MMI-Interaktiv Nr. 12, April 2007

Inhalt

Human-Machine Interaction in Virtual Environments – Recent Developments and Industrial Applications

JOHANN HABAKUK ISRAEL[1], ANJA NAUMANN[2]

[1]*Fraunhofer Institute for Production Systems and Design Technology (FhG IPK), Berlin, Germany*

[2]*Deutsche Telekom Laboratories, Technische Universität Berlin, Germany*

Virtual Reality (VR) has seen a major shift since its first public appearance in the beginning of the 1990s. What started as a vision, overloaded with expectations on how it would shape the way we will live and work has now found many applications mainly in the fields of medicine, assembly, design review, training, and also military applications. Many companies and academic research groups are working in the field, some on basic research and others developing systems which benefit from VR technology.

The picture of VR has been that of people wearing heavy Head Mounted Displays (HMDs) and cumbersome data gloves, leaving the user fully immersed and often lost in the virtual environment. While this is still the case in some fields, usability has also become an important issue in VR, for reasons of sheer survival. User acceptance for traditional wired VR boxes has been low. But as new technologies have emerged over recent years, which now allow lightweight systems that are easy to use, together with trends to merge real and virtual environments (Augmented and Mixed Reality, AR, MR) which include natural interaction, VR is spreading and entering new domains, such as tourism, business applications, and end user operating systems. The emerging field of Digital Human Modeling builds a bridge between VR and human factors engineering, applied ergonomics, and computer-aided engineering design, allowing questions to be addressed which relate to the early phases in the product development process.

VR does not necessarily mean immersive stereoscopic visualization. In fact, many emerging VR applications are desktop or table based and not stereoscopic. Thanks to the computer gaming industry, almost every personal computer now has a powerful graphics card so that it could act as a desktop VR instance. Cheap video projectors,

HMDs and digital humans will do the rest to bring VR into everyday live and work. This raises ethical issues of how we will live with or in virtual environments in the future. The ongoing discussion about the online VR game Second Life shows how relevant these questions are also for our social structure.

This special issue attempts to show different views of where we are with VR today and to present exemplary work and research in the field. In the first article, Bauckhage, Thurau, Gormand & Humphrys present a novel approach to naturally acting artificial agents in virtual environments. By applying machine learning techniques to recordings of network traffic of distributed games, they aim to generate "human-like" behavior of artificial characters. As the results show, their approach generates a significantly more human-like impression than traditional ones.

Herbon & Rötting describe mechanisms of detection and processing of visual information in three-dimensional space. They present a theoretical model of 3-D spatial interaction and give an overview of fundamental research on spatial attention and applied studies in the fields of aviation and car driving. These studies concern the location of information presentation in three-dimensional space for the completion of different types of tasks. Herbon and Rötting integrate this research in the model of 3-D spatial interactions and discuss resulting research challenges.

Naumann & Rötting discuss the use of Digital Human Modeling for design and evaluation of human-machine systems. They describe the advantages of the application of digital human models in product design and manufacturing and give an overview of current research on digital human modeling. Consequences of a wider application of digital human modeling in the area of Human Factors are being discussed. Up to now, the use of anthropometric models dominates. In contrast, the authors see the main research challenges in combining anthropometric and cognitive models.

A tabletop system for displaying large geographical scenes is introduced by Peinsipp-Byma, Eck, Bader & Geisler. It provides seamless integrated zoom functionality and is currently being used in the context of surveillance tasks. The seamless integration of physical tools and virtual imagery makes it a prominent example for non intrusive Mixed Reality systems.

This edition's community section has two articles. Amditis, Bimpas, and Blach describe the INTUITION network of excellence funded under the 6th EU Framework which aims towards networked and structured VR in industry. The prime objective of INTUITION is to promote and facilitate the development and application of VR/VE technology in industrial domains, establishing European excellence in this area. Amditis, Bimpas, and Blach give an overview of barriers to applying VR in industry, drivers for change, and resultant research challenges. The second community article by Israel gives a short summary of the 2007 IEEE VR conference and highlights main trends and developments.

We would like to thank all authors and reviewers for their contributions and their cooperation. All articles of this special issue are available online at www.useworld.net.

Johann Habakuk Israel & Anja Naumann

Learning Human Behavior from Analyzing Activities in Virtual Environments

C. Bauckhage[1], B. Gorman[2], C. Thurau[3] & M. Humphrys[2]

[1] *Deutsche Telekom Laboratories, Berlin, Germany*
[2] *Dublin City University, Dublin, Ireland*
[3] *Czech Technical University, Prague, Czech Republic*

Abstract

Present day multiplayer video games offer an interesting perspective for researching artificial cognitive systems. In this contribution, we focus on the problem of learning believable behavior models for artificial characters. Recordings of the network traffic of modern games allow for applying machine learning techniques to realize artificial agents that act more human-like than conventional current game characters. We detail an imitation learning approach and present the results of an extensive believability study that was carried out on the Internet.

1. Introduction

Computer- and video games have turned into an integral part of our popular culture. Given their short history, it occurs that video games must exert a deep fascination otherwise their success would be inexplicable. Indeed, thanks to the technical performance of current computing hardware, well selling genres like action games, adventure games and simulation games do nowadays create a haunting and engaging experience for the player. They are set in dynamic, atmospheric virtual worlds of high complexity and they display an astounding level of physical accuracy and graphical detail. Moreover, data transfer over local networks or the Internet enables sharing the game experience with other human players making it a lot more unpredictable and exciting.

This amazing state of the art in game technology correlates with the fact that game development has become big business. Although modern game development requires

considerable intellectual and financial efforts because it involves large teams of programmers, authors, designers and marketing specialists, the game business yields considerable revenues. Recent reports actually see the world market for video games and edutainment software rapidly closing in on $20 billion a year[1].

Apart from entertainment and financial gains, however, present day computer games also provide interesting perspectives for research in disciplines such as sociology, psychology, or computer science. In this contribution, we elaborate the latter claim. More specifically, we discuss benefits computer games might offer for *machine learning* and explore the problem of behavior learning for game characters.

Our interest in the topic arose from our background in artificial cognitive systems designed for dynamic real world settings. It occurred to us that observing human players performing tasks of different complexity in a virtual 3D world might provide new insights into intelligent behavior modeling. Accordingly, we were surprised to find that –even in modern games– rather old-fashioned ideas such as preprogrammed scripts, finite state machines, or tree searches dominate behavior programming. Of course, maturity does not imply ineptitude of a programming technique. But none of these methods is well known for their generalization capabilities. Consequently, common approaches to artificial intelligence (AI) for games may lead to ennui and frustration for experienced players. After some time of playing, the actions of computer controlled characters tend to appear artificial and lack the element of surprise human opponents would provide. If a human player acts in a way not envisaged by the game programmers, game characters simply appear to behave 'dumb' (Cass 2002).

This might be different if game characters (often called *gamebots*) were to learn from experienced human players. In fact, the idea of learning from demonstration to produce more human-like behavior is popular in cognitive systems research (cf. e.g. Schaal 1999). Until now, however, this research focused on autonomous machines intended for deployment in the physical world. This focus led to a situation where research aimed at behavior representation and learning still first and foremost struggles with issues arising from uncontrollable environmental dynamics and noisy sensors. Unlike present day robotics, however, virtual environments and computer games allow for actually concentrating on cognitive aspects of complex behaviors. While in robotics the problem of sensor noise widely prevents investigating *reactive*, *tactical*, and *strategic* decision making, computer games offer a less cumbersome avenue.

As a consequence, we find ourselves amongst a growing number of researchers who are discovering that game worlds provide challenges and opportunities for intelligent systems research. In contrast to most recent contributors, however, we pursue an approach of statistical machine learning rather than of deliberative AI. In the following, we will outline basic concepts from statistical learning and discuss how they may apply to human-like behavior modeling for virtual characters. Then, we shall survey related work in this area and present some of our results which were evaluated in an extensive online survey.

---

[1]See, for instance, a study released in 2004 by the British Dept. of Trade and Industries: http://www.dti.gov.uk/sectors/games/index.html.

2. Machine Learning and Video Games

The capability to learn from what we perceive and experience is essential for our everyday live. Just consider the fact that almost everything that constitutes our personality had to be learned at some point in our lives. When born, none of us knew how to walk, how to talk, or how to behave in public. As these examples indicate, learning enables flexibility and adaptation. Once we learned how to walk in our nursery, we were able to transfer this knowledge to other terrains. Thus, whenever we refer to learning in this paper, we are interested in the phenomenon generalizing from something known in order to act appropriately in a novel situation or to better perform in a familiar one.
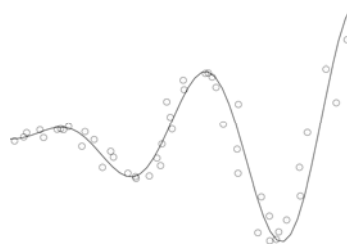
Furthermore, learning is based on examples. Without the analysis of exemplary input or role models there will be no extension of knowledge and capabilities. Given the importance of this mechanism, it is no surprise to find it to be innate and even 'hard-wired' into our brains. Especially if it comes to behavior learning, experiments in behavioral science document that already newborn infants endeavor to reproduce activities they observe in their surroundings (Rao & Meltzoff 2003). Recent neuro-physiological examinations even indicate that there are particular brain areas specialized in imitation (Kohler, Keysers, Umiltà, Fogassi, Gallese & Rizzolatti 2002).

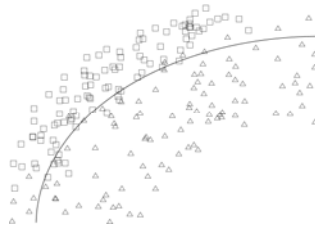### 2.1 Machine Learning and Pattern Recognition

Machine learning (ML) is an area of computer science that tries to mimic the flexible learning capabilities of the human brain. It deals with the development of algorithms that learn from examples and apply this knowledge in order to produce reasonable output if confronted with input they never saw before. Note, however, that even though the performance of an algorithm that has learned from examples rather depends on the analysis of data sets than on the intuition of engineers, human intuition cannot entirely be abandoned. The designer of a ML system still must specify the data to learn from as well as the method to analyze it.

Acquiring knowledge requires mechanisms to represent knowledge. Structural methods like rule bases or grammars, for instance, encode relations among pieces of *symbolic* information. *Statistical* machine learning, in contrast, deals with numerical data. The standard approach is to consider vectors $\mathbf{x} = [x_1, x_2, ... x_n]^T$ whose components encode numerical values of features that characterize certain entities. Given a training set of exemplary data, the task of a statistical ML system is to find mathematical functions which provide an abstract description of the examples. This is usually done by adapting the parameters of a given method for function approximation. Common such methods are Gaussian mixture models, neural networks, or support vector machines[2].
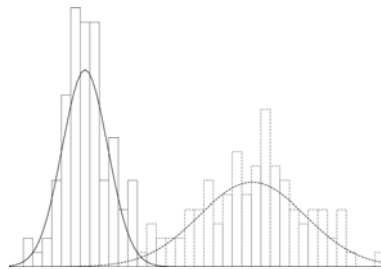
---

[2] For the sake of completeness, we should note that there also are hybrid machine learning techniques. So called *graphical models* such as Bayesian networks or Hidden Markov Models process numerical data but come along with an underlying graph structure with weighted edges. Suitable values for the weights are learned from examples.

(a) Regression

(b) Classification

(c) Density estimation

Figure 1: Three topics of statistical machine learning: 1(a) regression fits a function into a set of data points; 1(b) classification searches for boundaries between classes of data points; 1(c) density estimation determines how data points are distributed.

Figure 1 exemplifies what statistical ML may accomplish. Given a training set of pairs $\{(\mathbf{x}^{\alpha}, \mathbf{y}^{\alpha})\}$, $\alpha=1, ..., N$, the *regression* task tries fitting a function f to the data such that a new input $\mathbf{x}$ will yield the most plausible output $\mathbf{y} = f(\mathbf{x})$. More formally, if we assume the in- and output of the system to be random variables X and Y, respectively, the objective is to estimate the expected value $E(Y \mid X = \mathbf{x})$. A typical application for this would be time series forecasting in stock market analysis.

If the n dimensional vector space V^{n} of input data is partitioned into K different classes, one might want to know to which class an input vector $\mathbf{x}$ belongs. Instances of this problem are automatic speech recognition or object recognition in computer vision. For training, the *classification* task requires a set of pairs $\{(\mathbf{x}^{\alpha}, y^{\alpha})\}$ where $y^{\alpha} \in \{1,...,K\}$ denotes the class index of the pattern vector $\mathbf{x}^{\alpha}$. The goal is to learn a function $f : V^{n} \rightarrow \{1,...,K\}$ that partitions the input space and maximizes the probability $P(Y \mid X = \mathbf{x})$.

Finally, given a data set $\{\mathbf{x}^{\alpha}\}$, statistical ML can produce a functional description of the distribution $p(\mathbf{x})$ of the data. Applications of this task of *density estimation* can, for instance, be found in several data compression technologies.

Machine learning algorithms are often categorized with respect to the training data that is provided. *Supervised learning* characterizes algorithms that generate a function which maps inputs to desired outputs. The tasks of regression and classification would thus typically be dealt with by supervised learning techniques. Algorithms for *unsupervised learning* generate a model for a set of inputs; density estimation would hence be an example for unsupervised learning. In *reinforcement learning*, the algorithm itself creates pairs of input/output vectors and has to apply a trial-and-error strategy to determine whether they lead to a desired goal.

(a) 3D game environment

(b) Training data generation at a LAN party.

Figure 2: Complex 3D computer game worlds are popular among players.

2.2 How Does It Relate to Video Games?

Our tendency to imitate successful behavior certainly also apply to the way we learn to play a video game. Since imitating tricks and routines of experienced players leads to more success, learning from demonstration may also provide an avenue to programming engaging behaviors for computer game characters.

For the reminder of this contribution, we will consider behavior learning for the game QUAKE II® in which the player moves through a virtual 3D world (also called a

map) which (s)he perceives from the first person perspective (see Fig.2(a)). The map is loosely based on the physics of the real world. Players can move freely only constrained by the game physics. Though tactical variations exist, the player's task is to gain as many points as possible by battling other characters. In doing so, the player will loose health, armor, and ammunition but can compensate it by collecting corresponding items distributed all over a map. Items will reappear at the same position shortly after having been picked up. This induces strategies into game play. Winning will be facilitated by *item control*, which means moving through the map such that you secure the best items for yourself and leave the weaker ones for your adversaries.

