User-System Interaction: Trends and Future Directions

Prof. Dr. Matthias Rauterberg
IPO—Center for User-System Interaction
Eindhoven University of Technology (EUT)
Den Dolech 2, NL-5600 MB Eindhoven
Email: g.w.m.rauterberg@tue.nl

Introduction

The embodiment of computers in the work place has had a tremendous impact on the field of user-system interaction. Mice and graphic displays are everywhere, the desktop workstations define the frontier between the computer world and the real world. We spend a lot of time and energy to transfer information between those two worlds. This could be reduced by better integrating the virtual world of the computer with the real world of the user. The most promising approach to this integration is *Augmented Reality* (AR) [29]. The expected success of this approach lies in its ability to build on fundamental human skills: namely, to interact with real world subjects and objects!

We present a new interaction style, called Natural User Interface (NUI). Several prototypical applications already implement NUIs and demonstrate its technical feasibility.

Today several dialogue techniques are developed and in usage. The following dialogue techniques and dialogue objects can be distinguished with regard to traditional user interfaces: command language, function key, menu selection, icon, and window [23]. These techniques can be summarized into three different *interaction styles:*

- *Command language:* This interaction style (including action codes and softkeys) is one of the oldest way of interacting with a computer. *Pros*: In the command mode the user has a maximum of direct access to all available functions and operations. *Cons*: The user has no permanent feedback of all actual available function points.
- *Menu selection:* This includes rigid menu structures, pop-up and pull-down menus, form fill-in, etc. This interaction style became technically possible only with those terminals which, essentially, can reproduce only the ASCII character set. With this type of interaction style function keys are often used in addition to manage the dialogue. Pros: All available functions are represented by visible interaction points. Cons: Finding a function point in deeper menu hierarchies is cumbersome.
- *Direct manipulation*: This type of interaction style did not spread until bit mapped graphic displays came on the market; the development of this interaction style was based on the desktop metaphor which assumes that by depicting the work environment (i.e. of the desk: files, waste-paper basket, etc.) as realistically as possible on the user interface, it would be particularly easy for the user to adjust to the virtual world of electronic objects. Pros: All functions are continuously represented by visible interaction points (e.g. mouse sensitive areas, etc.). The activation of intended functions can be achieved by directly pointing to their visible representations. Cons: Direct manipulation interfaces have difficulty handling variables, or distinguishing the depiction of an individual element from a representation of a set or class of elements.

In all these traditional interaction styles the user can not mix real world objects with virtual objects in the *same* interface space. They also do not take into considerations the enormous potential of human hands to interact with real and virtual world objects. This aspect was one of the basic ideas to develop data gloves and data suits for interactions in a virtual reality system (VR). The other basic idea, to realize a VR system, was the 3D output capabilities in the usage of head mounted displays. However, in VR systems several serious problems are inherently present; they are:

- The lack of tactile and touch information and consequently the mismatch with the proprioceptive feedback.
- The lack of information for depth perception, since visual displays generate 2D data. Many informational concepts are generating possibilities to reconstruct 3D pictures from these 2D data.
- There is always a time delay in the user-computer control loop, which may yield severe problems with reference to the perceptual stability of the vestibular apparatus in the ear.
- The strong influence of continuous communication--based on a shared social space--on social interaction is of paramount importance. Not only the shared sound space, but also the shared social nearness--in the real world--influences the human to human interaction [20].

The general advantage and disadvantage of immersive VR are the necessity to put the user into a complete modeled virtual world. This concept of immersing the user in the computer's world ignores the on-going process of interacting with the real world. In the same interface space the mixing of real and virtual objects is not possible. But, humans are--most of their time--part of a real world and interact with real objects and other real humans.

