

The Digital Playing Desk: a Case Study for Augmented Reality

M. Rauterberg¹, T. Mauch² and R. Stebler²

¹Work and Organisation Psychology Unit
Swiss Federal Institute of Technology
Nelkenstrasse 11, CH-8092 ZURICH
<http://www.ifap.bepr.ethz.ch/~rauterberg>

²Computer Science Department
Swiss Federal Institute of Technology
Clausiusstrasse 49, CH-8092 ZURICH

Abstract

To compare the advantages and disadvantages of a "Natural User Interface" (see Rauterberg & Steiger 1996) a field study was carried out. During five days of the largest computer fair in Switzerland four different computer stations with (1) a command language, (2) a mouse, (3) a touch screen, and (4) a Digital Playing Desk (DPD) interface was presented for public use. In this version of the DPD the user has to play a board game by moving a real chip on a virtual playing field against a virtual player. The task was to win the computer game "Go-bang". The reactions of the virtual player were simulated by "emoticons" as colored comic strip pictures with a corresponding sound pattern. We investigated the effects of these four different interaction techniques with two different methods: (a) an inquiry with a questionnaire, and (b) a field test for public and anonymous use. (a) Results of the inquiry: 304 visitors rated the usability of all four stations on a bipolar scale. The touch screen station was rated as the easiest to use interaction technique, followed by the mouse and DPD interface; the "tail-light" was the command language interface. One very important result was a significant correlation between "age" and "DPD usability". This correlation means that older people prefer significantly more a graspable user interface in form of the DPD than younger people. (b) Results of the field test: The analysis of 9'006 automatically recorded contacts to one of the four stations shows that the highest chance to win the game could be observed for the DPD interface technique. We conclude that the DPD as a graspable user interface with emotional, non-verbal feedback is a promising candidate for the next generation of dialog techniques.

1 Introduction

There are two main contrary directions for new interface technology: (1) [immersive] virtual reality (cf. [9]), and (2) augmented reality or ubiquitous computing, resp. (cf. [12]). Enthusiasts of the virtual reality approach (VR) believe "that all constraints of the real world can be overcome in VR, and physical tools can be made obsolete by more flexible, virtual alternatives" ([13] p. 87-88). We restrict the notion of "virtual reality system"--in the context of this paper--to systems with head mounted display and data gloves or suits. In these VR applications the user has to leave his natural physical and social environment and to

immerse in the simulated world. The following two unsolved problems are important: (1) how to simulate tactile and haptic feedback, and (2) how to overcome the social isolation for collaborative tasks. The effect, that the social nearness between real persons is of tremendous importance for collaboration, was investigated and shown in [6]. As we stated in [8], that the effects of tactile and haptic senses are important, we are looking for a realization of a user interface where the user can control the human-computer interaction by his hands dealing with real and virtual objects in the same interface space. One of the first systems with such a user interface was the DigitalDesk of Wellner [13].

Inspired by the ideas of Newman and Wellner [4], we were interested in a way to test empirically the advantages or disadvantages of the DigitalDesk in comparison with established interaction techniques. The DigitalDesk has the following three important features: (1) it projects electronic images (virtual objects) down onto the desk and onto real objects, (2) it responds to interaction with real objects (e.g., pens or bare fingers: hence *DigitalDesk*), and (3) it can interpret the scene on an appropriate semantic level (e.g., read paper documents placed on the desk; cf. [13]).

If we generalize the DigitalDesk approach to augment real-world objects, then it leads us directly to "Graspable User Interfaces" (cf. [1] [2]). We follow the argumentation of Fitzmaurice et al. [2] that a Graspable User Interface has the following advantages: "(1) It encourages two handed interactions; ... (3) allows for more parallel input specification by the user, ... (4) leverages off of our well developed skills ... for physical object manipulations; ... (6) facilitates interactions by making interface elements more 'direct' and more 'manipulable' by using physical artifacts; ... (9) affords multi-person, collaborative use" ([2] p.443).

These arguments sound convincing. But, all the published approaches--caused by the complexity of the implemented tasks--are research prototypes, so that only the experiences of preliminary user testing are reported (cf. [1], [2], [4]).