Obviously, the state of a game character can be characterized by its current position and view on the map and its current armament and health conditions as well its distance to possible foes. If these features are thought of as components of a *state vector*, the current state of the character corresponds to a point in a high dimensional state space. The history of states a character assumes during a game will form a path in this state space. Neglecting a possible dependency on former actions and assuming the state of player p at time t to be given by a vector $\mathbf{x}_t^p$, a simple first order approximation of the player's state at the next time step $t+1$ could hence be modeled as $\mathbf{x}_{t+1}^p = F(\mathbf{x}_t^p, \mathbf{y}_t^p(\mathbf{x}_t^p), \mathbf{e}_t)$ where F is some unknown function, $\mathbf{e}_t$ denotes environmental influences at time t and $\mathbf{y}_t^p(\mathbf{x}_t^p)$ represents the action player accomplishes according to his current state. Restating this expression as $\mathbf{y}_t^p = f(\mathbf{x}_{t+1}^p, \mathbf{x}_t^p, \mathbf{e}_t)$ reveals that this model corresponds to what Arkin (1998) calls *reactive behaviors*. The actions of a player only depend on his or her state and on the current environmental influence. We also recognize that our model resembles the regression task in machine learning. Thus, given suitable training data, prototypical actions $\mathbf{y}^{p,t}$ or situated behaviors might be *learnable*. Since they simply correspond to sequences of actions $\{\mathbf{y}_{t_i}^p, \mathbf{y}_{t_{i+1}}^p, ..., \mathbf{y}_{t_{i+n}}^p\}$, techniques like neural networks, support vector machines, or Bayesian learning may apply.

Since a demo contains recordings of the network traffic of a multiplayer game, it encodes the series of states $\mathbf{x}_0^p, \mathbf{x}_1^p, \mathbf{x}_2^p, ..., \mathbf{x}_T^p$ the recording player p underwent during a game. It also includes information about nearby items and other players as well as temporary entities. There is no need for a visual analysis of a game scene, since all necessary information is already available on a cognitive higher level. The same applies to the player actions; they are included as simple velocity and position vectors.

3. Related Work

During the past two years, we could witness an increased academic interest in the problem of believable computer game characters. One of the driving factors behind this interest was already mentioned in the introduction and has also been noted by authors such as Cass (2002) or Nareyek (2004): on the one hand, commercial games still mainly rely on well seasoned, deliberative AI techniques like finite state machines or game trees. On the other hand, statistical machine learning as a tool to produce believably acting game agents has been largely neglected by the scientific community. This, however, seems to be changing.

Recent work by Spronck, Sprinkhuizen-Kuyper & (2003) introduced reinforcement learning to the task of rule selection for agent behavior in a commercially available role playing game. Earlier, the same authors reported on a hybrid coupling of genetic algorithms and neural networks for offline learning in a simple strategy game (Spronck, Sprinkhuizen-Kuyper & 2002).

The idea of using human generated data to train game agents was first reported by Sklar, Blair, Funes & (1999) who collected the key-strokes of people playing *Tron* in order to train neural networks. Just recently, Le Hy, Arrigioni, Bessière & (2004) described advanced probabilistic action selection for a commercial game using Bayesian networks which are trained by means of human generated input.

Next, we will summarize some of our own results in using machine learning techniques for producing human-like bot behavior for modern video games.

4. Imitation Learning

In this section, we outline our current approach to imitating human movement and strategic behavior in QUAKE II®. The model discussed by Gorman, Thurau, Bauckhage & (2006) focuses on two core aspects of human behavior; *strategic planning* and *motion modeling*. Several investigations (Laird 2001, Livingstone 2006) have found that the ability of an agent to exhibit long-term strategic planning faculties is a crucial factor in determining how human-like its behavior appears. The importance of *motion modeling* is equally evident because human players frequently exhibit actions other than simply moving along the environment surface. In many cases, the player can only attain certain goals by performing one or more such actions at the appropriate time; they therefore have an important *functional* element. From the perspective of creating a believable agent, it is also vital to reproduce the *aesthetic* qualities of movements and activities.

4.1 Learning Goal-Oriented Strategic Behaviors

In QUAKE II®, experienced players roam the environment methodically, controlling important areas of the map and picking up *items* to strengthen their character. Thus, a player's long-term goal can be seen in systematically collecting items found at certain points of a map. By learning the mappings between the player's status and his subsequent item pickups, the agent can adopt observed strategies when appropriate, and *adapt* to situations which the player did not face.

We first read the set of all player locations $\mathbf{l}=[x,y,z]$ from the recording, and cluster them to produce a *goal-oriented* discrimination of the level's topology. We also construct an $n{\times}n$ matrix of edges E, where n is the number of clusters, and $E_{i,j}=1$ if the player was observed to move from node i to node j and 0 otherwise. The player's *inventory* –the list of what quantities of which items he currently possesses– is also read from the demo and unique state vectors are obtained; these *inventory prototypes* represent the varying situations faced by the player during a game. We can now construct a set of *paths* which the player followed while in each such situation.

Having obtained the different paths pursued by the player in each inventory state, we turn to reinforcement learning to learn his or her behavior. The topological map of

the game environment may now be viewed as a *Markov Decision Process* (MDP), with the clusters corresponding to states and the edges to transitions. In this scenario, the MDP's actions are considered to be the choice to move to a given node from the current position. Thus, the transition probabilities are $P(c' = j \mid c = i, a = j) = E_{ij}$ where c is the current node, c' is the next node, a is the executed action, and E is the edge matrix. We assign an increasing reward to consecutive nodes in every path taken under each prototype, such that the agent will be guided along similar paths to the human when facing similar situations.

To model player's intuitive *weighing* of strategic objectives, and his understanding of *object transience*, we introduce a weighted *fuzzy clustering* approach and an *item activation* variable $m_p(\mathbf{s})$. Its membership distribution implicitly specifies the agent's current goals, which will later facilitate integration with the Bayesian motion-modeling system. The final utilities thus result from:

$$U(c) = g^{e(c)} \sum V_p(c) m_p(\mathbf{s}), \quad c_{t+1} = \max_y U(y), \quad y \in \{x \mid E_{c,x} = 1\} \tag{1}$$

where $U(c)$ is the final utility of node c, γ is the discount, $e(c)$ is the number of times the player has entered cluster c, $V_p(c)$ is the original value of node c in state prototype p, and E is the edge matrix.

4.2 Bayesian Motion Modeling

It is not sufficient to simply identify the player's goals and the paths along which (s)he moved to reach them; it is also necessary to capture the actions executed by the player in pursuit of these goals. Here, we apply a Bayesian inverse-model for action selection in infants and robots due to Rao & Meltzoff (2003). The choice of action at each time step is expressed as a probability function of the subject's current position c_t, next position c_{t+1} and goal c_g:

$$P(a_t \mid c_t, c_{t+1}, c_g) = \frac{P(c_{t+1} \mid c_t, a_t) P(a_t \mid c_t, c_g)}{\sum_u P(c_{t+1} \mid c_t, a_u) P(a_u \mid c_t, c_g)}$$

$$\tag{2}$$

This model fits into the strategic navigation system almost perfectly; the clusters c_t and c_{t+1} are chosen by examining the utility values, while the current goal state is implicitly defined by the membership distribution. In order to derive the probabilities, we read the sequence of actions taken by the player as a set of vectors $\mathbf{v}$. We then cluster these action vectors to obtain a set of *action primitives*, each of which amalgamates a number of similar actions performed at different times into a single unit of behavior.

Several important adaptations must be made in order to use this model in the game environment. Firstly, Rao's model assumes that transitions between states are instantaneous, whereas multiple actions may be performed in QUAKE II®while moving between successive clusters; we therefore express $P(c_{t+1} \mid c_t, a_t)$ as a soft-distribution of all observed actions on edge $E_{ct,ct+1}$ in the topological map. Secondly, Rao assumes a single unambiguous goal, whereas we deal with multiple weighted goals in parallel. We thus perform a similar weighting of the probabilities across all

active goal clusters. Finally, Rao's model assumes that each action is independent of the previous action. In QUAKE II[®], however, each action is constrained by the action performed on the preceding time step; we therefore introduce an additional dependency in our calculations. The final probabilities are computed as follows:

$$\sum_g m_g P(a_t | c_t, c_{t+1}, c_g) \frac{P(a_t | a_{t-1})}{\sum_u P(a_u | a_{t-1})}$$

(3)

5. Believability Testing

Still, there exists no standard method of gauging the 'believability' of game bots. Given that one of the central aims of our work lies in improving this quality of such agents, we need to address this shortcoming. The most obvious means of determining the degree to which agents are perceived as human is to conduct a survey. This, of course, raises questions of subjectivity, experimenter influence, and so on. In order to produce a credible assessment of agent believability, any proposed system must be designed with these concerns in mind. Our aims, then, are as follows: i) to construct a framework which facilitates rigorous, objective testing of the degree to which game agents are perceived as human; ii) to formulate a *believability index* expressing this 'humanness', and allowing comparisons between different agents.

To counteract any potential observer bias, our test takes the form of an anonymous Internet-based survey (see Fig. 3 for a screenshot of one of the forms). Respondents are presented with detailed instructions covering all aspects of the test. They are not asked for personal data such as age or gender, but are required to estimate their gaming experience at one of five different levels:

1. Never played, rarely or never seen
2. Some passing familiarity (played / seen infrequently)
3. Played occasionally (monthly / every few months)
4. Played regularly (weekly)
5. Played frequently (daily)

Upon proceeding to the test itself, respondents are presented a series of pages, each of which contains a group of video clips. Each group shows similar sequences of game play from the perspective of the in-game character. Within each group, the clips may depict any combination of human and artificial players; the respondent is required to examine the behavior of the character in each clip, and indicate whether (s)he believes it is a human or artificial player. The clips are marked on a gradient, as follows: 1: Human, 2: Probably Human, 3: Don't Know, 4: Probably Artificial, 5: Artificial

This rating is the central conceit of the survey and will later be used to compute the believability index. Additionally, the respondent may provide an optional comment explaining his/her choice. In cases where (s)he indicates that (s)he believes the agent to be artificial, (s)he will be further asked to rate how "human-like" (s)he perceives its behavior to be, on a scale of 1 to 10. This more subjective rating is not involved in the computation of the believability index, but may be used to provide additional insight into users' opinions of different agents.
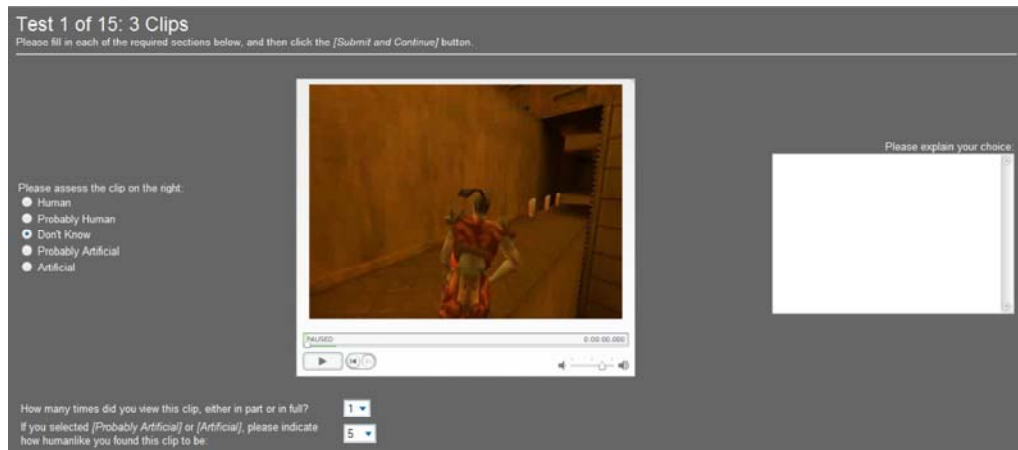
Figure 3: Extract from the believability test questionnaire.

5.1 Subjectivity, Bias and Other Concerns

Aside from the observer effect, there are several areas in which the potential for subjectivity and the introduction of bias exist. Since our aim is to provide an objective measure of believability, these must be eliminated or minimized.

The first obvious pitfall lies in the selection of video clips. The selector may deliberately choose certain clips in an effort to influence the respondents. To guard against this, we first ensure that the number of samples is sufficient to embody a wide variety of behaviors, and secondly, we cede control of the selection of the specific behaviors to an unbiased arbiter. In our case, we wished to compare the believability of our imitation agents against both human players and traditional rule-based bots; thus, we first ran numerous simulations with the traditional agent –over whose behavior we had no control– to generate a corpus of game play samples, and then proceeded to use human clips embodying similar movements and activities both in the believability test and to train our imitation agents.

Similarly, the order in which the videos are presented could conceivably be used to guide the respondents' answers. To prevent this, we randomize the order in which the groups of clips are displayed to each user, as well as the sequence of clips within each page; the test designer thus has no control over the order of the samples seen by the user.

Another issue concerns the possibility that users will choose the 'Probably' options in a deliberate effort to artificially minimize their error and 'beat' the test, or that they will attempt to average out their answers over the course of the survey – that is, they may rate a clip as 'human' for little reason other than that they rated several previous clips as 'artificial', or vice-versa. To discourage this, we include notes on the introduction page to the effect that the test does not adhere to any averages, that the user's ratings should be based exclusively upon their perception of the character's behavior in each clip, and that the user should be as definitive as possible in their answers. A related problem is that of user fatigue; as the test progresses, the user may begin to lose interest, and will consequently invest less effort in each successive clip. We address this by including a feature enabling users to save their progress at any point, allowing them to complete the survey at their convenience.

