A gesture based interaction technique for a planning tool for construction and design

M. Rauterberg[1], M. Bichsel[2], M. Meier[2] & M. Fjeld[1]

[1]Institute for Hygiene and Applied Physiology (IHA)
[2]Institute of Construction and Design Methods (IKB)
Swiss Federal Institute of Technology (ETH)
Clausiusstrasse 25, CH-8092 Zurich, SWITZERLAND
http://www.ifap.bepr.ethz.ch/~rauter/science.html

ABSTRACT

In this article we wish to show a method that goes beyond the established approaches of human-computer interaction. We first bring a serious critique of traditional interface types, showing their major drawbacks and limitations. Promising alternatives are offered by Virtual (or: immersive) Reality (VR) and by Augmented Reality (AR). The AR design strategy enables humans to behave in a nearly natural way. Natural interaction means human actions in the real world with other humans and/or with real world objects. Guided by the basic constraints of natural interaction, we derive a set of recommendations for the next generation of user interfaces: the Natural User Interface (NUI). Our approach to NUIs is discussed in the form of a general framework followed by a prototype. The prototype tool builds on video-based interaction and supports construction and plant layout. A first empirical evaluation is briefly presented.

1 Introduction

The introduction of computers in the work place has had a tremendous impact on task solving methods in that area. Mouse based and graphical displays are everywhere, the desktop workstations define the frontier between digital (computer) and analogue ('real') worlds. We spend a lot of time and energy transferring information between such worlds. This effort could be reduced by better integrating the virtual world of the computer with the real world of the user and vice versa.

In the past, several dialogue techniques were developed and are now in use. The following dialogue techniques and objects can be distinguished: command language, function key, menu selection, iconic, and window [15]. These five essential terms can be cast 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 currently available function points.

- *Menu selection:* This includes rigid menu structures, pop-up and pull-down menus, fill-in forms, etc. It is characterised by dual usage of the function keys. They support dialogue management as well as application functionality.

Pros: All available functions are represented by partly or fully visible interaction points.

Cons: Finding a function point deeper in the menu hierarchies is cumbersome.

- *Direct manipulation*: This type of interaction only took on weight as the bit mapped graphical displays were introduced. The development of this interaction style is based on the desktop metaphor, assuming that realistic depiction of the work environment (i.e. the desk with its files, waste-paper basket, etc.) helps users adjust to the virtual world of electronic objects.

Pros: All functions are continuously represented by visible interaction points (e.g. mouse sensitive areas). 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 cannot combine real world and virtual objects within the *same* interface space. Nor do they incorporate the human hands' enormous potential for interaction with real and virtual objects. This aspect was one of the basic incitements to develop data gloves and data suits. Users equipped with such artefacts can interact in an immersive, virtual reality (VR) system. Another reason to realise VR systems, was the emergence of the of head mounted displays with 3D output capabilities . However, VR systems are still subject to serious, inherent limitations, such as:

- The lack of tactile and touch information, leading to a mismatch with the proprioceptive feedback. Special techniques are proposed to overcome this problem [4].

- The lack of depth perception, due to visual displays only generating 2D output. Many informational concepts offer a remake of the 3D impression by superimposing 2D pictures [12].

- A consistent delay in the user-computer control loop, often yielding severe problems with reference to the perceptual stability of the ear vestibular apparatus [5].

- An influence from communication on social interaction. A shared sound space, as well as a shared real social world, stimulates humans to mutual interaction [11].

The advantage, but at the same time disadvantage of immersive VR, is the necessity to put the user into a fully modelled, virtual world. Bringing users into the computer world, ignores their on-going interaction with the real world, because mixing of real and virtual objects is not yet possible. Nevertheless, humans are most of the time part of a real world where they interact with real objects and humans.

To overcome the drawbacks of immersive VR, the concept of *Augmented Reality* (AR) [18] was introduced. This approach is promising because it incorporates fundamental human skills: interaction with real world subjects and objects. Hence, the AR design strategy enables humans to behave in a nearly natural way; we call this way natural interaction.

Guided by the AR approach *and* the basic constraints of natural interaction, we derive a set of recommendations for the next generation of user interfaces: the *Natural User Interface* (NUI). The NUI approach is discussed in form of a general framework and in the form of a prototype. The prototype tool builds on video-based interaction, and supports construction and plant layout. A first empirical evaluation will be briefly presented.

2 Behaviour in the Real World

Interaction with real world objects is constrained by the laws of physics (e.g. matter, energy, mechanics, heat, light, electricity and sound). In a more or less similar way, human interaction is based on social and cultural norms.