Augmented Reality and Ubiquitous Computing

By definition, the goal of augmented reality is to make computational devices ubiquitous in the everyday world. Although oft described as making these devices "invisible," most efforts have focused on replication and dissemination. The vision is something like "Computing devices will be everywhere. They will be smaller so that we can carry them around with us and they will be able to communicate back and forth with each other." [28] The end result is that although computational devices are more likely to be seen as ubiquitous, they are far from invisible. Our world is now cluttered with various computational things that vie for our attention.
Augmented reality and ubiquitous computing is an inevitably evolving topic in computing. Many devices, which are already equipped with computers, can be found in our homes or offices. Nevertheless people still use their washing machines and telephones not considering them as a computers. Additionally, growing communication possibilities and the advent of small computers like Personal-Digital-Assistance (PDA) let Mark Weiser's vision come true [27] [28] (see figure 1). In the ubiquitous computing research we consider technology and applications, using PDAs, other mobile devices and extended "dump" machines (like TV) to set up prototypes for the demonstration of ubiquitous computing scenarios.

Figure 1: The world of net-worked mobile computing.

A common focus shared by researchers in mobile, ubiquitous and wearable computing is the attempt to break away from the traditional desktop computing paradigm. Computational services need to become as mobile as their users. Whether that service mobility is achieved by equipping the user with computational power or by instrumenting the environment, all services need to be extended to take advantage of the constantly changing context in which they are accessed.
Attaining the goals of augmented reality and ubiquitous computing will require a highly sophisticated infrastructure. In the ideal system, a real-time tracking mechanism will derive the locations and operational status of many system components and will use that context to deliver messages more intelligently. Users will be able to choose from among a variety of devices to gain mobile, high-bandwidth access to data and computational resources anywhere on the network. These devices will be intuitive, attractive and responsive. They will automatically adapt their behavior to suit the current user and context.
The wireless communication revolution is bringing fundamental changes to telecommunication and computing. Wide-area cellular systems and wireless LANs promise to make integrated networks a reality and provide fully distributed and ubiquitous mobile computing and communications, thus bringing an end to the tyranny of geography. Furthermore, services for the mobile user are maturing and are poised to change the nature and scope of communication
Recent advances in wireless networks and computer down-sizing technologies have led to the development of the concept of mobile computing. In the near future, millions of mobile users will be equipped with small, powerful and battery-operated palmtops. Through the wireless networks, this portable equipment will become an integrated part of existing distributed computing environments, and mobile users can access to data stored at information servers located at the static portion of the network even while they are on the move. In our increasingly mobile world, the ability to access information on demand at any location can satisfy people's information needs as well as conferring on them a competitive advantage. As such, the potential market for mobile computing applications is estimated to be billions of dollars annually. For example, passengers will access airline schedules, investors will access stock activities, travelers will access weather or traffic conditions.

For wireless computing to be widely accepted, there are two obstacles to be cleared. First, for palmtops that operate on AA batteries, power conservation is a key issue. For an `average user', the power source is expected to last 2 to 3 hours before replacing or recharging becomes necessary. What makes it worse is the predictions by battery experts of the modest improvement in battery capacity of only 20%-30% over the next 5-10 years. Second, the bandwidth of the wireless channel is also very limited. The bandwidth of a wireless channel can vary from 1.2 Kbps for slow paging channels, through 19.2 Kbps (e.g. Cellular Digital Packet Data) to about 2 Mbps for the wireless LAN. Therefore, mobile computers would frequently be disconnected from the network or be kept in a weak connection status to conserve energy, and transmission be kept at a minimum to avoid channel contention. These two issues pose a great challenge to researchers in the community.
What are the relevant research fields in ubiquitous and mobile computing?

Handheld Devices

Handheld devices like PDAs are a key technology and build the base for many mobile computing and ubiquitous computing projects (see [2]); see figure 2).