2 System Description

To run a laboratory investigation or a field study we need a fast, reliable and robust implementation of the whole system. First, we decided to minimize the task complexity and to restrict the user's action space to cognitive planning processes. For public use a simple computer game

seems to be best. We implemented a version of the computer game "Go-bang". The user has to play the game by moving a *real* chip on the *virtual* playing field (see Fig. 1). To compare this interface type with the most established dialog techniques we implemented the same game algorithm on three other stations with (1) a command language, (2) a mouse, and (3) a touch screen interface.

Command interface (CI): This station run on a 386er PC with a color screen (17") in an upright position. The user has to enter the co-ordinates of the desired place of a playing field with 12 by 12 positions (e.g., A1, L12). To start a new or to cancel a game she or he has to enter the command NEW. The internal state of the algorithm is presented as text in a special output field (e.g., "Make the next move").

Mouse interface (MI): This interface run on a 386er PC with a color screen (17") in an upright position. To move the user has to click with the mouse on the desired place. To start a new or to cancel a running game she or he has to click on the button NEW. The internal state of the algorithm is presented as text in a special popup window.

Touch screen interface (TI): This station run on 386er PC with a color touch screen (21") in an inclined position of 30 degrees. To make a move the user has to touch with a finger the desired place. To start a new or to cancel a running game she or he has to touch the button NEW. The virtual player was shown on a second colored screen (17") served by a second 386er PC in a client-server architecture.

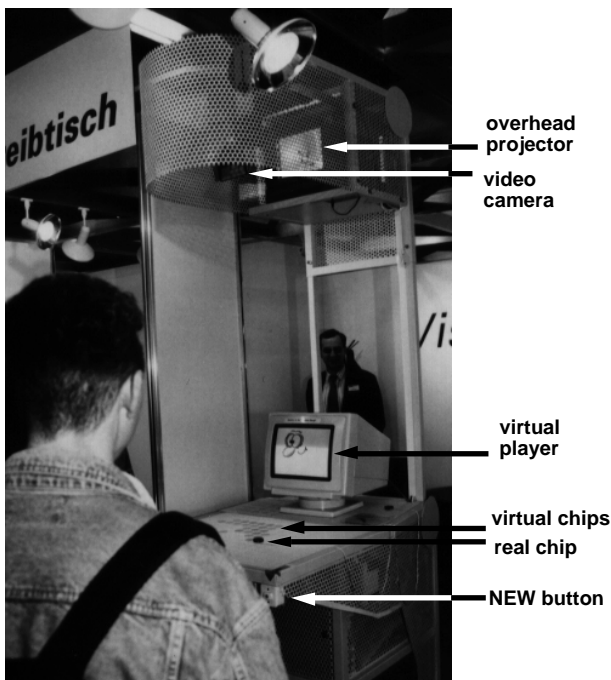


Fig. 1: The Digital Playing Desk--the front view for the users.

Digital Playing Desk (DPD): This station was completely realized in C++ on standard hardware components: (1) a Pentium PC, (2) an overhead projector of high luminous intensity and the projection panel GehaVision™, (3) a high resolution video camera, and (4) the video board MovieMa-

chinePro™. For the virtual player a second 386er PC was connected in a client-server architecture. A user has to make a move by putting a *real* chip on the desired place of the *virtual* playing field (see Fig. 1). The computer's output is the projection of a *virtual* chip on the desk. If a user wanted to cancel or restart a game, then he or she had to press the real NEW button in the front of the station.

The Virtual Player: The CI and MI have their output screens in an upright position. This upright position makes it impossible to give an opponent--in the metaphor of a game--an individual representation. The classical solution is, to give feedback about the machine's internal states as text or graphics in defined areas on the screen. This solution leads always to a partition of the screen into a working area (e.g., the playing field) and the feedback area. This superimposing of qualitatively different feedback's in the same output space can be overcome, if we use an additional output device. This can be done, if we separate the working area (the playing field) from the feedback area and if we change the upright position of the working area to a horizontal position. Therefore we composed the system of a flat table to project the playing field onto and of a second screen to present the virtual player.



Fig. 2: The two emoticons "reasoning" and "yawning".

All--from the user's point of view--important internal states of the game algorithm were presented by the virtual player with six different facial expressions and a corresponding sound (see Fig. 2; partially animated comic strip pictures: "Reasoning" = a face with a balloon of animated turning wheels and a machine-like sound, "Waiting for the next move" = a yawning face with a corresponding sound, "Initial state" = a face with blinking eyes, "Incorrect move" = an angry facial expression with an indignant cry, "Be the winner" = a happy facial expression with an arrogant laughter, "Be the loser" = a shrinking face with a disappointed cry). We call these six comic strip pictures "emoticons" (acronym for "emotional icons"). All emoticons of the virtual player are shown on a second color screen (17").