Task related activities have been a topic in various behavioural approaches. Mackenzie [8] introduced prehensile behaviour as "... the application of functionally effective forces by the hand to an object for a task, given numerous constraints." Sanders [14] proposed certain classes of motor movements: "(1) *Discrete movements* involve a single reaching movement to a stationary target, such as reaching for a control or pointing to a word on a computer screen. Discrete movements can be made with or without visual control. (2) *Repetitive movements* involve a repetition of a single movement to a stationary target or targets. Examples include hammering a nail or tapping a cursor on a computer keyboard. (3) *Sequential movements* involve discrete movements to a number of stationary targets regularly or irregularly spaced. Examples include typewriting or reaching for parts in various stock bins. (4) *Continuous movements* involve movements that require muscular control adjustments of some degree during the movement, as in operating the steering wheel of a car or guiding a piece of wood through a band saw. (5) *Static positioning* consists of maintaining a specific position of a body member for a period of time. Strictly speaking, this is not a movement, but rather the absence of movement. Examples include holding a part in one hand while soldering, or holding a needle to thread it".

In the context of this article we are primarily interested in purposeful motor activities. These activities are executed by a person to achieve some goal (in contrast to erroneous or exploratory behaviour). Actions (e.g. motor based movements) will be functionally, but not anatomically nor mechanically defined. The catching of a ball could be carried out by either the left or the right hand, the starting position of the approach and the catching position of the ball might change from one reach to the next, and no two reaching trajectories will look exactly alike. However, these movements are classified as the same action because they share the same function.

Following the argumentation of Fitzmaurice, Ishii and Buxton [7] a grasp-based user interface has the following advantages:

• it encourages two handed interactions

• it shifts to more specialised, context sensitive input devices

• it allows for more parallel input specification by the user

• it leverages off of our well developed skills … for physical object manipulations

• it externalises traditionally internal computer representations

• it facilitates interactions by making interface elements more 'direct' and more 'manipulable' by using physical artefacts

• it affords multi-person, collaborative use.

Summarising the above discussion about real world behaviour, we come to the following design recommendation: To enhance human computer interaction, users must be able to behave in a *natural way*, bringing into action all of their body parts (e.g. hands, arms, face, head and voice). To interpret all of these expressions we need very powerful and intelligent pattern recognition techniques.

3 A Framework for Natural User Interfaces (NUI)

Augmented Reality (AR) recognises that people are used to the real world, which strictly cannot be reproduced by a computer. AR is based on the real world, augmented by computer characteristics. It is the general design strategy behind "Natural User Interfaces" (NUI) [13].

A NUI based system supports the fusion of real and virtual objects. It understands visual, acoustic and other human input forms. It also recognises physical objects and human actions like speech and hand writing in a natural way. Its output is based on pattern projection such as

video, holography, speech synthesis and 3D audio strips. NUI necessarily implies inter-referential I/O [6], meaning that the same modality is used for input and output. Hence, a projected item can be referred directly by the user as part of his or her non-verbal input behaviour. Figure 1 gives an overview of what a NUI based system could look like.

The spatial position of the user is monitored by one or more cameras. This could also create a stereoscopic picture for potential video conference partners. Speech and sound are recorded by several microphones, enabling the system to maintain an internal 3D user model. From above, a close-up camera permanently records the state of the user activity taking place in the horizontal working area. In this very area, virtual and physical objects are fully integrated.
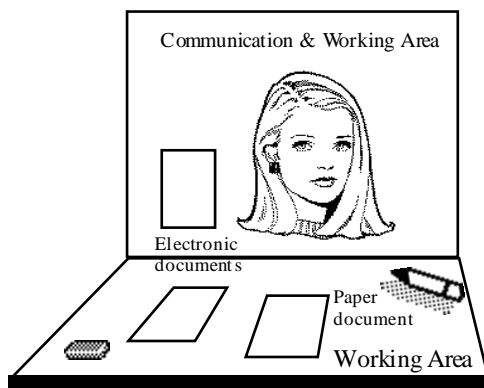

Figure 1: Architecture of a Natural User Interface.

The set-up of several parallel input channels makes it possible to communicate multiple views to remote partners, such as a 3D face view [17], and a view of shared work objects [20]. Multimedia output is provided by a) the vertical display, b) the projection device illuminating the working area, and c) a multichannel audio system. Free space in the communication area can be used for other work (see Figure 1). Of course, traditional I/O devices can be added. As required by Tognazzini [16], NUIs are multimodal, so users are allowed to (re-)choose their personal and appropriate interaction style at any moment.