It is also imperative to ensure that the test is focused upon the variable under investigation – namely, the believability of the agent's movement and behavior. As such, the survey must be structured so as not to present 'clues' which might influence the respondents. For instance, the tester should ensure that all clips conform to a standard presentation format, so that the respondent cannot discern between different agents based on extraneous visual cues. To this end, we run a script over the demo files to remove all such indicators. In the resulting clips, all agents are rendered using the same model, they are given the same name, and the display perspective is homogenized to a common point of view. In the specific case of our imitation agents, this requirement that all extraneous indicators be removed raises a conflict between two of our goals in conducting the survey. If the players in two of the three clips we use on each page begin from the same location and exhibit near-identical behavior, the respondent may conclude through pure logical deduction that (s)he is probably viewing a human and imitation agent, and consequently that the remaining clip is more likely to be a traditional artificial agent. Note that this might not necessarily be true, but even an incorrect answer based on factors other than believability will adversely affect the accuracy of the results. We circumvent this problem by training imitation agents with different (but similar) samples of human game play to those actually used in the test. The resulting clips are therefore comparable, but do not 'leak' any additional information; respondents must judge whether or not they are human based solely on their appearance. At the same time, however, we obviously wish to test how accurately our agents can capture the aesthetic appearances of their human exemplars. To satisfy both requirements, a small minority of imitation agents are trained using the same human data as presented in the survey; in the experiments described below, 2 of the imitation agents were direct clones, while the remainder were trained on different data.

5.2   Evaluation of Results

Before evaluating the results of the survey, one should ensure that there have been a substantial number of responses with a decent distribution across all experience levels; a good 'stopping criterion' is to run the test until the average experience level is at least 3 (i.e. a typical, casual games player). Standard analyzes (precision, recall, etc) can be carried out on the results; however, as mentioned earlier, we also wish to formulate a believability index which is specifically designed to express the agent's believability as a function of user experience and the certainty with which the clips were identified.

Recall that each clip is rated on a scale of 1 (definitely human) to 5 (definitely artificial). Obviously, the true value of each clip is always either 1 or 5. Thus, we can express the degree to which a clip persuaded an individual that the visualized character was human as the normalized difference between that person's rating and the value corresponding to 'artificial':

$$h_p(c_i) = \frac{|r_p(c_i) - A|}{\max(h)} \tag{4}$$

where $h_p(c_i)$ is the degree to which person p regarded the clip as depicting a human, $r_p(c_i)$ is person p's rating of clip i, A is the value on the rating scale which corresponds to 'artificial', and $max(h)$ is the maximum possible difference between a

clip's rating and the value of 'artificial'. In other words, $h_p(c_i)$ will be 0 if the individual identified a clip as artificial, 1 if he identified it as human, and somewhere in between if he chose one of the 'Probably' or 'Don't Know' options. We now weight this according to the individual's experience level:

$$w_p(c_i) = \frac{e_p h_p(c_i)}{\max(e)} , \quad (5)$$

where e_p is the experience level of person p and $\max(e)$ is the maximum experience level. Thus, the believability index is conditioned upon a sufficient level of expertise among respondents; if their average experience level is 1, for instance, then their responses will be weighted into insignificance and the believability index will be correspondingly low. Finally, we sum the weighted accuracies across all clips and respondents, and take the average:

$$b = \frac{\sum_{p}^{n} \sum_{i}^{m} w_p(c_i)}{nm} , \quad (6)$$

where b is the believability index, n is the number of individual respondents, and m is the number of clips. The believability index is, in essence, a weighted representation of the degree to which a given type of clip was regarded as human. In the context of the survey, then, a 'good' result for an AI agent would involve a high value of b for both the agent and human clips.

6. Experiments

The main purpose of the experiment described in this section was to examine how believable our imitation agents were in comparison with human players and traditional rule-based artificial agents. It consisted of 15 groups of video clips, with 3 clips in each; these clips were, on average, approximately 20 seconds in length. We first ran numerous simulations involving the rule-based artificial agent to derive a set of game play samples, and then used similar samples of human players both in the test itself and to train our imitation agents. The rule-based agent used was the QUAKE II® Gladiator bot, which was chosen due to its reputation as one of the best bots available.

With the video clips in place, the URL of the survey site was distributed to the mailing lists of several colleges in Ireland and Germany. After a one-week test period, we had amassed a considerable number of responses. After discarding incomplete responses, we were left with 20 completed surveys, totaling 900 individual clip ratings.
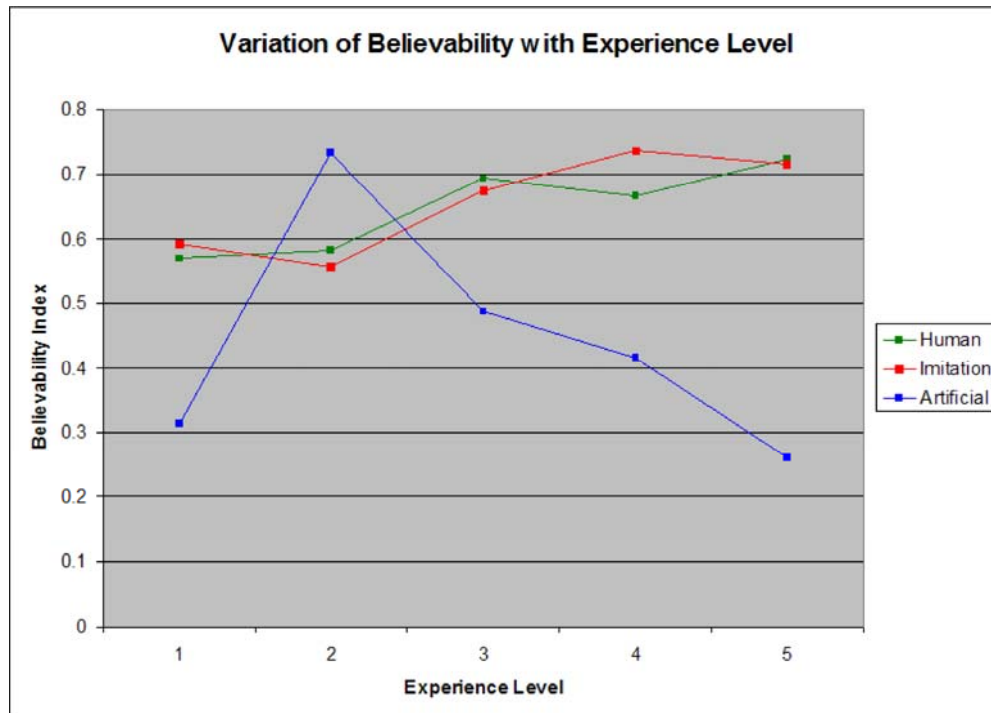
Figure 4: Variation of perceived believability with experience level.

The survey produced a very favorable impression of our imitation agents compared to the artificial agent. The believability indices for human, imitation and traditional artificial clips were 0.69, 0.69 and 0.35, respectively. In other words, the imitation agents were misidentified as human 69% of the time, while the rule-based agents were mistaken as human in only 35% of cases (weighted according to experience). Clips which actually *did* depict human players were also identified 69% the time. Essentially, it seems that respondents were generally unable to discern between the human players and our imitation agents. These results are corroborated by the recall values, which indicate that both the human and imitation clips were classified as human in approximately 68% of cases, while the rule-based agent was classified as human only 36.69% of the time. Since the human sources used to train the imitation agents were different than those human clips presented as part of the test, this implies that the results are based on the general abilities of the imitation mechanism, rather than any factors unique to the clips in question.

Further indication of the imitation agents' effectiveness is evident in the graph of believability against experience level shown in Fig. 4. While an in-depth psychological explanation of the curves displayed there is beyond the scope of our work, it is noticeable that, as the experience level rises, respondents correctly identify human clips as human more frequently, and misidentify the traditional agent as human less frequently. The identification of imitation agents as human, by contrast, closely parallels that of genuine human clips. These trends may be explained by the fact that more experienced players have a greater knowledge of characteristically human behaviors – smooth strafing, unnecessary jumping, pausing to examine the environment, and similar idiosyncrasies – which the traditional agent would not exhibit, but which would be captured and reproduced by the imitation bots.

In summary: the results of the believability study suggest that our imitation agents exhibit far greater 'humanness' than even a well-regarded rule-based agent, and indeed are comparable to genuine human players. We consider this to be strong evi-

dence in support of our original premise; namely, that imitation learning has the potential to produce more believable game agents than traditional AI techniques.

7. Conclusion

In this paper, we considered virtual computer game worlds as a testbed that allows for studying the problem of modeling human behavior. We reviewed concepts in statistical machine learning and described one of our own approaches to behavior learning from human generated data. Also, we proposed a formal method of quantifying the degree to which different agents are perceived as 'human-like', in the form of a web-based survey and an objective metric based on both the respondents' level of experience and the accuracy with which the players/agents were identified. Through our experiments, we showed that our imitation-learning approach produces game bots which are capable of conveying a significantly more human-like impression than traditional rule-based agents, and are often almost indistinguishable from genuine human players.

References

Arkin, R. C. (1998). *Behavior-Based Robotics*. Cambridge, MA: MIT Press.

Cass, S. (2002). Mind Games. *IEEE Spectrum, 39,* (12), 40–44.

Gorman, B.; Thurau, C.; Bauckhage, C. & Humphrys, M. (2006). Believability Testing and Bayesian Imitation in Interactive Computer Games. In: *Proc. Int. Conf. on Simulation of Adaptive Behavior*. Berlin, Germany: Springer Verlag.

Kohler, E.; Keysers, C.; Umiltà, M.; Fogassi, L.; Gallese, V. & Rizzolatti, G. (2002). Hearing Sounds, Understanding Actions: Action Representation in Mirror Neurons. *Science*, 297, 846–848.

Laird, J. E. (2001). Using a Computer Game to develop advanced AI. *IEEE Computer, 34,* (7), 70–75.

Le Hy, R. ; Arrigioni, A. ; Bessière, P. & Lebeltel, O. (2004). Teaching Bayesian behaviours to video game characters. *Robot. Auton. Syst.*, 47 (2–3), 177–185.

Livingstone, D. (2006). Turing's test and believable AI in games. *Comput. Entertain.*, *4,* (1), 6.

Nareyek, A. (2004). Artificial Intelligence in Computer Games – State of the Art and Future Directions. *ACM Queue*, 1 (10), 58–65.

Rao, R. & Meltzoff, A. (2003). Imitation learning in infants and robots: Towards probabilistic computational models. In: *Proc. AISB 2003 Convention: Cognition in Machines and Animals, Aberystwyth, UK*. Society for the Study of Artificial Intelligence and the Simulation of Behaviour.

Schaal, S. (1999). Is Imitation Learning the Route to Humanoid Robots? *Trends Cogn. Sci.*, 3 (6), 233–242.

Sklar, E.; Blair, A.; Funes, P. & Pollack, J. (1999). Training intelligent agents using human internet data. In: *Proc. 1st Asia-Pacific Conference on Intelligent Agent Technology*. Hong Kong: Hong Kong Baptists University.

Spronck, P.; Sprinkhuizen-Kuyper, I. & Postma, E. (2002). Improving Opponent Intelligence through Machine Learning, In: *Proc. GAME-ON*. London, UK: EOROSIS.

Spronck, P.; Sprinkhuizen-Kuyper, I. & Postma, E. (2003). Online Adaptation of Game Opponent AI in Simulation and in Practice. In: *Proc. GAME-ON*. London, UK: EUROSIS.

Detection and Processing of Visual Information in Three-Dimensional Space

ANTJE HERBON & MATTHIAS RÖTTING

Center of Human-Machine-Systems, Berlin University of Technology, Germany

Keywords: three-dimensional, depth, spatial interaction, spatial cognition, visual information presentation, visual space, attention, task-specificity

Abstract

This paper reviews literature on visual information presentation in depth. A theoretical model of 3-D spatial interactions by Previc (1998) is presented. Fundamental research on spatial attention and applied studies in the fields of aviation and automobile driving are described, which were concerned with the task-specificity of information presentation in space. Necessary research is identified and an experimental set-up for intended future studies is presented.

1. Introduction

When Human Machine Interaction (HMI) in virtual or augmented reality is concerned, frequently asked questions are: How are objects to be implemented? How do users view the environment? How will they interact, i.e. what kind of tools are to be used? The question of *where* in three-dimensional space, especially in depth, information should be presented or interaction should take place has been rather neglected, excepting aviation and automobile driving. In those fields of application, many researchers have been concerned with the utilization of either head-up displays (HUDs) or head-down displays (HDDs) for information presentation (e.g. Foyle, Sanford & McCann 1991, Horrey, Alexander & Wickens 2003, Liu & Wen 2004).

This paper will present evidence, which indicates that the matter of *where* information should be presented in depth, is worth investigating. First, a neuropsychologically based model of spatial interaction by Previc (1998) will be introduced. Secondly, results from empirical studies examining attentional issues in depth in both,

laboratory and field settings, will be presented. Finally, open research questions are discussed and an experimental set-up for future studies is presented.

2. A theoretical model of 3-D spatial interactions

When considering information presentation in three-dimensional space, it is of importance to make assumptions on whether detection and processing of information differ according to location in space. Integrating neuropsychological work on three-dimensional spatial interaction, Previc (1998) suggested the division of the space surrounding us into four realms (see figure 1). Each realm has specific properties concerning neurological processing of stimuli and reactions, influencing human perception and behavior.
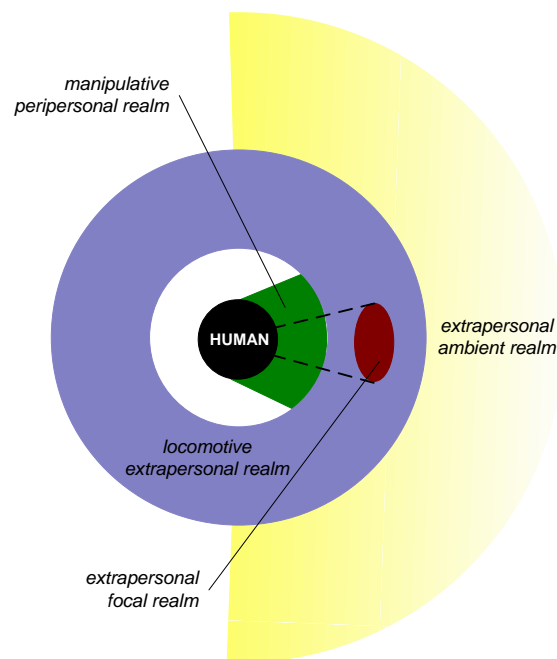

Figure 1: A theoretical model of 3-D spatial interactions, adapted from Previc (1998)

The first realm is the peripersonal or – extended by May (2006) – the manipulative peripersonal realm. It extends from zero to two meters into depth and is used for visual grasping and object manipulation. Its lateral extent is 60° (central). The peripersonal realm is to be understood as the realm inside hand-reaching distance, even though it has been shown to be able to expand with tool use, when the manipulable region expands therewith (Berti & Frassinetti 2000). As far as visual perception is concerned, this realm is specialized for global form, depth and motion. There is a lower-field bias, meaning that attention is more readily allocated to and manual reaction times are faster in the lower field (Sheliga, Craighero, Riggio & Rizzolatti 1997).