3 Validation of the Digital Playing Desk

To present our four dialog techniques to a broad population of heterogeneous users, four special stations were constructed. All stations were presented in a central exhibition area during five days at the largest computer fair of Switzerland in September 1995. The official number of visitors was approximately 70'000. Most of these people passed the exhibition area and many of them came into close contact with one of the stations (e.g., playing at one of the stations or observing other people playing).

3.1 Inquiry with a questionnaire

With an additional questionnaire we got some personal information's of several users of at least one of our four stations. The stand personal was instructed to request users to answer the questionnaire. It was especially necessary to ask women, because they behaved very reserved.

Subjects: The questionnaire was answered by 304 visitors (61 females, 243 males, 5 anonymous data). The average age of the women was 31 ± 13 years, and of the men 30 ± 14 years (T-Test: $p \leq .724$). As a *control variable* the "computer experience" was measured in millimetre on the corresponding bipolar rating scale ["no experience": 0 ... 90 mm: "expert"]. There was a significant gender difference in computer experience; the computer experience of men was higher than the experience of women (men: 58 ± 22 mm; women: 48 ± 23 mm; T-Test: $p \leq .002$).

Dependent Measures: The questionnaire consisted of two parts: (a) personal data (age in years, gender, computer experience in form of a bipolar rating scale, and (b) usability ratings. For each of the four interfaces the following five aspects were asked with a multiple choice question: "did you play", "did you loose", "did you win", "did you play draw", "did you cancel". The *dependent measure* per interface type is the number of millimetres of the user's marking on the bipolar rating scale ["very easy to use": 0 ... 70 mm: "very difficult to use"]. We also differentiated between "real station contact" and "only observer status".

Results: All persons, who answered that they had no real contact to one of the stations, had a significant higher amount of computer experience ("no contact": 70 ± 17 mm, $N = 10$; "with contact": 55 ± 23 mm, $N = 276$; T-Test: $p \leq .04$). We can not find a significant difference between the "contact" and "no contact" group in the usability ratings.

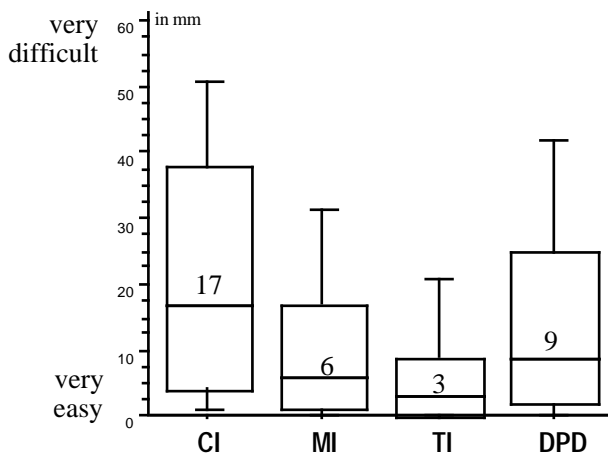


Fig. 3: Box Plots of the results of the subjective usability ratings for the four different interface types (CI: command line, MI: mouse interface, TI: touch screen, DPD: Digital Playing Desk).

The usability rating was best for the touch screen interface (see Fig. 3; median = 3; mean rank = 1.9), followed by the mouse interface (median = 6; mean rank = 2.4) and the Digital Playing Desk (median = 9; mean rank = 2.5). The "tail-light" was the command language interface (median =

17; mean rank = 3.3). These differences are significant (Friedman Test, $df = 3$, χ^2 corrected for ties = 89.1, $p \leq .0001$). One very important result was a significant correlation between "age" and "Digital Playing Desk's usability". This correlation means that older people prefer significantly more a graspable user interface in form of the Digital Playing Desk than younger people. ($R = -0.202$, $N = 179$, $p \leq .006$). All other correlations are not significant.

