using keyboards, pointing devices (together with GUIs), voice, and in some cases body movements (in front of video camera or in vicinity of other movement detectors), or body warmth (warm up a sensor and something happens).

The equivalence of TOWER interfaces A and B requires evidently that *the same information is mediated through both interface A and B.* This means that if a certain piece of information can be output through interface A it can also be output through interface B and vice versa. The same holds for the input. For instance, if a colour display is used as part of interface A and B then this mapping is easy. Both use vision as the sense to transmit the information such as picture or video. Whatever the actual information content is, it can be transmitted to the user through both interfaces, provided both support the same data types (video, image). One can at most argue, whether from the user's point of view the information contained in a video presented on the screen with size 50mm*50mm is indeed the same as if presented on a larger screen with 250mm*250mm.

The situation is similar with other output streams (e.g., audio) that are received by the same sense (hearing) at different interfaces A and B. If on the one hand a user listens to news using a next generation Telecom (mono) handset with an ear plug, or on the other hand to a car radio with eight loudspeakers, the perception is different, but the user still gets the same information from the news. However, listening to stereo music is not the same, because 3D-space perception is missing.

Although the comparison between different interfaces relying on the same sense is already quite complicated, it is yet much more complicated if the user's sense or body property used for interface A is not the same as used for interface B. Nevertheless, equivalency between interfaces should also be possible if different senses and different types of physical interaction (writing on keyboard, speaking, or pointing) are used. In many cases the output information can be mediated through different senses at A and B. This is also true for input, but there are less possibilities. On the output side, for instance, text can be both shown on the screen and read (vision) or spoken through loudspeakers or other similar devices and listened to (hearing). And similarly, speech (voice stream) can be transformed into text and be shown on the screen and the information carried with it can be perceived by the vision. However, there are some absolute barriers in inter-sense transformations. Video streams cannot reasonably be encoded into a form so that by 'listening' to them the same information can be mediated as by watching them. The same holds for music; it is impossible to represent a piece of music so that just by looking at the presentation on the screen, whatever it is, the user would be able to hear the actual music (well, some professional musicians might still be able to do it, if the stream of corresponding notes is displayed to them on the screen instead of playing the stream itself). Images, on the other hand, could be represented as spoken descriptions.

Looking at these examples, it seems to be a reasonable condition for the equivalence of interfaces A and B that *both the type and the amount of information exchanged over them* is the same. The definition would actually abstract from other aspects. Aspects such as usability, look-and-feel, and so forth are highly important from a pragmatic point of view, but the information mediated through two interfaces should be the same, otherwise they cannot be equivalent. This informational equivalence understood as capability of mediating the same information with respect to both type and amount is thus a necessary, but not a sufficient, condition for the interfaces to be equivalent.

Should we really measure the amount of information in each case or can we find an easier approach? There is the problem that for some users a picture, a video stream or other artefacts do not mean anything, for some others they mean thing C, for yet some others they mean thing U. Thus, the amount of information mediated through an interface is not the same among people, although they get the same data output (e.g., watch the same video simultaneously). So, when defining the equivalence of two interfaces in the above sense, we should not require that each individual person acquires or exhibits the same amount of information through both of them. Rather, we must require that any particular person gets the same information output and inputs the same information through A and B. Thus, one can postulate:

User-centred informational equivalence of UIs: Assume that through an interface IF_i the user gets (inputs) I(i) units of information of type X while interacting through it T(i) seconds. IF_j and IF_i are *equivalent* in output (input) sense, iff the user gets (inputs) over IF_j and IF_i the same piece of information (I(i) = (I(j)) during any two equally long time intervals (T(i) = T(j)). IF_j and IF_i are *weakly equivalent*, iff for any T(i) there is T(j) such that I(i) = I(j). Interfaces are *strongly equivalent* if they are equivalent and the user considers them to be equally usable in any thinkable sense.

The above definition tries to go around the question, what is information actually for a user and how its quantity is measured. It also abstracts from different components used such as sound, video or Lego men. It postulates that whatever the user considers as piece of information, strongly equivalent interfaces are able to provide it to her in the same time (or she is able to provide the information to the system) and with the same look-and-feel. Equivalent interfaces do this in the same time, but maybe with a different look-and-feel. Weakly equivalent interfaces are able to output/absorb the same information, but it may take more or less time to get or input the information and also in other respects the look-and-feel differs.

It is worth noting here that the equivalence above is dependent on certain threshold values that vary form case to case, but are physiologically anchored (like perceiving a stream of still images as continuos movie). We do not go further into details in this context.