The second realm is the extrapersonal focal realm. It radially (i.e. in depth) extends from 0.2 meters to the distance at which a respective object is no longer resolvable, while the lateral extent is 20-30° (central)[1]. It is mainly used for visual search and object recognition, which entails that areas and objects of interest at a certain loca-

---

[1] Note that the red ellipse represents a focused object/position, while the dashed line depicts the lateral border of the realm.

tion inside this realm are focused for identification and classification. As far as visual perception is concerned, this realm is specialized for color and form, including high-resolution contour analysis. There is an upper field bias.

The third realm is the extrapersonal action or – as May (2006) called it – locomotive extrapersonal realm. It extends from two to 30 m or more in the full 360°. Positions in this realm can be reached by walking or other movements of the body. It is mainly used for navigation, target orientation (incl. motion-based actualization of object location) and scene memory. There is an upper field bias.

The fourth realm is the extrapersonal ambient realm. It extends up to a few kilometers and is mainly used for spatial orientation, postural control and locomotion beyond the locomotive extrapersonal realm. Its lateral extent is 180° and there is a lower field bias.

This model of three-dimensional space has been examined and supported mainly by neuropsychological studies (e.g. McCourt and Garlinghouse 2000; Weiss et al. 2000; see Halligan, Fink, Marshall and Vallar (2003) for a review). Studies in the field of cognitive psychology are rather scarce. However, literature from HMI-research, especially aviation and automobile driving reports many studies on information presentation in different depth-planes. There, focal and ambient vision are mainly differentiated and can alternatively be interpreted as peripersonal, focal extrapersonal or ambient extrapersonal realms most of the time. Some of those studies are being presented in the following.

3. Empirical Literature on Visual Attention in Depth

There are quite a number of laboratory studies which were concerned with the spread of attention in depth. Drive for most of this research was the question of whether attention allocation differs depending on the number of dimensions (two or three) that are concerned. However, just a few experiments directly examined allocation of attention in (and between) peripersonal and extrapersonal space.

3.1 Fundamental research on attention allocation in depth

It has been shown that attention does indeed spread in depth and is not limited to two-dimensional space: Atchley, Kramer, Andersen and Theeuwes (1997) found that reaction times were slower when subjects had to switch attention in x-, y-, and z-dimension (depth) than only in x- and y-dimensions. However, this effect was only apparent when distractors, i.e. increased perceptual load, were present.

Many researchers (e.g. Gawryszewski, Riggio, Rizzolatti & Umilta 1987; Andersen & Kramer 1993; Kimura, Miura, Doi & Yamamoto 2002) reported an asymmetry in the allocation of attention in depth where subjects were able to switch attention faster from far to near objects than from near to far objects. Arnott and Shedden (2000) found that this viewer-centered asymmetric depth gradient is dependent on perceptual load, i.e. it is not apparent when perceptual load is low because in this case a narrow attentional focus is not necessary.

As in two-dimensional space, attention in three-dimensional space can be either object- or space-based; a trade-off between the two possibilities is suggested (Atchley & Kramer 2001).

3.2   Peripersonal and extrapersonal space

Couyoumdjian, di Nocera and Ferlazzo (2003) conducted three experiments on allocation of attention within and between peripersonal and extrapersonal space. They presented four pairs of LED-cubes, one on the left and one on the right, at 40, 80, 120 and 160 cm from the observer and cued their onset either validly (at the same location) or invalidly (at a different location). They found that reaction times were significantly faster when invalidly cued and target locations were in the same realm than when subjects had to shift their attention across realm boarders, distances between cue light and target cube being equal. Results remained consistent when fixation point and target distances where manipulated also. These findings strongly suggest that – on top of the time needed to switch attention between two points in depth – there is an added cost when attention has to be switched between two perceptual realms.

In spite of this supportive evidence on perceptional issues, there have been contradictive results on the behavioral part of the model of 3-D space presented above. Schoumans, Kappers and Koenderink (2002) could not find any differences in a pointing task in 40 and 120 cm distance from their subjects. Moreover, they replicated systematic context-based errors in both distances. However, it would be of interest if using the pointer lead to an extension of the manipulative peripersonal realm and if effects were due to the fact that the two distances were actually part of one instead of two realms.

3.3   Applied studies in aviation and automobile driving

With the technical maturation of head-up displays (HUDs), research on information presentation in aviation and automobile driving has increased greatly. HUDs provide the opportunity of moving information from displays inside a vehicle to the windshield, resulting in reduced eyes-off-the-road time. Comparative research on HUDs and conventional head-down-displays (HDDs) has lead to the assumption that the optimal location of information presentation is task-specific.

Summarizing previous simulator studies, Horrey and Wickens (2004) suggested, that certain combinations of multiple tasks associated with operating a vehicle can be time-shared more efficiently than others. They proposed that the reason for this difference is that some tasks utilize focal, whereas other tasks utilize ambient vision. The multiple resources model (Wickens 2002) would then predict that two focal tasks will interfere and lead to poorer performance while a focal and an ambient task can be completed in parallel. Tasks utilizing ambient vision thereafter are lane-keeping and speed control, whereas hazard detection utilizes focal vision. This finding is very much in line with Previc's (1998) model, where the extrapersonal ambient realm is mainly used for spatial orientation and the extrapersonal focal realm is very sensitive for object recognition.

However, this generalization can be problematic. Liu and Wen (2004) for example showed in a goods delivery task with commercial vehicle operators, that reacting to

urgent events and *speed keeping* require focal vision and do therefore interfere with a side-task that requires focal vision also. This discrepancy to Horrey's and Wickens' findings can be explained by the specifics of the presentation of the information that was needed in order to complete the task: Speedometer information was presented in form of a number on an HUD or an HDD and could therefore not be perceived using ambient vision, since this high-resolution form identification utilizes focal vision. Thus, speed-maintenance using information on the outside-world moving by (as in Horrey, Alexander & Wickens 2003) is very different from speed-maintenance as operationalized by Liu and Wen. The importance of this distinction was also apparent in a study conducted by Foyle, Sanford & McCann (1991), who had subjects complete a flight-task and presented altitude information necessary for an altitude-maintenance task in two different ways: as digital information on an HUD and as indirect source of information using sketched buildings super-imposed on the sides of the flight path. Results indicated that reading digits utilizes focal, while using the sketched buildings as height information utilizes ambient vision. In this study, altitude-maintenance was competitive to path-keeping, which most interestingly utilized focal vision. However, this is likely to be explained by the curvature of the path, because workload has been shown to modulate the allocation of visual resources (Horrey, Alexander & Wickens 2003).

Recently, Crawford and Neal (2006) have reviewed selected literature dealing with perceptual and cognitive issues associated with HUDs in aviation. They identified cognitive tunneling as one of the main attentional problems in the use of HUDs, which causes an impairment of the pilots' ability to detect events outside their vehicle because of their attention being captured by the information on the windshield. This is especially important when two tasks both depending on focal vision are competing for limited resources. Levy, Foyle and McCann (1998) found that linking HUD-symbology to the outside world (i.e. displaying information as if it were located on the flight path, for example, instead of on the windshield) can solve this problem by directing attention to a different depth. In their experiment, it did not matter where exactly on the path the symbology – in this case an analog gauge – was located. Performance on the main focal task was always better than in the traditional HUD-situation, where information was projected directly onto the windshield without additional depth information.

The empirical literature cited above gives strong evidence for an allocation of attention in depth. Task-type, which supposedly influences attention has been investigated in applied HMI-studies, but not yet in fundamental research. However, in order to generalize findings to other different fields where Augmented Reality (AR) technology is utilized, further research on task-specificity in attention allocation is needed. For example, AR is widely-used in production (e.g. Reiners, Stricker, Klinker & Müller 1998; Sarval, Baker & Filipovic 2005), has been utilized for navigation (e.g. Biocca, Tang, Owen & Fan 2006) and has found its way into Smart Homes (e.g. Hammond, Sharkey & Foster 1996; Intille 2002). The Tangible Media Group at MIT Media Lab have designed a number of applications for social interaction (Chang, Resner, Koerner, Wang & Ishii 2001; Bonanni, Vaucelle, Lieberman & Zuckerman 2006), sports (Ishii, Wisneski, Orbanes, Chun & Paradiso 1999), infotainment (Ishii & Ullmer 1997; Ishii 2004) and work (Ishii, Wisneski, Brave, Dahley, Gorbet, Ullmer & Yarin 1998). Tasks in these applications differ a lot, but might still be classified according to their requirements on information detection and processing.

4. Task-Specificity

Lacey and Lacey (1970) have described different stressors they used in physiological studies, which produced task-specific physiological response patterns. Mental arithmetic, reversed spelling, making up sentences and noxious stimulation lead to an increase in heart rate and heart rate variability, while attending to photoic flashes, white noise or a dramatic recitation resulted in a decrease of both parameters. The authors characterized the reaction to the two groups of stimuli as *rejection* and *intake of the environment*, respectively. In the first group with cardiac acceleration, tasks require internal cognitive elaboration (calculating, putting together letters and words) and in the case of the cold pressure test (noxious stimulation) simply the suppression of the unpleasant feeling. Perception of the environment is therefore not necessary and not wanted, thus the environment is *rejected*. In the second group with cardiac deceleration, tasks require attention on visual and auditory stimuli in the environment, i.e. the environment is *taken in*.

This classification can be compared to Norman's (1993) notion on experiential and reflective cognition. Tasks that are completed in experiential mode require environmental intake. A person has to perceive and/or react to stimuli from the outside world. Responses in this mode of cognition are automatic and not reflected upon. By contrast, reflective cognition utilizes conscious engagement in the task, e.g. planning or decision making. Once input from the outside world is attained, environmental intake is no longer needed but instead hinders and detracts from the problem solving process.

5. Intended Research

Task-specificity of attention allocation in depth is to be examined further. A main research goal is to draw implications as to where to present information in augmented and virtual realities independent of a specific application-field. Lacey's and Lacey's (1970) and Norman's (1993) notions on task classification will form a basis for experimental task design.

Previc's model on 3D-spatial-interaction will be used in order to define perceptual realms in an experimental set-up. It has been shown to be in line with previous experimental laboratory findings on allocation of attention in depth and can be used to interpret HMI-research results also. Even though cited applied-studies were not based on the model, they did give evidence for the division of the space surrounding us into different perceptual realms.

An experiment is currently being designed to investigate coherences between task-characteristics as described above (tasks with different requirements on visual perception as in the applied studies, reflective vs. experiential tasks) and perception of information in depth. As shown in figure 2, different tasks will be presented in two different depths, one on a transparent screen close to the observer and a second one on a far projection screen, several meters away from the observer.
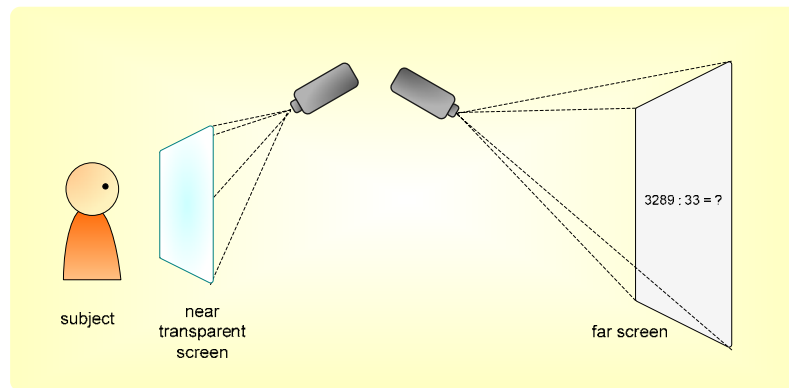
Figure 2: Intended experimental set-up

Suitable tasks will be identified and classified according to their requirements on visual perception and reflective thought. Elements from all groups of tasks will then be presented in both depths. Parameters to be inspected are reaction times, error rates and workload. Additional collection of physiological parameters, such as eye movement for investigation of attention allocation or stress-level indicators such as skin resistance or heart rate is also possible.

Acknowledgements

References

Andersen, G.J. & Kramer, A.F. (1993). Limits of focused attention in three-dimensional space. *Perception & Psychophysics*, 53 (6), 658-667.

Arnott, S.R. & Shedden, J.M. (2000). Attention switching in depth using random-dot autostereograms: Attention gradient asymmetries. *Perception & Psychophysics*, 62 (7), 1459-1473.

Atchley, P. & Kramer, A.F. (2001). Object and space-based attentional selection in three-dimensional space. *Visual Cognition*, 8 (1), 1-32.

Atchley, P.; Kramer, A.F.; Andersen, G.J. & Theeuwes, J. (1997). Spatial cuing in a stereoscopic display: Evidence for a „depth-aware" attentional focus. *Psychonomic Bulletin & Review*, 4 (4), 524-529.

Berti, A. & Frassinetti, F. (2000). When Far becomes Near: Re-mapping of Space by Tool Use. *Journal of Cognitive Neuroscience,* 12 (3), 415-420

Bonanni, L.; Vaucelle, C.; Lieberman, J. & Zuckerman, O. (2006). PlayPals: Tangible Interfaces for Remote Communication and Play. *Proceedings of CHI '06*.

Chang, A.; Resner, B.; Koerner, B.; Wang, XC & Ishii, H. (2001). LumiTouch: An Emotional Communication Device. *Extended Abstracts of Conference on Human Factors in Computing Systems*.

Couyoumdjian, A.; di Nocera, F. & Ferlazzo, F. (2003). Functional representation of 3D space in endogenous attention shifts. *The Quarterly Journal of Experimental Psychology,* 56A (1), 155-183.

Crawford, J. & Neal, A. (2006). A Review of the Perceptual and Cognitive Issues Associated With the Use of Head-Up Displays in Commercial Aviation. *The International Journal of Aviation Psychology*, 16 (1), 1-19.

Foyle, D.C.; Sanford, B.D. & McCann, R.S. (1991). Attentional Issues in Superimposed Flight Symbology. *Proceedings of the Sixth International Symposium on Aviation Psychology*, 577-582.