Discussion of the inquiry: To carry out an inquiry with a questionnaire only for scientific purposes in the context of a commercial fair was more difficulty than we expected. The only motivation for the user to fill out such a questionnaire--in contrast to all the lotteries of the commercial issuers around us (with their very attractive prizes)--was to bring in his or her personal opinion into an scientific research process. This argumentation was the most convincing reason to participate. Overall we got more filled out questionnaires than we expected, but much less than observed visitors at one of the four stations. To increase the number of answered questionnaires the stand personal has to be active: to go to the user and to ask for participation.

We could find two main results: (1) the touch screen interface was estimated as the easiest to use, and (2) the significant correlation between age and the usability ratings for the Digital Playing Desk. If we assume, that the average age of the populations in all high industrialised countries will increase in the next two or three decades, then this result will be of tremendous importance for the development of modern computer technology for elderly people! To compare the results of the user's subjective ratings with their concrete performances--as an objective measure--we analysed the outcome of all played games. To do this we carried out a field study.

3.2 Field study

The four stations were ready for unconstrained public and anonymous use during five days. The stand personal was instructed to hold back with help. To test the effect of non-verbal feedback we realized a virtual player on a second screen for the touch screen and the Digital Playing Desk station.

Subjects: All passing visitors had unconstrained access to all stations. It is difficult to get a valid picture of all users in a situation of public and anonymous use. We can assume that the results of the inquiry give us representative information about the gender and age structure of our visitor population.

Task: The user's task was to play the game Go-bang against the computer (five chips in a row win). This task requires a significant amount of mental planning activities on a strategic level. The syntax is quite simple: to put a chip on a free field. But, to finish a game needs attention and concentration for a couple of minutes (see Fig. 5).

Dependent Measures: The main *dependent measure* is the user's "winning chance" (the frequency of winnings) for each station separately. This frequency can be interpreted as a measure of the "quality" of the task result. The analysis of user's "winning chance" includes all games where the

computer would win in the next move, but in this state the game was canceled by the user (to avoid losing). We call these games "quasi winnings". The frequency--of real winnings and of quasi winnings--is a reliable and valid measure to compare precisely the different interaction techniques on the performance level. As *control measures*, the average computing time per move, the total playing time per game, and the number of moves per game was recorded and analysed.

Results: We analysed our data with the statistic tool StatView (vers. 4.02). We recorded automatically 9'006 contacts with in all 96'739 moves: 1'128 contacts with CI, 3'645 with MI, 2'881 with TI, and 1'352 with DPD. In all we observed 3'801 completed games. Each visitor had three different classes of behaviour to react to each station: (1) to pass; (2) to stand still and to observe what happens; and (3) to interact with one of the stations. For the last class of behaviour we could differentiate five separate actions: (A1) to start a new game even if the system is in the initial state; (A2) to cancel a game that was started --either by himself or by another user before; (A3) to finish a game with a draw; (A4) to finish a game with a winning result, either by the user or by the computer; and (A5) to finish a game with a "quasi winning" situation: a user canceled before the computer could win in the next move. These five actions are not independent of the interface type ($\chi^2 = 420$, $p \leq .0001$, see Tab. 1).

Tab. 1: Percentages of all observed user actions for each station separately. [CELL CONTENT: observed column percentage]

User's Action	CI	MI	TI	DPD
A1: start a game	43 %	49 %	60 %	54 %
A2: interrupt a game	8 %	2 %	4 %	13 %
A3: draw result	1 %	1 %	0 %	0 %
A4: completed games	38 %	41 %	30 %	29 %
A5: quasi winning	10 %	7 %	6 %	4 %
Total no. of actions:	1128	3645	2881	1352

For the TI and DPD interface type the two actions A1 and A2 could be significantly observed more often than expected (observed frequencies see Tab. 1; expected frequencies: A1 x TI: 52%; A1 x DPD: 52%; A2 x DPD: 5%). On the other side, in the CI and MI condition the two actions A4 and A5 were more often observed than expected (expected frequencies: A4 x CI: 35%; A4 x MI: 35%; A5 x CI: 7%). This result does not mean that the user had a better chance to win in the CI or MI condition than in one of both other conditions; it means only that for both conditions a game was more often completed, or nearly completed (the "quasi winnings"). This result can be interpreted in two ways: (1) a significant number of users had serious difficulties to detect the system in its initial state, caused by the ambiguous emoticon, the facial expression of the virtual player in the waiting state; and (2) users try to avoid the additional sound feedback after a lost game, caused by being afraid of the reactions of the people around.