Since humans often and easily manipulate objects in the real world with their hands, they have a natural desire to bring in this faculty when interacting with computers. NUIs allow users to interact with real and virtual objects on the working area in a 'literally' direct manipulative way. Since the working area is basically horizontal, the user can place real objects onto its surface. So there is a direct mapping of the real, user manipulated object onto its corresponding virtual object. We can actually say that perception and action space coincide, which is a powerful design criterion, described and empirically validated by Rauterberg [10].

4   The Prototype "Build-it"

In a first step, we designed a system primarily based on the concept of NUIs. However, we did not support the communication aspects of a computer based, co-operative work environment. As our task context, we chose that of planning activities for plant design. A prototype system, called "BUILD-IT", was realised. This is an application that supports engineers in designing assembly lines and building plants. The realised design room (see Figure 2) enables users, grouped around a table, to interact in a space of virtual and real world objects. The vertical working area in the background of Figure 2, gives a side view of the plant. In the horizontal working area there are several views where the users can select and manipulate objects.

The hardware comprises seven components:
- A table with a white surface is used as horizontal working area.
- A white projection screen provides the vertical working area.
- An ASK 960 high resolution LCD projector projects the horizontal views vertically onto the table.
- An ASK 860 high resolution LCD projector projects the vertical view horizontally onto the projection screen.
- A CCD camera with a resolution of 752(H) by 582(V) pixels looks vertically down to the table.
- A brick, size 3 cm x 2 cm x 2 cm, is the physical interaction device (the universal interaction handler).
- A low-cost Silicon Graphics Indy (IP22 R4600 133 MHz processor and standard Audio-Video Board) provides the computing power for digitising the video signal coming from the camera, analysing the user interactions on the table, and rendering the interaction result in the two views.

Figure 2: The design room of BUILD-IT.

The software consists of two independent processes communicating via socket connection:

- A real time process for analysis of the video images. This process extracts and interprets contours of moving objects [2, 3], and determines the position and orientation of the universal interaction handler (the brick).

- An application built upon the multi-media framework MET++ [1]. Based on the position and orientation of the interaction handler, it interprets user actions. It modifies a virtual scenario, and renders the above (and side) view of the new scenario via the vertical (and horizontal) projector (Figure 2).

The application is designed to support providers of assembly lines and plants in the early design processes. It can read and render arbitrary CAD models of machines in VRML format. The input of a 3D model of the virtual objects is realised by connecting BUILD-IT with the CAD-System CATIA. Thus, the original CAD-models were imported into BUILD-IT.

Geometry is not the single aspect of product data. There is a growing need to interact in other dimensions, like cost, configurations and variants. Therefore, it will be possible to send and receive additional metadata from BUILD-IT.
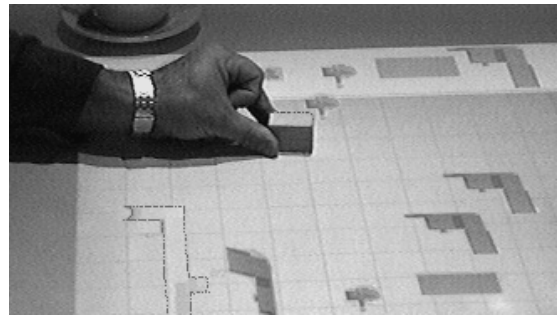

Figure 3: The two working areas and their views.