Gawryszewski, L.D.G.; Riggio, L.; Rizzolatti, G. & Umilta, C. (1987). Movement of attention in the three spatial dimensions and the meaning of "neutral" cues. *Neuropsychologia*, 25 (1A), 19-29.

Halligan, P.W.; Fink, G.R.; Marshall, J.C. & Vallar, G. (2003). Spatial cognition: evidence from visual neglect. *TRENDS in Cognitive Sciences*, 7 (3), 125-133.

Hammond, J.C.; Sharkey, P.M. & Foster, G.T. (1996). Integrating augmented reality with home systems. *Proc. 1st Euro. Conf. Disability, Virtual Reality & Assoc. Tech.*

Horrey, W.J.; Alexander, A.L. & Wickens, C.D. (2003). Does Workload Modulate the Effect of In-Vehicle Display Location on Concurrent Driving and Side Task Performance? *DSC North America Proceedings*.

Horrey, W.J. & Wickens, C.D. (2004). Focal and Ambient Visual Contributions and Driver Visual Scanning in Lane Keeping and Hazard Detection. *Proceedings of the Human Factors and Ergonomics Society 48th annual meeting*.

Ishii, H. (2004). Bottles: A Transparent Interface as a Tribute to Mark Weiser. *IEICE Transactions on Information and Systems*, E87-D (6), 1299-1311.

Ishii, H. & Ullmer, B. (1997). Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms. *Proceedings of CHI '97*.

Ishii, H.; Wisneski, C.; Brave, S.; Dahley, A.; Gorbet, M.; Ullmer, B. & Yarin, P. (1998). ambietROOM: Integrating Ambient Media with Architectural Space. *Proceedings of CHI '98*.

Ishii, H.; Wisneski, C.; Orbanes, J.; Chun, B. & Paradiso, J. (1999). PingPongPlus: Design of an Athletic-Tangible Interface for Computer Supported Cooperative Play. *Proceedings of CHI '99*.

Intille, S. S. (2002). Designing a Home of the Future. *IEEE Pervasive Computing*, 1, 76-82.

Kimura, T.; Miura, T.; Doi, S. & Yamamoto, Y. (2002). *Top-down and Bottom-up allocation of attention in three-dimensional space when observers are moving forward*. Technical Report on Attention and Cognition No. 16, http://L.u-tokyo.ac.jp/AandC/

Lacey, J.I. & Lacey, B.C. (1970). Some Autonomic-Central Nervous System Interrelationships. In: P. Black (ed.). *Physiological Correlates of Emotion*, 205-227. New York: Academic Press.

Liu, Y.-C. & Wen, M.-H. (2004). Comparison of head-up display (HUD) vs. Head-Down-Display (HDD): driving performance of commercial vehicle operators in Taiwan. *International Journal of Human-Computer Studies*, 61, 679-697.

May, M. (2006). Raum. In: J. Funke & P.A. Frensch (eds.). *Handbuch der Allgemeinen Psychologie - Kognition*, 66-74. Göttingen: Hogrefe.

McCourt, M.E. & Garlinghouse, M. (2000). Asymmetries of visuospatial attention are modulated by viewing distance and visual field elevation: Pseudoneglect in peripersonal and extrapersonal space. *Cortex*, 36, 715-731.

Norman, D.A. (1993). *Things that make us Smart*. Reading, MA: Addison-Wesley Publishing Company.

Previc, F.H. (1998). The Neuropsychology of 3-D Space. *Psychological Bulletin*, 124 (2), 123-164.

Reiner, D.; Stricker, D.; Klinker, G. & Müller, S. (1998). Augmented Reality for Construction Tasks: Doorlock Assembly. *Proceedings of the 1st International Workshop on Augmented Reality (IWAR'98).*

Sarval, A.; Baker, C. & Filipovic, D. (2005). Head-Worn Display-Based Augmented Reality System for Manufacturing. *Proceedings of SPIE 2005*.

Schmidt, L.; Wiedenmaier, S.; Oehme, O. & Luczak, H. (2005). Benutzerzentrierte Gestaltung von Augmented Reality in der Produktion. In: C. Stary (ed.). *Mensch & Computer 2005: Kunst und Wissenschaft – Grenzüberschreitungen der interaktiven ART*, 51-60. München: Oldenbourg Verlag.

Shoumans, N.; Kappers, A.M.L. & Koenderink, J.J. (2002). Scale invariance in near space: pointing under influence of context. *Acta Psychologica*, 110, 63-81.

Sheliga, B.M.; Craighero, L.; Riggio, L. & Rizzolatti, G. (1997). Effects of spatial attention on directional manual and ocular responses. *Experimental Brain Research*, 114, 339-351.

Weiss, P.H.; Marshall, J.C.; Wunderlich, G.; Tellmann, L.; Halligan, P.W.; Freund, H.-J.; Zilles, K. & Fink, G.R. (2000). *Brain*, 123, 2531-2541.

Wickens, C.D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3 (2), 159-177.

Digital Human Modeling for Design and Evaluation of Human-Machine Systems

Anja Naumann[1] & Matthias Roetting[2]

[1] *Deutsche Telekom Laboratories, Technische Universität Berlin*
[2] *Department of Human-Machine Systems, Technische Universität Berlin*

Keywords: Digital Human Modeling, Human-Machine Interaction, Human Factors, Usability Testing, Ergonomics, Virtual Prototyping

Abstract

Digital Human Modeling is an emerging area that bridges computer-aided engineering design, human factors engineering and applied ergonomics. Especially for ergonomics questions, digital humans are already being used. The paper describes different fields for the application of digital human modeling in product design, assembly and manufacturing. Current research on digital human modeling in training, risk assessment of workplaces, hand-object interaction, modeling eye movement, and the assessment of comfort is highlighted. The case is made for a combination of anthropometric models and cognitive models. In addition, the consequences of a wider application of digital human modeling in product design, assembly and maintenance are discussed, especially with regard to the education of engineers and human factors specialists.

1. Why Use Digital Humans in Engineering?

Countless organizations in a variety of industries are facing the same problem: the human element is not being considered early or thoroughly enough in the design, assembly and maintenance of products. More importantly, this is having a devastating impact on cost, time to market, quality and safety. However, for a growing number of organizations, factoring the human element into design, manufacturing and maintenance is no longer a problem but rather a competitive advantage. Using digital human models realizes benefits such as shorter design time, lower development costs, improved quality, and enhanced productivity.

Digital human modeling enables engineers in product development to address questions of ergonomics and human factors early on in the development process. At the same time, digital human modeling reduces the need for the production of real prototypes and can even make it obsolete (Cappelli & Duffy 2006).

In research and development, commercial human models are already being used. These models are up to now mainly restricted to anthropometric issues. Two often used human models are JACK and RAMSIS (Duffy 2006):

- JACK enables users in industry to position biomechanically accurate digital humans of various sizes in virtual environments, assign them tasks, and analyze their performance. The digital humans JACK (and his female counterpart JILL) can tell engineers what they can see and reach, how comfortable they are, when and why they're getting hurt, when they're getting tired and other important ergonomic information. This information helps organizations design safer and more effective products (see also http://www.ugs.com/Products/tecnomatix/human_performance/jack/) faster and for less cost.

- RAMSIS helps manufacturers and engineering services providers to do substantial design studies during the design phase. The core functions of this software are the realistic display of international anthropometric data and the efficient analysis of ergonomic questions concerning, for example, sight, maximum force, reachability, and comfort. RAMSIS is already used by more than 60% of all automotive manufacturers worldwide for the ergonomic design and analysis of passenger compartments and work places (see also http://www.human-solutions.com/automotive_industry/).

In the following, some examples for the use of digital humans in product design and manufacturing will be given.

1.1 Digital Humans in Product Design

During the product design phase, organizations face several key challenges related to the physical attributes and behavior of humans. They need to develop products centered on humans and evaluate designs based on ergonomic factors, account for different sizes and shapes of people, and consider human factors in the design before building physical prototypes. Human simulation allows, for example, questions concerning positioning and comfort to be answered:

- Comfort: Is the design optimized for the comfort of the envisioned user groups?
- Visibility: What can people of different sizes see when they operate a piece of equipment or a vehicle?
- Ingress and egress: Can the target population easily climb in and out of the equipment or vehicles?
- Reaching and grasping: Are controls placed so that everyone can reach and operate them?
- Multi-person interaction: How do multiple people interact with the product?
- Strength assessment: Does operating the product require inordinate strength or create the potential for injury?

1.2 Digital Humans in Manufacturing

Organizations face several key challenges related to the physical attributes and behavior of humans in manufacturing. They need to bring factories on-line faster, to optimize manual workflow, to improve worker safety, and to reduce training costs. In the manufacturing phase of the product lifecycle, human simulation allows the following questions to be answered:

- Work cell layout: Are machines and other equipment positioned to optimize cycle time and avoid hazards?
- Workflow simulation: Are the manufacturing processes designed to eliminate inefficiencies and ensure optimal productivity?
- Assembly accessibility: Can all assembly personnel access the parts and equipment needed to assemble the product?
- Safety analysis: Can all of the assembly tasks be performed safely?
- Lifting: Do all lifting tasks fall within the strength guidelines?
- Energy expenditure: How much energy will workers expend over time as they perform repetitive tasks?
- Simulation based training: How can real-time simulation and virtual reality be used to train workers?

2. Current Research on Digital Human Modeling

Digital human modeling is an emerging area that bridges computer-aided engineering design, human factors engineering and applied ergonomics. It is increasingly getting attention from research and development (Cappelli & Duffy 2006).

Up to now, most research efforts have concentrated on anthropometric aspects of human modeling. Allen, Curless, and Popovic, for example, work on improving models of body shapes. They examined body deformation during movements (Allen, Curless, & Popovic 2002) on the one hand and individual variations in body shapes on the other hand (Allen, Curless, & Popovic 2004).

Besides the continuous optimization of measurement methods and the technical aspects of 3-D-Modeling, one particular research topic is the verification and validation of digital human models. Methods and procedures are thus being developed which make sure that the human models used lead to reliable results. One example is the development and application of digital human models for training purposes, for example, for the development of the diagnostic competence of medical students. In one project, haptic virtual models of the spine serve as educational aids and this especially at early stages of the study of medical science (Howell, Williams, Burns, Eland, & Conatser 2006; Chen, Williams, Conatser, & Howell 2006). One further research topic here is motion analysis for the ergonomic assessment of risky work places or dynamic tasks (e.g. lifting of heavy items; Cappelli & Duffy, 2006).

A relatively new area concerns the modeling of hands and their interaction with objects. An effective ergonomic evaluation often requires a realistic simulation of the hand-object interaction and a reliable estimation of performance. The aim is to use digital mock-ups and digital hand models for usability tests in the early phases of

system development. This saves costs and delivers results almost as realistic as from tests with real users (Endo, Kanai, Kishinami, Miyata, Kouchi, & Masaaki 2006). In this way, for example, usage performance, usage durations, and the handling (grasp) qualities of PDAs, mobile phones, and MP3 players can be measured (Endo et al. 2006). In addition, a judgment of shape, color, texture, keys or jog-dials of a mobile device can be made (e.g., DhaibaHand; Miyata, Kouchi, & Mochimaru 2006). Also, based on motion prediction, it can be estimated which elements can be and which can not be reached by finger motion (Grasp Quality Index; Yang, Pitarch, Kim, & Abdel-Malek 2006).

Modeling eye movement deals with the development of adequate models of human vision for the prediction of human behavior using new concepts and technologies. It addresses (amongst other things) basic aspects which are connected to eye movements and the reaction of the person concerned (e.g., worker or operator) to environmental conditions. In this context, examinations are possible into how reading rate, eye movement parameters, light adaptation, planning of eye movements, visibility of objects, and perception of depth and movements are correlated to human movement and posture. Kim et al., for example, examined the role of visual and manual requirements when planning movements (Kim, Martin, Dukic, & Hanson 2006), and the prediction of movements when moving the head (Kim, Martin, & Gillespie 2006). For visibility analyses, commercial digital human models proved to be useful. For example, JACK (see 2.) was used successfully in combination with a CAD model of a car for testing the visibility of a child beneath a car. The deviation of the simulation results from a real situation with a real car and a paperboard model of the child were minimal (Ruspa 2006).

An important research area in the automotive, and more generally in the transportation, sector is comfort and discomfort. Both, drivers and passengers of cars and airplanes expect a high level of comfort. Human-comfort models have thus been developed which can replace test panels for comfort tests (Reynolds & Wehrle 2006; Bidal, Bekkour, & Kayvantash 2006; Parakkat, Pellettier, Reynolds, Sasidharan, & El-Zoghbi 2006). Another important area of research is the modeling of motor behavior and motion sequences, for example for assessing a vehicle, a work place or a task (Monnier, Renard, Chameroy, Wang, & Trasbot 2006; Reed, Faraway, Chaffin, & Martin 2006). Research questions in this context concern, for example, ingress/egress (Cherednichenko, Assmann, & Bubb 2006), driving motions (Parkinson & Reed 2006; Rider, Chaffin, & Martin 2006), accessibility in passenger compartments (Wang, Chevalot, Monnier, & Trasbot 2006), step motions (Wagner, Reed, & Chaffin 2006), and sitting behavior (Wirsching, Junker, & Nitzsche 2006).

First steps towards a combination of anthropometric models and cognitive modeling can be found in the modeling of human behavior. The models used contain representations of the processes how people think and behave. These representations are especially relevant for the planning of operational processes in complex human-machine systems. In recent years, cognitive and behavioral models have become more and more meaningful and allow for human behavior to be modeled at a high level. Using such models, the simulation of autonomous virtual humans in a virtual environment is possible (e.g., Delleman 2006; Rasmussen, Christensen, Siebertz, & Rausch 2006). Thereby, cognitive aspects are already beginning to be considered (e.g., Narkevicius, Bagnall, Sargent, & Owen 2006; Gore & Milgram 2006). In addition, attempts in implementing multitasking models have been undertaken (e.g.,

Albeck & Badler 2006). Anthropometric models show already a good performance - almost comparable to real humans. However, the combination of cognitive and anthropometric models is not yet powerful enough to be applied and implemented in industry.