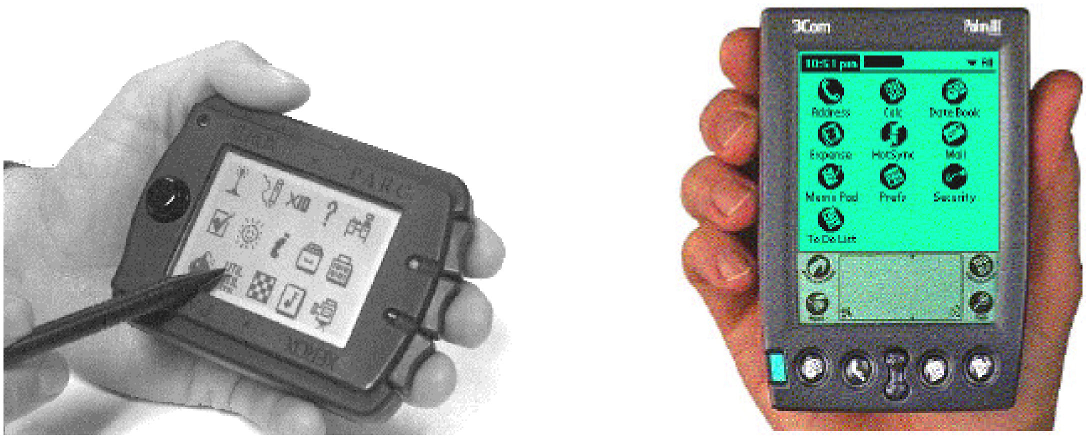

Figure 2: The unsuccessful PARCtab and the successful PALMpilot.

Location

Knowledge about the location of a person or device is useful for many applications in ubiquitous computing. An examples is the routing of phone calls; to provide this service a user has to be located all around a building or area (absolute location). But even relative location, where only devices and persons in the direct proximity are detected (e.g. in a room), could be useful. One example is the ad-hoc meeting, supported by PDAs, where social communication is supported by PDAs. The research in this area focuses on how to get this location information and how to present this information in the background computer system.

Context of use

There is more to context than location **-** For example, context is also the social situation the user actually was in. In this research area it is worthwhile to find out what sensor information could be useful to detect situations and how applications could profit from this information.

Information Access

The relation between direct and indirect information access is one of the most exciting areas of research in ubiquitous computing. Ubiquitous computing environments give a lot of possibilities to retrieve, to input and to output information, so combining only these possibilities leads to a lot of new challenges for new applications. Different information access scenarios are built to find out how information access works in ubiquitous computing scenarios.

Privacy

When building global communication and network infrastructures and "computerizing" the user as described in ubiquitous computing the question of privacy arises. Technologies like active badges or even GSM mobile phones enable computer systems to track any person using such technologies. These issue of privacy and security as well as individualism is underestimated by current ubiquitous computing proposals and collides with the real world. As a solution we suggest a concept that integrates a personal device (PD) into ubiquitous computing to overcome these problems.

Encryption

New network technologies like infrared or wireless radio communication need encryption technology to ensure security to its users. But available bandwidth of these networks is small and therefore a precious resource. The ongoing research tries to find solutions, based on available and well introduced technologies.

A framework for Natural User Interfaces (NUI)

Augmented Reality (AR) recognizes that people are used to the real world and that the real world cannot be reproduced completely and accurately enough on a computer. AR builds on the real world by augmenting it with computational capabilities. AR is the general design strategy behind "Natural User Interfaces" (NUI).
A system with a NUI supports the mix of real and virtual objects. As input it recognizes (visually, acoustically or with other sensors) and understands physical objects and humans acting in a natural way (e.g., speech input, hand writing, etc.). Its output is based on pattern projection such as video projection, holography, speech synthesis or 3D audio patterns. A necessary condition in our definition of a NUI is that it allows inter-referential I/O [8], i.e. that the same modality is used for input and output. For example, a projected item can be referred directly by the user for his or her nonverbal input behavior (see figure 3).

Figure 3: An intelligent and ambient office environment.

The spatial position of the user is monitored by two cameras. This also creates a stereoscopic picture for potential video conference partners. Speech and sound are recorded by several microphones, again allowing the system to maintain it's internal 3D model of the user. A third close-up camera on the top, records permanently the content state of the users work taking place on the horizontal working area. There, virtual and physical objects are fully integrated.
This set-up of several parallel input channels allows to show multiple views to remote communication partners, such as a (3D) face view and a view of shared work objects. Multimedia output is provided through the vertical communication area display, the projection device from the top down to the working area and through the four loudspeakers, producing a spatial impression on the user. Free space in the communication area can be used for (content) work, too. Of course, traditional input and output devices still can be used in addition. As required by Tognazzini [25], NUIs are multi modal and therefore allow users to choose for every action the appropriate and individually preferred interaction style.
Since human beings manipulate objects in the physical world most often and most naturally with hands, there is a desire to apply these skills to user-system interaction. In fact, NUIs allow the user to interact with real and virtual objects on the working area in a--literally--direct manipulative way! The working area is primarily horizontal, so that user can put real objects on the surface. Users get the feedback of the state of manipulated objects exactly at the same place where they manipulate these objects: perception space and action space coincide!