3.2 Encoding of the information within the system

We have seen above that the same information can be represented using different senses and thus different output and input devices as part of the interface. Within a digital system, all information and algorithms are encoded in finite bit streams. These are then transformed by the output devices in analogue forms (pictures, videos, music, speech, and text with Latin alphabet) that can be perceived by human senses. Correspondingly, human interaction with input devices (e.g., keyboard, touch screen, GUI with pointing device) are transformed into bit sequences and interpreted correspondingly. Thus, we could give a measure for the information content by measuring the length of the bit streams needed to produce the output or produced by input. Intuitively, the more bits are needed to encode the information, the more information is mediated through the interface.

But the bit string length or processing speed is not a very good measure for the amount of information. For instance, a written picture description might require 548 bytes of disk space. The speech generator would use, say 30 seconds to read this text aloud. Assuming that it would encode the text in phone quality, each second would require 64 kbits data, that is the overall encoding of the text would be 1920 kbits long. And the size of the actual graphics would vary of a few kB to maybe tens of kB, depending on the tool we are using to draw it.

Although the number of bits is not an exact measure for the amount of information, it still gives a hint of the correct magnitude. For instance, video contains more information than pure audio and, therefore, video files are usually much larger than audio files that play the same time. The magnitude of length of the encoding is, however, not fixed, because the encoding data can be compressed down to, say, one per cent of the uncompressed size using either compression with or without loss. For the wireless data transfer compression of the data is often of high importance [10].

3.3 Transformations of interfaces

The equivalence as seen from the users' point of view and discussed above is a crucial criterion. Measuring the bit stream that is passed through an interface in a time unit, e.g., in bits/s, gives important information about the technical requirements of a system. In particular, this bit stream has to be transmitted from the server to the client or vice versa on time. The programs decoding the data onto the output devices must be able to process the data on time and must run in the terminal. *Therefore, the digital encoding of the* *information (program, data) can be used as a technical measure for the space and time requirements.*

One can now look at the technical level, what transformation of interface means. From a system's point of view, transforming information from one interface to another primarily means that the *information encoding in data, and programs interpreting it, must be transformed accordingly, preserving the equivalence of the interface as discussed above.*

The first evident idea is that we simply replicate the software (e.g., image viewers, VRML tools, audio tools and the data they use in the PC environment) on the mobile device and let it run with the same data. If this works, we have most probably been successful in replicating that part of the interface in an equivalent manner into the mobile environment. We use the same encoding of the data, and programs interpreting it, and one can easily argue that whatever information is buried for a user in the encoding, she would be able to receive/exhibit it in the environments. But what if it does not work? And how about the other parts of an ambient interface? Can they at all be represented in the mobile environment?

What we need here is to view the interfaces to be *composed of components,* IFi_1, $IFi_2,\ldots IFi_k$. And these components must be mapped on other components $IFj_1\ldots,IFj_h$ on the other device. The components must be modelled case-by-case. For instance, mapping from a computer monitor to a display of a mobile device includes mapping image data from the monitor to the same or smaller image data on the display; software component to a replicated software component or to a different component with equivalent functionality; and data to identical data (e.g., image, video stream) or data with different encoding but the same information contents (GIF to JPEG); and so forth. In addition, we need to capture the possible temporal relationships between the components in (IFi_s, IFi_k) and map them in a coherent way to the temporal relationships between interface components (IFj_y, IFj_z), and so forth. This is especially important for video and audio streams that must be presented with a certain speed. Here we see immediately that not only the interface capabilities of the terminal but also the transfer capacity of the (wireless) network are of importance. If it is not able to transfer the stream fast enough, there is no hope of seeing the video in the same way as in the standard environment. Thus, the network capabilities must also be captured somewhere.

Let us revisit the TOWER example: In the 3D multi-user TOWER world houses represent workspaces and avatars represent users working in the workspaces. Assume that we try to present this world on a WAP phone display. It is hardly possible to find a mapping between the computer environment and most of the currently existing WAP terminals, because there is no colour display and the display is too tiny to make the avatars visible. Also, the visualisation software components at the terminal do not have the functionality needed to represent the virtual world.

The following table provides a mapping between the different types of information from section 2 and makes suggestions as to which type of mobile device is suited. The mobile devices are divided into four categories. In the ascending order of their capabilities they are: usual GSM phones with voice and SMS support, smart phones with a WAP or i-mode browser, PDAs/Communicators with TCP/IP communication capabilities and WWW-browsers and Java runtime environment, and finally notebooks.

Table 1. Mapping information in TOWER between a PC and mobile devices.