3. Prospects of Digital Human Modeling in Human Factors

That digital human modeling is increasingly being addressed by research and development (Cappelli & Duffy 2006), becomes apparent, for example, by the fact that the First International Conference on Digital Human Modeling will be held as a part of Human-Computer Interaction International conference (HCII) in Beijing in 2007 (see http://www.hcii2007.org/home.html).

In addition, in 2006 a new journal titled *International Journal of Human Factors Modeling and Simulation* has been established (see https://www.inderscience.com/www/IJHFMS_leaflet.pdf). This new journal focuses on the development and use of computer simulations and computational algorithms to advance knowledge and understanding in the field of human factors. It puts a particular focus on human factors theory as related to computational models of human performance and interaction with virtual environments, simulator-based evaluations of human factor issues, computer simulation of human behavior and performance, digital human modeling and simulation, developments in simulation and virtual environments to address human factors and ergonomic issues, and the simulation of physiological behavior, measures, and predictions.

Also, a *Handbook of Digital Human Modeling* edited by Vincent G. Duffy, will be published (Duffy 2007). This book will reflect the multidisciplinary perspective required to participate effectively in research and applications in this area (Duffy 2006). The digital human modeling effort will focus on six main areas including the background to digital human modeling, modeling fundamentals, evaluation methods, tools, applications and future work. There are many open questions for researchers and practitioners since there is a tremendous need for the practitioners, especially in ergonomics and human factors engineering, to be able to integrate digital human modeling tools into their everyday work (Duffy 2005). An understanding about the conditions under which certain models can be applied is also important, as well as the respective limitations.

Up to now, human factors specialists have already been included in design teams, but typically they are likely to suggest that a new design cannot work well, without having enough knowledge or tools to suggest possible modifications to make the new design acceptable in terms of human factors and ergonomics (Meister and Enderwick 2001). The digital human modeling community appears to be providing access to a computer-based set of tools to enable an improved contribution from both ergonomists and human factors engineers in engineering design. Attempts are being made to gather the best understanding of the fundamentals, tools and applications in Digital Human Modeling. The results will be made available to researchers and practitioners in applied ergonomics and human factors engineering that may participate in engineering design teams (Duffy 2005). One example for this integration is the Society of Automotive Engineers Digital Human Modeling Conference (SAE-DHM, see http://www.sae.org/events/dhm/). It will be important in the future that the information from this group can be transferred to those involved in

human factors engineering. Thereby they will have access to computer-based tools to provide a common platform for their interaction with engineers in other disciplines participating in those design teams. It is also important that the driving forces in this area focus on the scientific fundamentals of this emerging multi-disciplinary field. Human factors and ergonomics is still not required in most engineering curricula. Digital human modeling initiatives provide an opportunity for researchers in human factors engineering and ergonomics to provide additional scientific support. There could be course offerings for undergraduate or graduate students such as *Digital Human Modeling* and *Virtual Interactive Machine Design*. These courses, when taught with a cross-disciplinary perspective, can be of interest to students, for example, in Industrial Engineering, Biomedical Engineering, Mechanical Engineering, Electrical Engineering, and Computer Science as well as Cognitive Science (Duffy 2005). It is clear that without access to the digital mock-ups that include the human aspects of models in advance of workstation and product design, costs will be incurred later, either in retrofit solutions or in lost market opportunities (Duffy 2005).

Until now, applied human models address mainly anthropometric issues. However, in recent decades, the nature of work has changed so that physical demands on workers and users have gotten less and, at the same time, cognitive demands have become more important. This requires new tools when designing products and workstations and consequently will provide more opportunities for cognitive models to be more frequently applied in the workplace and in product design. Thus, the main research challenge currently lies in the development of a combination of cognitive and anthropometric models. It is clear that the three worlds of Digital Human Modeling, Human Factors Engineering and Applied Ergonomics will have great opportunities for synergy (Duffy 2005).

References

Allbeck, J.M. & Badler, N.I. (2006). Automated Analysis of Human Factors Requirements. *Digital Human Modeling for Design and Engineering Conference,* Lyon, 4-6 July 2006. Warrendale: SAE International.

Allen, B.; Curless, B. & Popovic, Z. (2004). Exploring the space of human body shapes: data-driven synthesis under anthropometric control. *Digital Human Modeling for Design and Engineering Conference*, Rochester, MI, June 15-17. Warrendale: SAE International.

Allen, B.; Curless, B. & Popovic, Z. (2002). Articulated body deformation from range scan data. *ACM Transactions on Graphics (ACM SIGGRAPH 2002), 21,* 3, 612-619.

Bidal, S.; Bekkour, T. & Kayvantash, K. (2006). Scaling, Positioning and Improving a Human Model. *Digital Human Modeling for Design and Engineering Conference,* Lyon, 4-6 July 2006. Warrendale: SAE International.

Cappelli, T.M. & Duffy, V.G. (2006). Motion Capture for Job Risk Classifications Incorporating Dynamic Aspects of Work. *Digital Human Modeling for Design and Engineering Conference,* Lyon, 4-6 July 2006. Warrendale: SAE International.

Chen, M.-Y; Williams, R.L.; Conatser, R.R. Jr. & Howell, J.N. (2006). The Virtual Movable Human Upper Body for Palpatory Diagnostic Training. *Digital Human Modeling for Design and Engineering Conference,* Lyon, 4-6 July 2006. Warrendale: SAE International.

Cherednichenko, A.; Assmann, E. & Bubb, H. (2006). Computational Approach for Entry Simulation. *Digital Human Modeling for Design and Engineering Conference,* Lyon, 4-6 July 2006. Warrendale: SAE International.

Delleman, N. (2006). About Avatars and Maneuvring in Virtual Environments. *Digital Human Modeling for Design and Engineering Conference,* Lyon, 4-6 July 2006. Warrendale: SAE International.

Duffy, V.G. (2005). *Digital human modelling for applied ergonomics and human factors engineering.* Retrieved from: http://cyberg.wits.ac.za/cb2005/method2.htm [06/12/15].

Duffy, V.G. (2006). Inaugural address. *Digital Human Modeling for Design and Engineering Conference,* Lyon, 4-6 July 2006. Warrendale: SAE International.

Duffy, V.G. (Ed.). (to be published December 15, 2007). Handbook of Digital Human Modeling: Research for Applied Ergonomics and Human Factors Engineering. London: CRS Press.

Endo, Y.; Kanai, S.; Kishinami, T.; Miyata, N.; Kouchi, M. & Masaaki, M. (2006). An Application of a Digital Hand to Ergonomic Assessment of Handheld Information Appliances. *Digital Human Modeling for Design and Engineering Conference,* Lyon, 4-6 July 2006. Warrendale: SAE International.

Gore, B.F. & Milgram, P. (2006). The Conceptual Development of a Time Estimation Model to Predict Human Performance in Complex Environments. *Digital Human Modeling for Design and Engineering Conference,* Lyon, 4-6 July 2006. Warrendale: SAE International.

Howell, J.N.; Williams, R.L.; Burns, J.M.; Eland, D.C. & Conatser, R.R. Jr. (2006). The Virtual Haptic Back: Detection of Compliance Differences. *Digital Human Modeling for Design and Engineering Conference,* Lyon, 4-6 July 2006. Warrendale: SAE International.

Kim, K.H.; Martin, B.J. & Gillespie, B. (2006). Posture and Motion Prediction: Perspectives for Unconstrained Head movements. *Digital Human Modeling for Design and Engineering Conference,* Lyon, 4-6 July 2006. Warrendale: SAE International.

Kim, K.H.; Martin, B.J.; Dukic, T. & Hanson, L. (2006). The Role of Visual and Manual Demand in Movement and Posture Organization. *Digital Human Modeling for Design and Engineering Conference,* Lyon, 4-6 July 2006. Warrendale: SAE International.

Meister, D. & Enderwick, T.P. (2001). *Human factors in system design, development and testing.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Miyata, N. ; Kouchi, M. & Mochimaru, M. (2006). Posture Estimation for Screening Design Alternatives by Dhaiba Hand – Cell Phone Operation. *Digital Human*

Modeling for Design and Engineering Conference, Lyon, 4-6 July 2006. Warrendale: SAE International.

Monnier, G., Renard, F., Chameroy, A., Wang, X., & Trasbot, J. (2006). A Motion Simulation Approach Integrated into a Design Engineering Process. *Digital Human Modeling for Design and Engineering Conference,* Lyon, 4-6 July 2006. Warrendale: SAE International.

Narkevicius, J.; Bagnall, T.M.; Sargent, R.A. & Owen, J.E. (2006). Networking Human Performance Models to Further Utility of Process Models: The SEAPRINT Approach. *Digital Human Modeling for Design and Engineering Conference,* Lyon, 4-6 July 2006. Warrendale: SAE International.

Parakkat, J.; Pellettier, J.; Reynolds, D.; Sasidharan, M. & El-Zoghbi, M. (2006). Quantitative Methods for Determining U.S. Air Force Crew Cushion Comfort. *Digital Human Modeling for Design and Engineering Conference,* Lyon, 4-6 July 2006. Warrendale: SAE International.

Parkinson, M.B & Reed, M.P. (2006). Considering Driver Balance Capability in Truck Shifter Design. *Digital Human Modeling for Design and Engineering Conference,* Lyon, 4-6 July 2006. Warrendale: SAE International.

Rasmussen, J.; Christensen, S.T.; Siebertz, K. & Rausch, J. (2006). Posture and Movement Prediction by Means of Muscoloskeletal Optimization. *Digital Human Modeling for Design and Engineering Conference,* Lyon, 4-6 July 2006. Warrendale: SAE International.

Reed, M.P.; Faraway, J.; Chaffin, D.B. & Martin, B.J. (2006). The HUMOSIM Ergonomics Framework: A New Approach to Digital Human Simulation for Ergonomic Analysis. *Digital Human Modeling for Design and Engineering Conference,* Lyon, 4-6 July 2006. Warrendale: SAE International.

Reynolds, H.M. & Wehrle, J.H. (2006). ERL Seat Design – Occupied Specifications for the Unoccupied Deliverable Seat. *Digital Human Modeling for Design and Engineering Conference,* Lyon, 4-6 July 2006. Warrendale: SAE International.

Rider, K.A.; Chaffin, D.B. & Martin, B.J. (2006). Development of Active Human Response Model to Ride Motion. *Digital Human Modeling for Design and Engineering Conference,* Lyon, 4-6 July 2006. Warrendale: SAE International.

Ruspa, R. (2006). *A Validation Study of the Jack Manikin in a Visibility Task.* Presentation at SAE International Conference, Lyon, 4-6 July 2006.

Wagner, D.W.; Reed, M.P. & Chaffin, D.B. (2006). A Task-Based Stepping Behavior Model for Digital Human Models. *Digital Human Modeling for Design and Engineering Conference,* Lyon, 4-6 July 2006. Warrendale: SAE International.

Wang, X. ; Chevalot, N.; Monnier, G. & Trasbot, J. (2006). From Motion Capture to Motion Simulation: An In-vehicle Reach Motion Database for Car Design. *Digital Human Modeling for Design and Engineering Conference,* Lyon, 4-6 July 2006. Warrendale: SAE International.

Wirsching, H.-J.; Junker, J. & Nitzsche, F. (2006). An Integrated and Evaluated Simulation of Human-Seat Interaction. *Digital Human Modeling for Design and Engineering Conference,* Lyon, 4-6 July 2006. Warrendale: SAE International.

Yang, J.; Pitarch, E.P.; Kim, J. & Abdel-Malek, K. (2006). Posture Prediction and Force/Torque Analysis for Human Hands. *Digital Human Modeling for Design and Engineering Conference,* Lyon, 4-6 July 2006. Warrendale: SAE International.

Teamarbeit am Digitalen Lagetisch mit Fovea-Tablett®

ELISABETH PEINSIPP-BYMA, RALF ECK, THOMAS BADER & JÜRGEN GEISLER

Fraunhofer-Institut für Informations- und Datenverarbeitung IITB, Karlsruhe

Schlüsselwörter: Mensch-Maschine-Interaktion, kooperatives Arbeiten, Multi-Display-Technologie, Visualisierung

Zusammenfassung

Mit dem Digitalen Lagetisch und den Fovea-Tabletts wurde ein Arbeitsplatz geschaffen, welcher es einem Team von Experten ermöglicht, gemeinsam die Sicherheitslage in einer größeren geographischen Region (Szene) zu analysieren und Sicherheitsmaßnahmen zu planen. Dazu bietet der Digitale Lagetisch eine gemeinsame Szenenübersicht und durch den Einsatz der Fovea-Tabletts lagerichtige hochaufgelöste Detailsichten auf beliebige Szenenausschnitte. Über die Fovea-Tabletts erfolgt überdies die präzise Auswahl von ergänzender Information, die einzelnen Geländepunkten hinterlegt ist und auf einem hinter dem Tisch angebrachten Tafel-Display angezeigt wird. Der Digitale Lagetisch befindet sich im Stadium eines Labormusters. Experimente zur Benutzbarkeit sind in Vorbereitung.

1. Einleitung

In vielen Anwendungsbereichen besteht heute die Anforderung, dass ein Team von Experten die Sicherheitslage einer großräumigen Szene analysiert und daraus Sicherheitsmaßnahmen ableitet. Beispiele für solche Anwendungsbereiche sind die Einsatzplanung und -überwachung bei der Polizei, der Feuerwehr und beim Objektschutz sowie die Analyse kritischer Situationen im industriellen Umfeld oder im Kraftwerksbereich.

In allen geschilderten Fällen sind für die Aufgabenbearbeitung eine einheitliche Szenenübersicht für das gesamte Team und individuelle Detailsichten für ausgewählte Teammitglieder erforderlich („Übersicht versus Detail"). Hinzu kommt die Notwendigkeit, dass Zusatzinformationen individuell abrufbar sein müssen. Hier gilt es zu gewährleisten, dass den Experten ein unmittelbarer Zeige-Zugriff auf diese ortsbezogene Informationen ermöglicht wird, um den Mensch nicht durch unnötige Suchaufgaben oder eine zu große Informationsmenge zu überlasten.