'NUI'-like applications

There exist already several applications realizing different NUI-like systems. These prototypes are:

- *DigitalDesk*: Wellner developed the first DigitalDesk [30]. It is based around an ordinary desk and can be used as such, but has extra capabilities. A video camera is mounted above the desk, pointing down at the work surface. The camera's output is fed through a pattern recognition system that can detect where the user is pointing and it can read documents that are placed on the desk. A computer-driven projector as an output device is also mounted above the desk, allowing the system to project electronic objects onto the work surface and onto real paper documents (something that can not be done with flat display panels or rear-projection).
- *Mosaic* [16]: This application is based on the DigitalDesk. Mosaic addresses the problem of extending paper along the temporal dimensions. Mosaic provides an interface that combines the benefits of paper storyboards with computer-controlled video.
- *Ariel* [17]: The research team of Mackay explored new ways of supporting engineers with a NUI-based system. They analyzed the task of engineers constructing the world's longest suspension bridge in Denmark. In practice, engineers often work with physical paper copies of CAD-constructed engineering drawings. To update manually added annotations in the on-line version and to support communicating such information between engineers, physical drawings have been enriched with a digital media space. An engineer's personalized paper drawing becomes the interface to the computer system.
- *PlayingDesk* [21]: Rauterberg et al. realized a game to be played on a projected virtual playing field with *real* chips. First, the player puts a real black chip on a field. The system recognizes the position, calculates the position of its next move and marks that field with a virtually projected red chip. In a field trial (during a 5-day fair) we had installed 4 stations, each offering the same game with a different user interface style (command interface, mouse-based GUI, GUI with touch screen and the PlayingDesk/NUI). More than 8000 users played a game against one of our computers. We found a significant higher chance for the users to win if they played with the well-known game board (NUI, PlayingDesk) while the chance to win was smaller if they played at one of the other three stations with conventional user interfaces.
- *DoubleDigitalDesk* [31]: The DoubleDigitalDesk, an extended version of the DigitalDesk, makes it possible to share real paper documents. It allows both users in two separate locations to "share" their physical desks, for seeing, editing and writing on each other's paper documents.
- *InteractiveDesk* [3]: A computer image is displayed on the desk with an additional pen-input facility and an ordinary upright display with a keyboard, thus integrating conventional office systems and an augmented reality system. The prototype assists users in switching input method (based on location of the keyboard) and retrieving files using real objects.
- *BrightBoard* [24]: The BrightBoard concentrates on using video input to control a computer system. A camera records commands on a whiteboard, flip-chart or even a sheet of paper. Feedback is provided acoustically by sound or by a speech synthesiser. Though the BrightBoard is implemented based on the ideas of augmented reality, it is *not* a NUI in the definition given at the beginning of the previous section, since it does not support inter-referential I/O. The output is acoustical and the input is visual. The user has to learn a lot of commands (written on the board by hand) to control the system in the manner that is typical for a command user interface.
- *Build-It* [22]:BUILD-IT is a planning tool based on computer vision technology, supporting complex planning and composition tasks (see figure 4). The system enables users, grouped around a table, to interact in a virtual scene, using real bricks to select and manipulate objects in the projected output space on the table. The objects can be created, positioned, rotated and fixed by simple brick manipulations. To support bi-manual interaction and multi-user operations, the system features multi-brick interaction.

In all presented NUI-based applications the computer identifies the user's commands by video-based pattern recognition and executes the appropriate actions by projecting its output back into the same working area.