Figure 4: The object menu (white), the above view (grey), and the user hand moving the interaction handler (the brick).
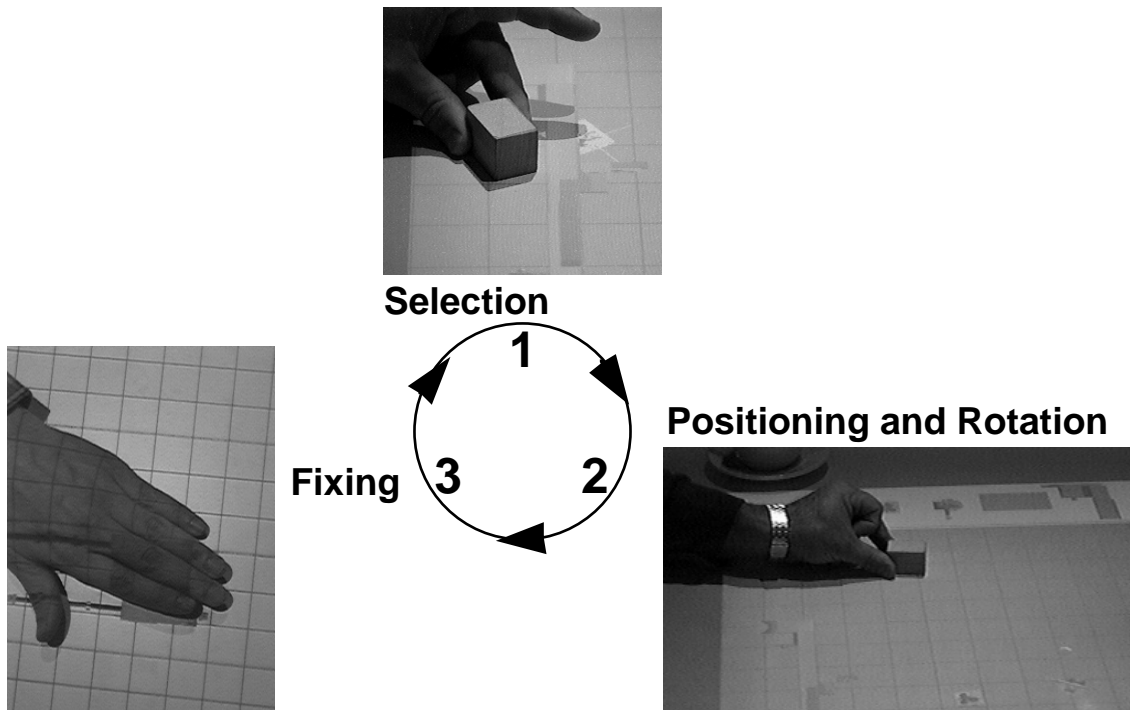

Figure 5: This cycle gives the three basic steps for user manipulations with the interaction handler (the brick).
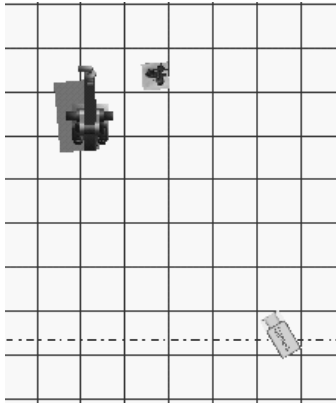
Figure 6: The above view showing the position of the camera and the robot, etc.
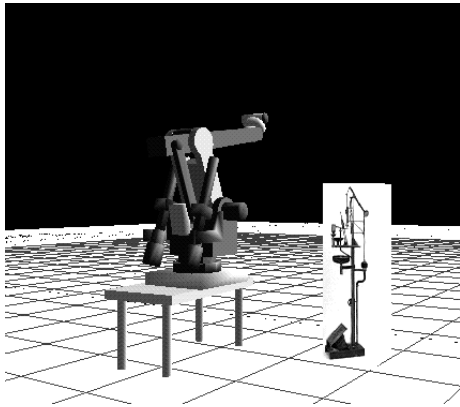

Figure 7: Side view perspective of the robot, etc. as seen with the same camera setting.

BUILD-IT currently features following user (inter-) actions (see Figure 3 to Figure 7):

- Selection of a virtual object (e.g. a specific machine) in a 'virtual machine store' by placing the interaction handler onto the projected image of the machine in the object menu.
- Positioning of a machine in the virtual plant by moving the interaction handler to the preferred position in the above view of the plant layout.
- Rotation of a machine is supported through a coupling of the machine and brick orientation.
- Fixing the machine by covering the surface of the interaction handler with the hand and removing it.
- Re-selection of a machine by placing the interaction handler onto the specific machine in the above view.
- Deleting the machine by moving it back into the object menu (the virtual machine store).
- Printing of the views, offered by a method menu icon.

- Saving of the working area contents, also offered by a method menu icon.
- Modification of object size and height by operators in the method menu applied on objects in the above view.
- Direct modification of object altitude in the height view.
- Scrolling of above view and menus.
- Automatic grouping of two or more objects along predefined contact lines within the above view.

In the above view (Figure 4) the user is permanently given a look from far above, giving the impression of a 2D situation. The camera is picked up in the method menu and manipulated like any other virtual object (Figure 6).

In the side view (Figure 7) a perspective is offered that gives the user an impression of a virtual human looking at a real situation. As the user moves the camera around, a real time update of the side view takes place.

5 First Empirical Evaluation

The system has been empirically tested with managers and engineers from companies producing assembly lines and plants. These tests showed that the system is intuitive and enjoyable to use as well as easy to learn. Most persons were able to assemble virtual plants after only 30 seconds of introduction to the system.