To answer the question with which interface type the user had the highest chance to win, we have to take a closer

look to all winning games. The user's chance to win a game was 11% in all (428 user's winnings divided by 3'801 winning games overall). We coded each game result--from the user's point of view--as follows: "0" for *lost*, "1" for *draw*, and "2" for *winning*. The average over all winning games is significantly different between the four interface types (see Fig. 4; ANOVA, $DF_{residual}=3821$, $DF_{interface}=3$, $F=17.8$, $p \leq .0001$). The highest chance to win a game is given for the video based interface type of the Digital Playing Desk (see Fig. 4; the Scheffe results for DPD). Looking to all other group by group comparisons, only the command line interface CI is better than the mouse interface MI, too. Therefore we can rank the four interface types as follows: in the first position DPD, second [CI and TI], and third MI. How can we explain this important outcome?

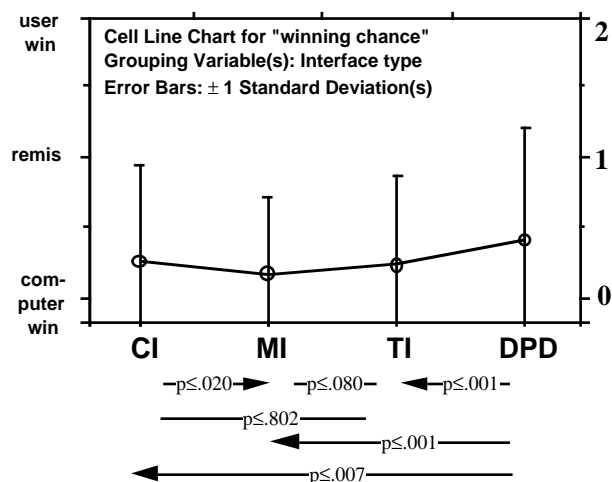


Fig. 4: The results per interface type for the dependent measure "winning chance" (below: the significance level of Scheffe tests).

The highest winning chance for DPD is *not* caused by different computing times: $TI_{time} = 55 \pm 32$ s versus $DPD_{time} = 56 \pm 31$ s, (ANOVA, $DF_{residual}=1488$, $DF_{interface}=1$, $F = 1.68$, $p \leq .195$). Therefore we can not assume that the DPD interaction technique performs best because an extended computing time gives the user more time to think about the next move than in the corresponding condition.

On the other side, we can observe a significant difference between all four interface types for the measure "playing time" (see Fig. 5). The shortest playing time can be found for MI 51 ± 60 s and--nearly--TI 63 ± 63 s. The longest playing time is given for DPD 216 ± 135 , followed by CI 168 ± 165 s. These differences are highly significant (see Fig. 5; ANOVA, $DF_{residual}=3821$, $DF_{interface}=3$, $F = 529$, $p \leq .0001$).

The outcome of the third control variable--the number of moves per game--is quite different to the outcome of the variable "playing time". We can observe a significant difference between the four interface types (see Fig. 6; ANOVA, $DF_{residual}=3821$, $DF_{interface}=3$, $F = 3.4$, $p \leq .01$). Only the number of moves for TI seems to be a little bit lower the average.

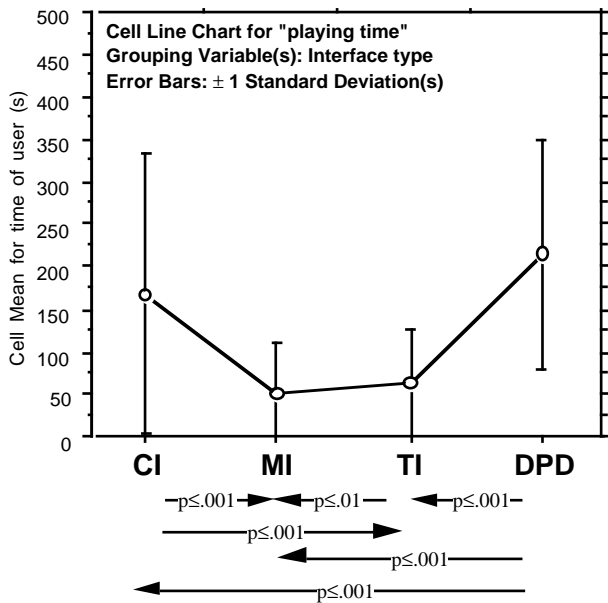


Fig. 5: The results per interface type for the dependent measure "playing time" (below: the significance level of Scheffe tests).