NUI-related technical research areas

To realise a NUI, input and output have to be supported by pattern recognition techniques. The user interacts in his natural environment with real or virtual objects. The system captures the scene and processes the image to recognise (in the context) meaningful objects. This could be (1) information on a piece of paper or a whiteboard, it may be (2) a human-being acting on the scene, it could be (3) any flat or 3D object, or it can be (4) a digital object projected into the scene. Therefore, efficient and robust algorithms are required from the following research areas:

- 2D and 3D [14] [26] object recognition,

- pattern recognition [5],
- scanning and optical character recognition [10],
- gesture recognition [11],
- voice [6] [33] and sound recognition [13],
- image understanding [7].

Is has to be determined which images are relevant and when the system has to take an action. E.g., if the user is moving his finger over a document to point on a particular word, and the system takes several images, how does it know on which image the user is arrived at his final position?
On the output side the following fields are involved:

- 2D projection,
- 3D projection [4] [12],
- 3D audio rendering [32],
- speech [1] and sound [18] synthesis.

2D projection is a well-settled discipline, used in everyday situations, such as slide projection, TV projection, etc. 3D projection is still a research issue. Promising approaches are digital holography [15] and volumetric display systems [4]. The real-time integration of a 3D-object into a filmed (live) scene is described by Kansy et al. [12].

Conclusion

We have shown that Natural User Interfaces (NUI) have many advantages over traditional interaction styles and [immersive] Virtual Reality (VR). Especially aspects of nonverbal communication between humans and capturing task-related activities (motor movements, voice and sound) are emphasized. Technology is already ripe enough that many research prototypes can show the feasibility of particular components of NUIs. Calling for multi-disciplinary research, we identified several domains we consider as key technologies to attack the remaining technical problems, which are mainly pattern recognition and projection. Providing people with support for their real-world tasks, NUIs open a wide new design space for the next generation of USI technology.

References

[1] Allen J, Hunnicutt S, Klatt D (1987). From text to speech - the MITalk system. Cambridge Univ. Press.

[2] Anderson S, Kobara S, Mathis B, Rosing D and Shafrir E (1995). Synergies--a vision of information products working together. In: CHI '95 Companion: Conference on Human Factors in Computing Systems: Mosaic of Creativity. ACM, New York.

[3] Arai T, Machii K, Kuzunuki S, Shojima H: (1995). InteractiveDESK: A Computer-augmented Desk Which Responds to Operations on Real Objects. In *Proc. of the CHI ´95 Companion*, pp. 141-142.

[4] Blundell B, Schwarz A, Horrell D (1993). Volumetric Three-dimensional Display Systems: Their Past, Present and Future. *Engineering Science and Education Journal*, October 1993, pp. 196-200.

[5] Chandrasekaran V, Palaniswami M, Caelli T: (1995). Pattern Recognition by Topology Free Spatio-Temporal Feature Map. In *Proc. of IEEE Intern. Conf. on Systems, Man and Cybernetics* '95, pp. 1136-1141.

[6] Cohen P, Oviatt S: The role of voice input for human-machine communication. *Proc. Natl. Acad. Sci. USA*, 92(22), 1995, pp. 9921-9927

[7] Covell M: Autocorrespondence: Feature-based Match Estimation and Image Metamorphosis. In *Proc. of IEEE Intern. Conf. on Systems, Man and Cybernetics*, 1995, pp. 2736-2741.

[8] Draper S (1986). Display Managers as the Basis for User-Machine Communication. In Norman D, Draper S (eds.) *User Centered System Design*. Lawrence Erlbaum, pp. 339-352.

[9] Gibson J (1979). The Ecological Approach to Visual Perception. New York: Houghton Mifflin.

[10] Govindaraju V, Srihari S: System for Reading Handwritten Documents. In *Proc. of IEEE Intern. Conf. on Systems, Man and Cybernetics*, 1995, pp. 2347-2352.

[11] Kämmerer B, Maggioni C: GestureComputer - Research and Practice. In *Proc. of 4th Intern. Conf. INTERFACE to Real & Virtual Worlds*, 1995, pp.251-260.

[12] Kansy K, Berlage T, Schmitgen G, Wisskirchen P: Real-Time Integration of Synthetic Computer Graphics into Live Video Scenes. In *Proc. of 4th Intern. Conference INTERFACE to Real & Virtual Worlds*, 1995, pp. 93-101.