Some typical user comments were:

- The concept phase is especially important in plant design since the customer must be involved in a direct manner. Often, partners using different languages sit at the same table. This novel interaction technique will be a mean for completing this phase efficiently and almost perfectly.
- This is a general improvement of the interface to the customer, in the offering phase as well as during the project, especially in simultaneous engineering projects.
- A usage of the novel interaction technique will lead to a simplification, acceleration, and reduction of the iterative steps in the start-up and concept phase of a plant construction project.

6 Conclusion

One of the most interesting benefits of a NUI-based interface is the possibility to combine real and virtual objects in the same interaction space [7, 16, 19]. Taking this advantage even further, we will implement two or three interaction handlers, allowing simultaneous interaction of several users grouped at one single table.

With this new interaction approach, customers whether CAD experts or not, can equally take part in discussions and management of complex 3D objects. Products and technical descriptions can easily be presented, and new requirements are realised and displayed within short time. The virtual camera allows a walk-through of the designed plant. Such inspection tours can give invaluable information about a complex system.

In the near future, one could imagine a direct, NUI-based information flow between customers and large product databases. It is conceivable that users wanting to change one detail of a machine, will have several configuration options presented on their table. As soon as one has been selected, the exact configuration cost will be calculated and displayed.

7 References

[1] Ackermann P: Developing Object-Oriented Multimedia Software Based on the MET++ Application Framework. Heidelberg: dpunkt Verlag für digitale Technologie, 1996.

[2] Bichsel M: Illumination Invariant Segmentation of Simply Connected Moving Objects. 5th British Machine Vision Conference, University of York, UK, September 13-16, 1994, pp. 459-468.

[3] Bichsel M: Segmenting Simply Connected Moving Objects in a Static Scene. Transactions on Pattern Recognition and Machine Intelligence (PAMI), Vol. 16, No. 11, Nov. 1994, pp. 1138-1142.

[4] CyberTouch, Virtual Technologies Inc., 2175 Park Boulevard, Palo Alto, CA.

[5] DIVE Laboratories Inc: Health and HMDs. On URL: http://www.divelabs.com /deeper.html

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

[7] Fitzmaurice G, Ishii H, Buxton W: Bricks: Laying the Foundations for Graspable User Interfaces. In Proc. of the CHI ´95, 1995, pp. 442-449.

[8] MacKenzie C, Iberall T: The grasping hand. Elsevier, 1994.

[9] Newman W, Wellner P: A Desk Supporting Computer-base Interaction with Paper Documents. In Proc. of the CHI ´92, 1992, pp. 587-592.

[10] Rauterberg M., Über die Quantifizierung software-ergonomischer Richtlinien. PhD Thesis, University of Zurich, 1995.

[11] 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.

[12] Rauterberg M, Szabo K: A Design Concept for N-dimensional User Interfaces. In Proc. of 4th Intern. Conf. INTERFACE to Real & Virtual Worlds, 1995, pp. 467-477.

[13] Rauterberg M, Steiger P: Pattern recognition as a key technology for the next generation of user interfaces. In Proc. of IEEE International Conference on Systems, Man and Cybernetics--SMC'96 (Vol. 4, IEEE Catalog Number: 96CH35929, pp. 2805-2810). Piscataway: IEEE.

[14] Sanders M, McCormick E: Human Factors in Engineering and Design. McGraw Hill, 1993.

[15] Shneiderman B: Designing the User Interface. Addison-Wesley, Reading MA, 1987.

[16] Tognazzini B: Tog on Software Design. Addison-Wesley, Reading MA, 1996.

[17] Watts L, Monk A: Remote assistance: a view of the work and a view of the face?. In Proc. of the CHI´96 Companion, 1996, pp. 101-102.

[18] Wellner P, Mackay W, Gold R: Computer-Augmented Environments: Back to the Real World. Communications of the ACM, 36(7), 1993, pp. 24-26.

[19] Wellner P: Interacting with Paper on the Digital Desk. Communications of the ACM, 36(7), 1993, pp. 87-96.

[20] Whittaker S: Rethinking video as a technology for interpersonal communications: theory and design implications. Intern. Journal of Human-Computer Studies, 42, 1995, pp. 501-529.

Proceedings

6th IEEE International Workshop on
Robot and Human Communication
29 Sept. – 1 Oct. 1997 Sendai

Hikaru Inooka (General Chairperson)

RO-MAN '97