Type of information	**TOWER features**	**Minimal terminal type required**	**Data encoding**
Ephemeral communication	Spontaneous chats in 3D TOWER world	Communicator, Notebook	2D/3D colour interface
Written communication	Email	Basic GSM voice terminal with SMS capabilities for ASCII emails; PDAs/Communicators for multimedia emails	Character display, no stringent timing requirements
Up-to-the-moment awareness	3D representation of shared documents and their evolution	Notebook	Pictures, 3D colour interface
Up-to-the-moment awareness	SMS notification about user activities and document changes	Basic GSM voice terminal with SMS capabilities	Character display, no stringent timing requirements
Up-to-the-moment awareness	WAP queries for user activities and document changes	Smart phones with a WAP browser	Display for formatted characters (incl. hyperlinks, tables) and images
Historical awareness	3D aggregated presentation of past events in TOWER	Notebook, 3G application-enabled handset	Video stream, 3D colour interface

3.4 Representation of the user-interface in XML

The latter thoughts raise one further question. If the interfaces should be mapped into the mobile channels, there should be a meta-description of the semantics of the interface components and metaphors. These semantics have different *encoding* in the system: in various virtual worlds, in textual descriptions, and so forth.

The meta-description would also help a new user to check what information different components of the interface convey. Let us take an example: avatars represent users; houses represent tower workspaces, folders, Web sites, etc.; the colour of the house represents the creator of the document; the height of the house represents the number of accesses to the document; etc. In practice, the meta-description should be formalised; e.g., in a form of XML DTD. The XML descriptions could then be used by both computers and humans to interpret the interface in the correct way.

The above considerations also raise the question of minimal encoding of the interface components. Such an encoding could be presented as part of the meta data. As an example, the person in the cyber space can be represented by an avatar or simply by an ID. For the representation it is enough that n different identities are represented if there are n people. Thus, minimal encoding of a person would be about log n bits. So very coarsely the interface description should be something like this:

```
<IF name>

<human_readable_descr>
<component_descr> An avatar symbolises an acting user; a building
  representes a document. Reading is symbolised as an avatar’s head
  movement in front of the respective building.
</component_descr>
<component_descr> 2D interface consists of a 2D graphical representation
  of the 3D world. Coloured circles symbolise users; rectangles
  represent documents. Reading is symbolised by a little image with
  reading glasses superimposed on the user’s circle.
</component_descr>
<component_descr> A Lego man represents… When it raises its hand,…
</component_descr>
</human_readable_descr>

<mappings_to_dev>
<mapping_to_Win>
  <human_dev_descr> An avatar is here mapped to avatar…
  </human_dev_descr>.
  <data_files>F1,F2,…</data_files>
  <progr>P1,P2,..</prog>
</mapping_to_Win>
<mapping_to_9210>
  <human_dev_descr> Avatars are not modelled here. …
</human_dev_descr>
<data_files>F1,F2,..</data_files>
<progr>P1,P2,..</prog>
</mapping_to_9210>
<mapping_to_R380>…
</mapping_to_R380>
</mappings_to_dev>

</IF>
```

The interface description contains a common human readable part that explains the artefacts and the meaning of their symbolic action. The interface is then ported to different devices. In the description it is explained which artefacts of the general interface are modelled and how the mapping is done. For each interface the necessary data and programs are referenced. The idea is that the technical description should be so complete that by dropping the programs and data onto the device either permanently or over network connection (applet-based approach) the interface would run.

As said earlier, the voice terminals also offer the possibility to mediate speech both in video and audio conferences and also as guidance. Synthetic speech could be used as part of the interface. In the most advanced case, the user interface is multimodal; that is, it can use graphics, voice, and text at the same time. The current terminals already allow simultaneous voice connection and running of applications such as a calendar application.

The above mapping is not only interesting and relevant from a hardware and software resources’ point of view, but also from a user’s point of view. In fact, one major challenge of groupware systems providing information and communication facilities as well as awareness information is the attention span and the concentration of the user. That is, these systems should provide functionality and information that is adequate to the tool on which it is presented and adequate for the user. Adequacy for the user means that the capabilities of the respective terminal type are optimally used, but that they should not be overused. Likewise, the usability and usefulness of the functionality and information provided strongly depends on the user's current situation. The system should therefore also take the user's context into account [5].

The ideas above must be refined and are for further study. The wireless world is now going towards XHTML [14] and Java as a unifying framework to describe Web and WAP interfaces. This solves in part the problems, but some problems such as support for the avatars and virtual worlds in mobile devices remain. When handsets begin to support Java the interfaces again become more complicated and should be describable by a meta-description. This would help both porting the interface between different devices and helping the user to understand the semantics of the interface. The Web Services Description Language (WSDL) is an XML-based language to define Web services and to describe how they can be accessed [1]. While it provides a nice syntax for describing services, its main limitations are that it only describes Web-based services and it only deals with syntactic rather than semantic questions. Other approaches and tools for transcoding—that is, translating functionality and data between platforms—such as the eXtensible Stylesheet Language for Transformations (XSLT) [4] or IBM WebSphere [7] offer nice syntactical translations, but again do not deal with the important question of semantic equivalency. Furthermore—despite the promise of standardised markup and other languages for Web and WAP—we do not believe that all differences between handsets would vanish. Thus, somewhere the servers must keep the knowledge about differences.