[13] Kramer G (ed.): Auditory Display: Sonification, Audification and Auditory Interfaces. *SFI Studies in the Sciences of Complexity, Proc. Vol. XVIII*, Addison-Wesley, 1994.

[14] Liu R, Durdle N, Peterson A: Three Dimensional Reconstruction of Trunk Surface Using Structured Light. In *Proc. of IEEE Intern. Conf. on Systems, Man and Cybernetics*, 1995, pp. 1085-1090.
[15] Lucente M, Galyean T: Rendering Interactive Holographic Images. In *Proc. of SIGGRAPH'95*, 1995, pp. 387-394.
[16] Mackay W, Velay G, Carter K, Ma C, Pagani D: Augmented Reality - Adding Computational Dimensions to Paper. *Communications of the ACM*, 36(7), 1993, pp. 96-97.
[17] Mackay W, Pagani D, Faber L, Inwood B, Launiainen L, Brenta L, Pouzol V: Ariel: Augmenting Paper Engineering Drawings. In *Proc. of the CHI ´95 Companion*, 1995, pp. 421-422.
[18] Munteanu E, Guggiana V, Darvishi A, Schauer H, Rauterberg M, Motavalli M: Physical modelling of environmental sounds. In: Proc. of '2nd Intern. Conf. on Acoustic and Musical Research'–CIARM' 95 (F. Pedrielli, ed.) Ferrara, 1995, pp. 107-112.
[19] Nickerson R S and Landauer T K (1997). Human-computer interaction: Background and issues. Handbook of Human-Computer Interaction, Second Edition, 3-31. Amsterdam: Elsevier.
[20] Rauterberg M, Dätwyler M, Sperisen M: From Competition to Collaboration through a Shared Social Space. In: *Proc. of East–West Intern. Conf. on Human–Computer Interaction (EWHCI '95)*, 1995, pp. 94-101.
[21] Rauterberg M, Stebler R, Mauch T: "Augmented Reality" in Contest with a Command, Desktop and Touch Screen Interface. In *Proc. of 5th Intern. Conf. INTERFACE to Real & Virtual Worlds*, 1996, pp. 95-103.
[22] Rauterberg M, Bichsel M, Fjeld M, Leonhardt U, Krueger H & Meier M (1998). BUILD-IT--A Planning Tool for Construction and Design. In the video program of CHI'98, ACM, New York.
[23] Shneiderman B: Designing the User Interface. Addison-Wesley, Reading MA, 1987.
[24] Stafford-Fraser Q, Robinson P: BrightBoard: A Video-Augmented Environment. In *Proc. of the CHI´96*, 1996, pp. 134-141.
[25] Tognazzini B: *Tog on Software Design*. Addison-Wesley, Reading MA, 1996.
[26] Wang Z, Ohnishi N: Deformable Template Based Stereo. In *Proc. of IEEE Intern. Conf. on Systems, Man and Cybernetics*, 1995, pp. 3884-3889.
[27] Weiser M (1991). The Computer for the Twenty-First Century. *Scientific American.* September 1991.
[28] Weiser M (1993). Some computer science issues in ubiquitous computing. Communications of the ACM, 36(7), 75-85.
[29] Wellner P, Mackay W, Gold R (1993). Computer-Augmented Environments: Back to the Real World. *Communications of the ACM*, 36(7), pp. 24-26.
[30] Wellner P (1993). Interacting with Paper on the Digital Desk. *Communications of the ACM*, 36(7), pp. 87-96.
[31] Wellner P, Freeman S (1993). The DoubleDigitalDesk: Shared Editing of Paper Documents. *Technical Report EPC-93-108, Xerox Research Centre Cambridge Laboratory*, Cambridge UK.
[32] Wenzel E (1992). Localization in Virtual Acoustic Displays. *Presence* 1(1), pp. 80-107.
[33] Zue V et al (1991). Spoken Language Systems for Human/Machine Interfaces. In *Intelligent Text and Information Handling RIAO91.* (A. Lichnerowicz, ed.) Elsevier, 1991, pp. 936-955.