Die heute kommerziell verfügbaren Darstellungstechnologien, zu nennen sind hier Projektoren und großflächige LCD- und Plasma-Monitore, ermöglichen die für ein Team ausreichend große Darstellung der zu analysierenden Szene. Da die Auflösung dieser Darstellungstechnologien jedoch ab einer gewissen Displaygröße für die Analyse von Detailbereichen der Szene zu schlecht ist, werden heute folgende Alternativlösungen praktiziert:

- Zoom: Der zu analysierende Detailausschnitt wird auf der ganzen zur Verfügung stehenden Displayfläche dargestellt. Nachteile an diesem Vorgehen sind, dass die Übersicht verloren geht und die Detailsicht in der Regel nur für einige Teammitglieder von Relevanz ist.

- Lupe: Der zu analysierende Detailausschnitt wird innerhalb der zur Verfügung stehenden Displayfläche vergrößert dargestellt. Nachteil an diesem Vorgehen ist, dass der Detailausschnitt Bereiche der Übersichtsdarstellung überdeckt.

- Fokus plus Kontext: Der zu analysierende Detailausschnitt wird in einem weiteren Display oder Fenster dargestellt. In diesem Fall können zwar unterschiedliche Detailsichten gleichzeitig dargestellt werden, der Bezug zwischen der Detailsicht und der Übersicht geht jedoch verloren.

Abbildung 1 zeigt als Beispiel für einen Teamarbeitsplatz mit unterschiedlichen Darstellungstechnologien das Gemeinsame Melde- und Lagezentrum von Bund und Ländern (GMLZ).

Abbildung 1: Gemeinsames Melde- und Lagezentrum von Bund und Ländern.

Zwar wurden in den letzten Jahren unterschiedliche Großbildschirmdarstellungen mit hoher Auflösung entwickelt, jede dieser Lösungen weist aber im Hinblick auf die oben aufgeführten Anwendungen und die Anforderungen, die bei deren Bearbeitung durch ein Expertenteam entstehen, Nachteile auf. Rückprojektionen mit hoher Auflösung erfordern neben einem hohen Anschaffungspreis auch einen hohen Kalibrierungsaufwand (Knöpfle & Stricker 2004), zusammengesetzte LCD-Monitore verhindern durch die Ränder zwischen den Monitoren ein durchgehendes Bild. Hinzu kommt, dass der Mensch ungünstigstenfalls durch die detailreiche Darstellung, welche sich an der besten Auflösung der zusammengesetzten Darstellungstechniken orientiert, die Übersicht verliert, da er durch die hohe Informationsdichte überfordert wird.

Zur Lösung des Problems „Übersicht versus Detail" werden folgende Lösungen vorgeschlagen:

- Mittels Auflichtprojektion wird die Szenenübersicht (Kontext) dargestellt und die Detailsicht eines kleineren Gebiets (Fokus) durch einen zweiten Projektor der Szenenübersicht überlagert (Ashdown & Robinson 2003: beide Projektoren sind dabei fest montiert und im Überlagerungsfeld wird das Bild des Übersichtsprojektors maskiert).

- Ähnlich ist der feste Einbau eines kleinen LCD-Displays in die Darstellungsfläche einer Rücklichtprojektion, welches die Detailsicht präsentiert (Baudisch et al. 2002).

Beide Methoden haben den Nachteil, dass der Nutzer Übersicht und Detailsicht nicht frei wählen kann. Immer muss zur Auswahl des gewünschten Details die Szenenübersicht verschoben werden.

Eine weitergehende Lösung stellen (Sanneblad & Holmquist 2006) vor. Mittels Auflichtprojektion wird die Szenenübersicht (Kontext) an die Wand projiziert und davor ein Hand-held Display gehalten, welches die Detailsicht anzeigt. Die Position für die Detailsicht ist hier frei wählbar. Das Hand-held Display (und damit die Detailsicht) kann jedoch nicht gedreht werden und die Anwendung ist auf nur ein Hand-held Display beschränkt.

2. Der Digitale Lagetisch mit Fovea-Tablett

Um den oben genannten Anforderungen gerecht zu werden, wurde der Digitale Lagetisch mit Fovea-Tablett entwickelt. Zum einen verbindet dieser Displays unterschiedlicher Auflösung miteinander und löst durch deren geschickte Kombination den Konflikt „Übersicht versus Detail" auf (Bader 2004). Zum anderen wurde für den Digitalen Lagetisch ein Interaktionskonzept entwickelt, welches das gemeinsame Bearbeiten einer Szene durch ein Team von Experten ermöglicht. Abbildung 2 zeigt den Einsatz des Digitalen Lagetisches am Beispiel des Szenarios „Planung einer Überwachungsaufgabe". Im Weiteren werden die Komponenten und die Funktionalität des Digitalen Lagetisches an Hand dieses Szenarios verdeutlicht.

Abbildung 2: Der Digitale Lagetisch mit Fovea-Tabletts.

Der Digitale Lagetisch besteht aus folgenden Komponenten:

- Horizontale Darstellungsfläche (Tischdisplay), welche als Durchlichtprojektion realisiert ist und zur Übersichtsdarstellung der Szene dient.

- Vertikale Darstellungsfläche (Tafeldisplay), welche für die Darstellung von Zusatzinformationen und Seitenansichten der Szene eingesetzt wird.

- Sogenannte Fovea-Tabletts, welche sowohl zur lokal hochaufgelösten Darstellung des darunter liegenden Szenenausschnitts als auch zur Interaktion mit dem Tisch- und dem Tafeldisplay dienen (siehe Abbildung 3).

Abbildung 3: Das Fovea-Tablett zeigt jeden beliebigen Szenenausschnitt des Tischdisplays in hoher Auflösung an und liefert damit die Funktionalität einer verdeckungsfreien Lupe.

Zur Szenendarstellung auf dem Tischdisplay und auf den Fovea-Tabletts werden Raster- und Vektordaten verwendet, welche die Szene in der Draufsicht darstellen. Um dem Menschen nur so viel Information darzustellen, wie er zur Analyse der Szene benötigt, und die Darstellung von unnützer Information zu vermeiden, werden auf dem Tischdisplay und den Fovea-Tabletts die Informationsdichten den jeweiligen Anforderungen angepasst. Während auf dem Tischdisplay entsprechend den Anforderungen an eine Übersicht nur die groben Strukturen der Szene visualisiert werden, wird auf den Fovea-Tabletts die höhere lokale Auflösung genutzt und Detailinformation eingeblendet. In der gegenwärtigen Ausführung besitzt das Tischdisplay eine Größe von 118 cm x 88,5 cm. Bei einem XGA-Bild (1024 x 768) wird damit eine Pixeldichte von 22 ppi (pixel per inch) erreicht. Als Fovea-Tabletts werden Tablet-PCs mit einer Bildschirmgröße von 24,5 cm x 18,4 cm verwendet, wobei bei einem XGA-Bild 110 ppi erreicht werden.

Damit die Fovea-Tabletts immer den darunter liegenden Ausschnitt des Tischdisplays darstellen, ist auf der Unterseite jedes Fovea-Tabletts eine sogenannte MC-MXT-Marke (Multi-Cursor-MarkerXtrackT) befestigt (Abbildung 4, links). Eine Kamera, die im nahen IR-Bereich (0,78 – 1,5 µm) das Tischdisplay von unten aufnimmt, liefert ein Bild (Abbildung 4, rechts), das mit einem automatischen Bildverarbeitungsverfahren ausgewertet wird. Das Verfahren berechnet die Position einer Marke mit einer Subpixelgenauigkeit von besser als 0,1 Pixel und die Orientierung der Marke mit einer Genauigkeit besser 2° (Rehfeld 2001).
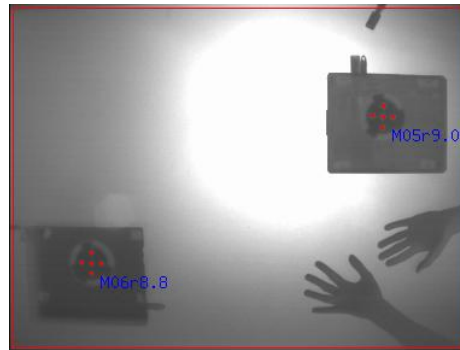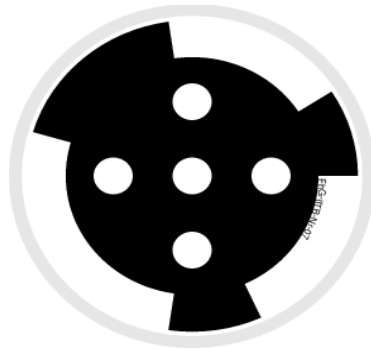
Abbildung 4: MC-MXT-Marke (links) und Visualisierung der Verfahrensergebnisse (rechts).

Die sogenannte Pose (Position und Orientierung) wird drahtlos an das zugehörige Tablett übertragen, welches den auf ihm darzustellenden Szenenausschnitt berechnet. Dieser wird so gewählt, dass der vom Fovea-Tablett verdeckte Szenenausschnitt dem Benutzer in hoher Auflösung dargestellt wird. (siehe Abbildung 3). Beim Verschieben des Fovea-Tabletts wird der korrekte Szenenausschnitt nahezu verzugsfrei aktualisiert.

Die Interaktion mit dem Digitalen Lagetisch und den Fovea-Tabletts erfolgt ausschließlich über die Fovea-Tabletts mittels Stift und Symbolleiste. Diese Symbolleiste befindet sich auf der rechten oberen Seite jedes Fovea-Tabletts (Abbildung 5).

Abbildung 5: Interaktion mit dem Digitalen Lagetisch über Stifteingabe am Fovea-Tablett.

Bisher realisierte Interaktionen sind u.a. das Verschieben und Zoomen der Szenendarstellung auf dem Tablett und auf dem Tischdisplay. Da gleichzeitig mehrere Fovea-Tabletts eingesetzt werden können, ist die Manipulation der Szenensicht am Tischdisplay nur vom Tablett des Teamleiters aus möglich. Auch kann nur vom Tablett des Teamleiters eine andere Szenensicht geladen werden.

Eine Funktionalität, die wiederum auf jedem Fovea-Tablett zur Verfügung steht, ist das Abrufen von Zusatzinformationen zu den in der Szene befindlichen Objekten. Als Zusatzinformationen stehen beschreibende Texte, Diagramme, Bilder und Online-Zugriffe auf „*Webcams*" und eigene Sensoren unterschiedlicher Typs zur Verfügung. Die Zusatzinformationen werden auf dem Fovea-Tablett durch Tippen mit dem Stift auf eine in der Karte annotierte Stelle ausgewählt und am Tafeldisplay dargestellt (Abbildung 6). Um Interessenkonflikte zwischen den Team-Mitgliedern zu

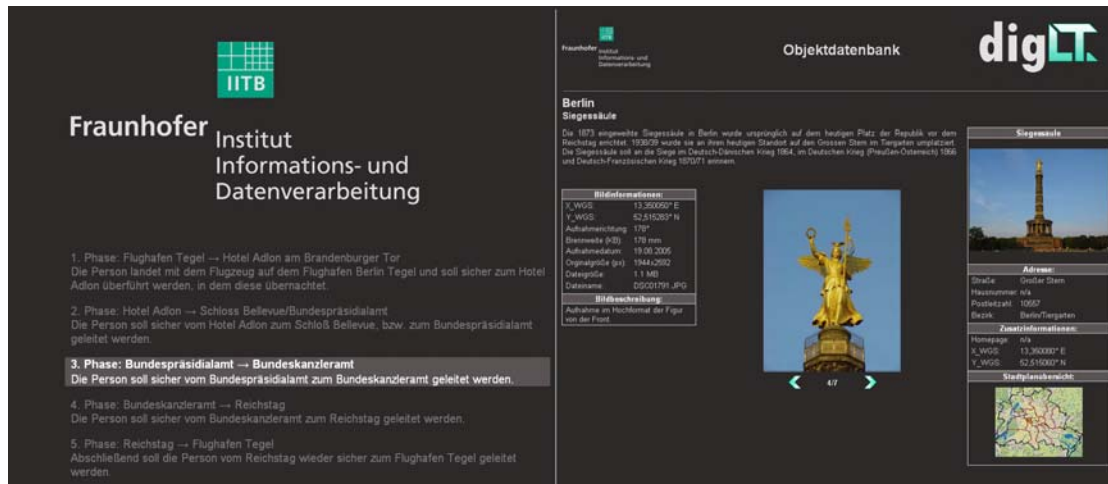vermeiden, können zu einem Zeitpunkt immer nur von einem Tablett aus Zusatzinformationen abgerufen werden.

Abbildung 6: Tafeldisplay zur Darstellung der Zusatzinformationen.

3. Ausblick

Erste Untersuchungen am Digitalen Lagetisch zeigten, dass eine Manipulation des am Tischdisplay dargestellten Szenenausschnitts, z. B. globales Verschieben oder Drehen nicht ausschließlich über die Fovea-Tabletts erfolgen sollte ist. Eine Möglichkeit der direkten Manipulation dieser Szensicht sind manipulatorische Handgesten, die direkt auf das Übersichtsbild wirken. Erste Arbeiten dazu sind abgeschlossen (Bader 2006). Die Integration mit der Interaktion über die Fovea-Tabletts ist gegenwärtig in Arbeit. Als weiteres Anwendungsgebiet wird gegenwärtig der interaktive Entwurf elektronischer Schaltkreise untersucht. Der Digitale Lagetisch würde dann zu einem digitalen Reißbrett.

Zahlen / Fakten

Größe des Lagetischs: 118 cm x 88,5 cm

Auflösung von Tischdisplay und Fovea-Tablett: jeweils 1024 x 768 Pixel

Fovea-Tablett® DE-Patentanmeldung 10 2004 046 151.1

Fovea-Tablett® Markenanmeldung 304 64 105.7 / 09

Literatur

Ashdown, M & Robinson, P (2003). The Escritoire: A Personal Projected Display. *Journal of WSCG, 11,* (1), 33-40

Bader, T. (2004). Echtzeitkopplung von Kartendarstellungen auf verteilten mobilen Displays, *Studienarbeit an der Fakultät für Informatik, Universität Karlsruhe.*

Bader, T. (2006). Handgestenerkennung für ubiquitäre Multi-User-Interaktion an großflächigen Displays, *Diplomarbeit an der Fakultät für Informatik, Universität Karlsruhe.*

Baudisch, P.; Good, N.; Bellotti, V. & Schraedly, P. (2002). Keeping Things in Context: A Comparative Evaluation of Focus plus Context Screens, Overviews, and Zooming. In: *Proceedings of SIGCHI 2002* (S. 259-266).