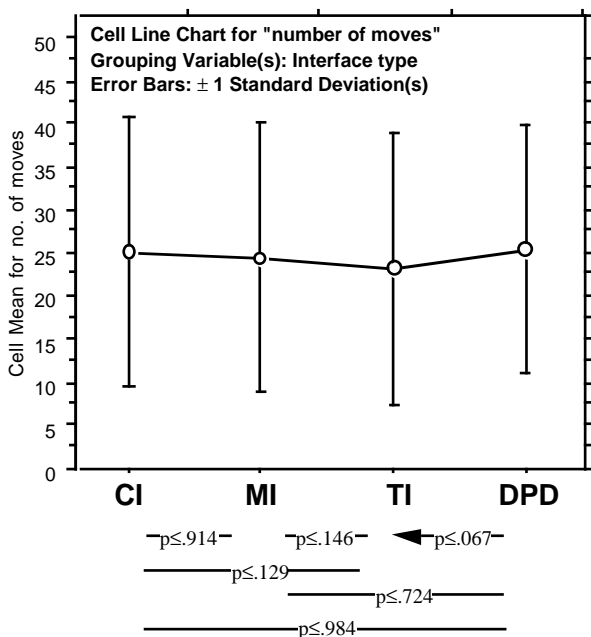


Fig. 6: The results per interface type for the dependent measure "number of moves" (below: the significance level of Scheffe tests).

Discussion of the field study: Informal observations during the fair show that several users of the DPD have difficulties to use this dialog technique without any kind of "starting" help: (1) to make the first move and to wait a moment, (2) not to leave one of the hands in the video controlled area; but, if a user overcomes these obstacles she or he was able to play without any further help. Now we can discuss the introduced advantages of graspable interfaces (see [2]). In the DPD interface we could observe two handed in-

teractions; there were no serious problems to move and remove chips with both hands. These two handed interactions enable more parallel input actions than with traditional interaction techniques.

The youngest DPD player was a four-year old girl carried by her father. This girl--well advised by her father--could interact without any serious problems. This easy to use effect is caused by the transfer of the very early and well-developed skills to manipulate physical artifacts.

We could also observe that users came together in larger groups--up to three and four persons--to collaborate and to discuss how to play and to move next. In one situation a circa ten years old boy was very proud to present his family that he was able to win.

Emoticons seem to be attractive especially for women. We could observe that several women tried to persuade their male partner to come to one of the emoticon stations, and to start playing. Most of male partners could be persuaded to start the game and to demonstrate their girl friends that it is totally harmless. At this moment the woman took over the leading role and played the game. This effect is in congruence with the criteria of "enjoyable interaction" [10].

The quantitative comparison between the four interface types is not quite fair, because the result for TI and DPD--compared with CI and MI--is superimposed by the effect of the virtual player's emoticons. The effect of the interaction technique itself and the effect of the emoticons superimpose each other. So, we have to take this effect into account.

On one side, we can try to estimate the effect size of the emoticons in our investigation. To differentiate between the influences of the interaction technique effect and the emoticon effect on the game outcome, we analyse the measure "user's chance to win" in the following way. The global average is 11%, so we have to divide the 9% advantage of the DPD technique (= 20% with DPD minus 11% global average) into two parts: (a) the emoticon effect of 3% (14% with MI plus DPD minus 11% global average), and (b) the interaction technique effect of 6% (= 20% with DPD minus 11% global average minus 3% emoticon effect). Overall, the DPD--adjusted by the emoticon effect--guarantees a significant increase of the user's performance.

On the other side, we can argue that the special design space for TI and DPD allows us to establish a new way of feedback: the virtual player. The complete--or nearly--horizontal work space of DPD and TI opens the vertical space in front of a user for additional use. The traditional screen blocks--caused by its upright position in front of the user--the free view to possible communication partners, as in an ordinary desk environment! Only the horizontal work space of a desk allows the user to communicate with other people in front of him.

4 General Discussion and Conclusion

The general advantage and disadvantage of immersive VR are the necessity to put the user into a complete modelled 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 (AR) recognises 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 a "Natural User Interface" (NUI).