4 Conclusions

We have discussed the possibilities of mapping information and functionality between devices with diverse hardware and software capabilities. We have presented an equivalence concept for user interfaces. As basis for the equivalence of two interfaces we proposed that the same information—from a particular user's point of view—could be transmitted through them with the same speed. Based on these ideas, some mappings between a full-fledged office environment and weaker mobile devices were discussed.

In this paper we omitted a discussion of contexts. It seems, however, that a system supporting multiple-channel access to a TOWER-like advanced CSCW environment must be aware both of the virtual context (i.e., the current situation and state in the 3D workspace, etc.), the terminal context (i.e., the current environment such as mobile, car, home, office), as well as the physical context (actual physical place), and of channel context. The latter is needed in order to determine what types of data can be sent to the user over the network (bandwidth, end-to-end delivery guarantees). Thus, the user is actually at the same time in different contexts (in a workspace (virtual context), at home (physical location), using wireless device (device context) over GPRS (channel context)) and the server should behave accordingly. The recording and combining of these contexts and the adaptation of the user interfaces components accordingly are for further study.

Acknowledgements. The work of the first author was mainly performed while he was visiting Fraunhofer FIT in Sankt Augustin, Germany, during 2000-2001. The economic and other support of FIT and especially of the TOWER project – EU/ IST-10846 – are highly appreciated.

References

1. Christensen, E., Curbera, F., Meredith, G. and Weerawarana, S. Web Service Definition Language (WSDL). W3C, http://www.w3.org/TR/wsdl, 2001. (Accessed 18/4/2002).
2. Devine, A. and Holmqvist, S. Mobile Internet Content Providers and Their Business Models. Master's thesis, Department of Industrial Engineering and Management, Royal Institute of Technology, Sep. 2001.
3. FIT. BSCW Home Page. Fraunhofer FIT and OrbiTeam, http://bscw.gmd.de, 2002. (Accessed 17/1/2002).
4. Global, X. XSLT.com. http://www.xslt.com/, 2002. (Accessed 18/4/2002).
5. Gross, T. and Koch, T. Ubiquitous Computing: New Challenges for the Design of User Interfaces of Mobile Devices. In IASTED International Symposia Applied Informatics - AI 2001 (Feb. 19-22, Innsbruck, Austria). ACTA Press, Calgary, Canada, 2001. pp. 549-554.
6. Gross, T. and Specht, M. Awareness in Context-Aware Information Systems. In Mensch & Computer - 1. Fachuebergreifende Konferenz - M&C 01 (Mar. 5-8, Bad Honnef, Germany). Teubner, 2001. pp. 173-182.
7. IBM. IBM WebSphere. http://www-3.ibm.com/software/info1/websphere/index.jsp, 2002. (Accessed 18/4/2002).
8. Japan Inc. Convergence Emergence. J@pan Inc., http://www.japaninc.net/mag/comp/2001/05/may01_converge2.html, 2001. (Accessed 20/8/2001).
9. Mobile electronic Transactions Initiative. MeT - Mobile electronic Transactions. http://www.mobiletransaction.org, 2001. (Accessed 20/8/2001).
10. Ojanen, E. and Veijalainen. Compressibility of WML and WMLScript Byte Code: Initial Results. In Proceedings of the Tenth IEEE Workshop on Research Issues in Data Engineering - RIDE 2000 (San Diego, CA). IEEE Computer Society Press, Los Alamitos, 2000. pp. 56-62.
11. Prinz, W. TOWER: Theatre of Work Enabling Relationships. http://www.fit.fraunhofer.de/projects/tower/, 2001. (Accessed 10/12/2001).
12. Prinz, W., Graether, W., Gross, T., Klein, K.-H., Kolvenbach, S. and Schaefer, L. Präsentation von Aktivitäten bei verteilter Zusammenarbeit. In Mensch & Computer - 2. Fachuebergreifende Konferenz -- Vom interaktiven Werkzeug zu kooperativen Arbeits- und Lernwelten - M&C 02 (Sep. 2-5, Hamburg, Germany). (accepted).
13. WAP Forum. Nokia Demonstrates First XHTML Microbrowser on Standard Mobile Phone. http://www.wapforum.org/new/20010321319Nok.htm, 2001. (Accessed 20/8/2001).
14. WAP Forum. Nokia, Motorola, Ericsson and Siemens to Further Advance the Development of Personal Mobile Services Through XHTML. http://www.wapforum.org/new/20010321318Nok.htm, 2001. (Accessed 20/8/2001).
15. WAP Forum. Wireless Application Protocol Forum Ltd. http://www.wapforum.org/, 2001. (Accessed 4/9/2001).