A system with a NUI supports the mix of real and virtual objects in the same interaction space. As input it recognises *and* understands physical objects and humans acting in a natural way (e.g., object handling, 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, i.e. that the same modality is used for input *and* output (see [8]). For example, a projected item can be referred directly by the user for his or her nonverbal input behavior. Fig. 7 provides an overview how a system with a NUI could look like.

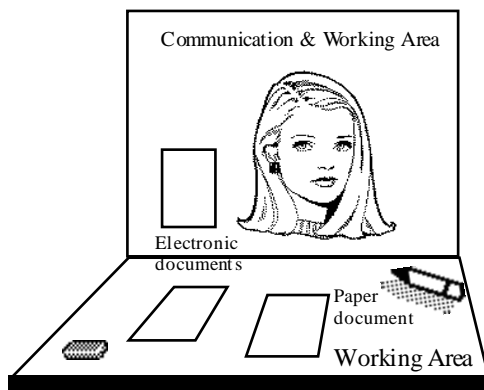


Fig. 7: The two different areas of a Natural User Interface

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 its internal 3D model of the user. A third close-up camera on the top, records permanently the content states 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 four loudspeakers, producing a spatial impression on the user. Free space in the communication area can be used for (content) work, too (see Fig. 7). Of course, traditional input and output devices still can be used in addition. As required by Tognazzini [11], NUIs are multimodal 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 human-computer interaction. 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!* (see [5]).

The outcomes of the field study presented here gives us a strong empirical evidence that a NUI with the possibility to interact with the computer with real world objects and with non-verbal feedback in form of emoticons has serious advantages over the traditional dialog techniques (e.g., command input, desktop with mouse, touch screen).

References

- [1] Arai, T., Machii, K., Kuzunuki, S. & Shojima, H. (1995) InteractiveDESK: A computer-augmented desk which responds to operations on real objects. In *Proc. of CHI'95 Companion*, ACM, pp. 141-142.
- [2] Fitzmaurice, G., Ishii, H. & Buxton W. (1995) Bricks: Laying the foundations for graspable user interfaces. In *Proc. of CHI'95*, ACM, pp. 442-449.
- [3] Krueger, M. W. (1991) *Artificial Reality II*. Addison.
- [4] Newman, W. & Wellner, P. (1992) A desk supporting computer-based interaction with paper documents. In *Proc. of CHI'92*, ACM, pp. 587-592.
- [5] Rauterberg, M. & Cachin, C. (1993) Locating the primary attention focus of the user. In: T. Grechenig & M. Tscheligi (Eds.) *Human Computer Interaction*. (Lecture Notes in Computer Science, Vol. 733, pp. 129-140), Springer.
- [6] Rauterberg, M., Dätwyler, M. & Sperisen, M. (1995) The shared social space as a basic factor for the design of group-ware. In: K. Brunnstein & P. Sint (Eds.) *Intellectual Property Rights and New Technologies*. (Schriftenreihe der Österreichischen Computer Gesellschaft Vol. 82, pp. 176-181), Oldenbourg.
- [7] Rauterberg, M. & Steiger, P. (1996) Pattern recognition as a key technology for the next generation of user interfaces. In: *Proceedings of IEEE International Conference on 'Systems, Man and Cybernetics' in Beijing (China)*, October 14 - 17, 1996.
- [8] Rauterberg, M. & Szabo, K. (1995) A design concept for N-dimensional user interfaces. In: *Proceedings of '4th International Conference INTERFACE to Real & Virtual Worlds'*. Montpellier: EC2 & C^{ie}, pp. 467-477.
- [9] Rheingold, H. (1991) *Virtual Reality*. Secker & Warburg.
- [10] Tamura, H. (1995) Human information technology for living oriented innovation. In: Y. Anzai, K. Ogawa & H. Mori (Eds.) *Symbiosis of Human and Artifact: Human and Social Aspects of Human-Computer Interaction*. (Advances in Human Factors/Ergonomics, Vol. 20A, pp. 121-124), Elsevier.
- [11] Tognazzini, B. (1996) *Tog on Software Design*. Addison.
- [12] Weiser, M. (1991) The computer for the twenty-first century. *Scientific American*, Sept., pp. 94-104.
- [13] Wellner, P. (1993) Interacting with paper on the DigitalDesk. In *Com. of the ACM*. 36(7), pp. 86-96.