Knöpfle, C. & Stricker, D. (2004). HEyeWall – Perfect Pictures for New Business Solutions. *Computer Graphik Topics 16,* 5-6.

Rehfeld, T. (2001). Codierte Marker als Mess- und Interaktionskomponente für das Kindermuseum ZOOM in Wien, In: *Fraunhofer IITB Jahresbericht 2001* (S. 32-33).

Sanneblad, J. & Holmquist, L.E. (2006). Ubiquitous Graphics: Combining Hand-held and Wall-size Displays to Interact with Large Images. In: *AVI 06, 2006.*

Towards Networked and Structured VR European Research Area: Intuition Network of Excellence and Future Research Challenges

ANGELOS AMDITIS[1], MATTHAIOS BIMPAS[1] & ROLAND BLACH[2]

[1]*Institute of Communications and Computer Systems (ICCS), Iroon Polytechneiou Str.9, Zografou, Athens, Greece*

[2]*Fraunhofer Gesellschaft für Arbeitswirtschaft und Organisation (FhG-IAO)*

Abstract

The massive research and development process concerning Virtual Reality (VR) technology has reached a degree which makes a pan-European structuring and integrating effort an absolute necessity. Despite the fact that VR/AR technology has started being used to an extent in different industrial applications, this has been processed though in an unorganised way, lacking of long-term vision and dealing with case-by-case scenarios. Thus, a critical milestone is to facilitate the adoption of Virtual Environments (VEs) in industrial processes and assess the impact of its "penetration" into the workplace and everyday life. INTUITION works towards this direction, with the prime objective of promoting and facilitating the development and application of VR/VE technology in even more industrial domains, establishing thereby European excellence in this area. Moreover, INTUITION partners attempted to investigate the existing barriers for further penetration of VR into industrial applications and recognise the most important drivers that will lead to enhanced and wider use of VR/AR. An additional scope of this paper is to propose the main research areas and challenges in the field of VR/AR as idientified through INTUITION Working Groups throughout the last couple of years.

1. Introduction

The expression "Virtual World" describes worlds existing with the support of computers or computer networks only. Virtual worlds can be completely fictive as well as representations of real worlds. Virtual worlds are not necessarily three dimensional and realistic. Some authors consider even text based environments (e.g., chatrooms, etc.) as virtual worlds. In the context of INTUITION, a Virtual World is composed of three dimensional objects, light sources, grids and three dimensional user interfaces. The all encompassing term "Virtual Reality" (VR) is used to describe the technology in general. It addresses the required hardware, basic software and applications, which are necessary to give a user access to a virtual world. A "Virtual Environment" (VE) is defined to consist of a VR-installation and at least one virtual world. A VE provides the users the immersed examination of and interaction with virtual worlds.

Within recent years research on the relative scientific field has expanded on the one hand. On the other hand, there were a number of reasons that prevented the broad establishment of VR technology and relevant tools in the product creation process. The high costs concerning infrastructure and needed equipment was probably the dominant reason. In addition the lack of major VR applications enhanced the industrial tendency to stick to more traditional tools. Having some applications as a background the major advantages that VR technology can introduce to extended services and product development approaches were easily highlighted. These advantages, along with the fact that VR technology became more "mature" and able to cope with increasing demands on interaction and virtual representation, led to a number of additional industry-supported research activities.

Although that was a step forward, these activities were more a result of individual efforts, leaving Europe to follow evolutions coming mostly from the other side of the Atlantic. Several research teams have been working all across Europe focusing on VEs from different aspects. Towards this direction, INTUITION, which is a Network of Excellence funded under 6th EU Framework, aims to combine and structure their efforts providing Europe with a pan-European network that can lead future evolutions on the relative sector, assist to establish VR as a major tool in product and process design and establish a training basis allowing new researchers to join this field or helping them to broader their knowledge on an international cross-application platform.

INTUITION includes 58 European partners, stemming from various fields such as industrial representatives, Small and Medium-sized Enterprises (SME's), key research institutes, universities and major international organizations or associations. INTUITION's major objective is to bring together leading experts and key actors across all major areas of VE understanding, development, testing and application in Europe, in order to overcome fragmentation and promote VE establishment within product and process design. Its major objectives include the integration of resources and VR equipment all around Europe, the structuring of European Research Area in VR and the promotion of Europe as a leading force in this field world wide.

To perform this, a number of activities have been already carried out in order to establish a common view of VE technology current status, open issues and future trends. These activities include integration of human and infrastructure resources,

research structuring, spreading of excellence and dissemination tasks. The quite large consortium is controlled by a firm managerial structure. Strong links with National relevant Networks, current National and EU-funded projects and clustering activities with new initiatives as well, assist in structuring the VR European Research Area (Blach et al. 2006).

This paper aims to provide the main barriers for further penetration of VR in industry, some of the drivers for change and the main research challenges in the field of VR, as identified by INTUITION working groups' activities. The main part of this paper content derives from INTUITION Research Roadmap that was the outcome from the specialised Working Groups roadmaps. Apart from that, the state of the art in the field has been identified in several deliverables, namely the Terms of References of the WGs, the common State of the Art report and the User Requirements Document. All these documents have been created by a broad range of European VR/AR experts from various disciplines and different organizational background reflected in the members of the INTUITION network.

The WG roadmaps have been derived via:

- State of the Art summary

- Focal point identification

- Research position papers

- Workshops on roadmaps

2. Identified Barriers for Further VR Penetration and Drivers for Change

2.1 Barriers to Change

The barriers and challenges which have to be tackled to overcome the limitations of today's systems and make them freely available can be structured in following major categories:

2.1.1 Technology

The main technological barrier is the inadequacy of both hardware and software development for simulation technology to provide simultaneously necessary accuracy and real-time data. Apart from this, the bandwidth for stimulation of the human sensory system is still relatively low (e.g. display resolution, colour depth, frame rate, field of view, etc.), while spatial registration and tracking of user and environment are not accurate and not available in a really ubiquitous manner. Additionally, the lack of efficient application/content development systems and the insufficient syn-

chronisation of multimodal and multi-sensory components is one extra barrier, that technology research has to tackle with. Finally, the premature hard- and software for VR-systems in terms of usability, stability and complexity and the necessary provision of ubiquitous access to virtual environments especially in AR, along with the improvement of power consumption in portable devices may be considered as secondary technological barriers.

2.1.2 Interaction concepts

The lack of immersive 3D-UI paradigm (comparable to the 2D WIMP paradigm) and the fact that spatial interaction is not properly understood at cognitive level are the main barriers in terms of interaction. Furthermore, the non existence of evaluation methodology for spatial interaction and presence constitutes a negative factor for further VR penetration. Finally, the missing knowledge of how to represent abstract data in space and whether VR/AR interaction techniques can assist in improving understanding and access to abstract data is the last identified barrier of this category (Kennedy et al. 1993, 1997).

2.1.3 Integration

Lack of integration and interoperability in VR systems and related applications such as CAD, CAE, PLM, is one of the most important barriers. This can be further decomposed in the lack of widely accepted standards (Data, Behaviour and Interaction), the lack of integration of VR/AR in existing applications and into the existing workflows.

2.1.4 Socio-Economic issues

Last but not least, some socio-economic issues such as the high cost of VR-systems are avertive factors towards VR penetration, especially concerning small SMEs and industries. Of course an important reason is also the fact that utilisation of new interfaces breaks with learned habits and there is still long way to go for making enough mature the ground for further VR penetration to industry.

These categories are interdependent and advances have to be achieved in more than one area to achieve a major improvement.

2.2 Drivers for change

In the previous section, the barriers for further penetration of VR were presented in terms of technology, interaction concepts, integration and socio-economic factors. On the other hand, some very important drivers for change of the current scenery in the field of VR have also been identified by INTUITION Working Groups and presented in the following paragraphs. Drivers outside the technology advancements themselves can be found in the further need of processing more data and more efficient access technologies to these data. Besides the hardware oriented point of view, we have identified the following other driving forces:

2.2.1 Industry requests

Several industries gain already a benefit from VR/AR-technology. Virtual prototypes pay off (from design and early prototyping to training applications and maintenance)

in companies where 3D digital data is ready at hand. The request for improvement with the will to invest is an important driving factor (Chryssolouris et al. 2004).

2.2.2 Socio-Economic

The socio-economic drivers may be summarised to the following hints:

- Complexity of information grows, therefore new access techniques are required.

- Next generation of computer users with VR/AR interface awareness is growing.

- There is increased interest of investors in novel technology opportunities.

2.2.3 Technology

From the technological point of view, the following drivers have been identified:

- Graphic cards develop fast and independent of VR/AR. The game industry is technology driver for real-time 3D graphics hardware.

- Projector and display technology develop fast and independent of VR/AR.

- Integration of VR/AR Technology in other human computer interfaces.

- More collected data is 3D (products, terrain, etc.) e.g. CAD goes 3D, Google Earth.

- Mobile phones and PDAs are wide-spread and can be the computing platform for VR/AR in the future. Even complex algorithms can be performed in future terminals.

3. Proposed Research Fields by INTUITION

In the previous chapter, a series of barriers that must be overcome has been presented. These barriers are principally related to the lack of visibility on real benefits of VR as an enabler to fast, cost-effective and valid product management, and to still difficult implementation of technologies and systems, which still present narrow application envelopes.

Moreover, the drivers which encourage investing into research activities have been presented. These drivers are related to the opportunity to extend the virtual approach to the whole product lifecycle management, increase the coverage of performances both by multi-domain simulation and interactive experience of users and engineers and contribute to the process agility and fast reaction. All these can be achieved through extended co-locate and remote collaboration among engineers, decision makers and users.

3.1 Interface Technologies

3.1.1 Components of multimodal interfaces

Visual display technology for immersive environments

Novel display technology for 3D-representation has to be developed. Unencumbered displays that free the user from unnecessary equipment, such as lightweight high resolution head mounted displays with wide field of view, flat panel active or passive stereo displays, auto-stereoscopic displays, volumetric display, true holographic displays, etc should be further explored, which can also be implemented for mobile VR systems.

Aural spatial display technology

Novel aural systems for low- and high-end spatial sound synthesis. Wave field synthesis is a novel technology which allows for true spatial aural rendering. Directed sound systems or noise cancellation systems could be exploited for interactive systems. New lightweight systems for unencumbered use should be developed.

Multi-user displays for the visual and aural channel

To support true co-located collaboration all users should have their own view or sensory experience (Experts Group on Collaboration@Work 2006). The generation of multi perspective views for the aural and visual perceptual system without personalized systems like ear-phones or head mounted displays, is still not solved although some approaches have been presented already.

Haptic display technology

Haptic displays consist of haptic devices and the haptic rendering software. Mechanics, electronics, automation and control, the integration of smart materials and technology, etc. are involved. Haptic rendering software (the computer-based contribution in the generation of a haptic feedback to the user) involves algorithmic components to design the haptic "core" library, including haptic specific collision detection, texture mapping, etc.

Tracking/Estimation of position and orientation

Tracking technology should become more accurate and more user friendly. New approaches to overcome these limitations comprise marker less and/or model based tracking, source less tracking as, e.g. gyroscopic systems, hybrid tracking and large area tracking. Moreover, tracking must evolve towards multi-user systems, to help enabling VR-based collaboration.

Gesture and Posture recognition

Gesture and posture recognition are based on tracking and other measurement technology where the dynamic spatial behaviour is interpreted as user commands sometimes clear sometimes supportive. In this field sophisticated dynamic pattern recognition algorithms that are able to be used without tedious individual calibration procedures should be developed.

Integration of Speech recognition

Speech recognition is an independent research field, which fits well in immersive environments. Existing systems have to be integrated and interaction concepts must

be developed, to allow even more natural communication with the virtual world or the underlying simulation system.

Integration of mobile devices

Existing mobile devices ought to be made 3D aware, while proper interaction techniques should be also developed.

3.2 Content Technologies

3.2.1 Realistic behaviour of VEs

Visual realism

Visual realism has to be improved in the field of virtual prototyping. New techniques as hardware based shader or real-time ray-tracing are still not optimal and it is essential that they are improved for general use. This is also related to material and lighting modelling which should capture more of the physical correct behaviour of surfaces. In non-real-time computer graphics, valuable models have been developed. These have to be adapted to real-time environments.

Physical behaviour and Haptic realism (Rigid body, deformable objects, EM, CFD)

To describe realistic environments, real–time algorithms for dynamic systems are indispensable. Although near real-time systems for restricted problems are available, there are no generalized solutions, especially not for large data sets. Also a good general description model for the physical behaviour and the capturing of the necessary parameters is not solved yet. For convincing haptic rendering these models are also crucial.

Data acquisition of physical properties

The transfer of real world data and physical properties for mapping real world in virtual environments is still a manual process with only some tools to support the process. Examples for such processes are texture generation or more sophisticated image based acquisition of material parameter, 3D-scanning based on different technologies, but also the parameterization of an acoustic or haptic virtual environment from real world objects. Therefore, there is a great need for the development or extension of algorithms and tools to render the process as automatic as possible.

3.2.2 Representation of real and virtual humans

Virtual Humans

The simulation of virtual humans is used for training, education and entertainment as representation of remote user or intelligent agents. Human models exist already but better physical behaviour has to be developed. The movements should be more natural, the programming should be simple, while the simulation should be efficient and scaleable, so as for multiple representatives to be used in the same virtual environment (Badler et al. 1998).

Virtual humans are also used in the industry for ergonomics and model based product interaction studies. Better behavioural representations, environment awareness and

adaptation, and possibly cognitive and emotional capabilities (in the long term) are of great application potential.

3.2.3 Content Management

Content development tools

To combine the interface technology with the content and to utilize multimodal interaction concepts specific application development systems are necessary. In the last decade very specialized tools have been developed. In the future they have to integrate into the standard workflow of content or software development. New architectures or novel open integration schemes have to be explored. Also, these content development tools have to enable designers (not programmers) and domain experts to create contents for VR applications (Musse & Thalmann 2001).

3.3 System and Integration Technologies

3.3.1 Software Architectures

Architectures of VR/AR systems

Under this sector, protocols and drivers for multi-domain real time product simulation have to be developed while at the same time research must particularly focus on Plug and play architecture for multi-domain simulations and define standardized architectures for VR/VE systems with clear interfaces among components. This would certainly lead to improvement of technologies, specifically mobile and remote.

Architectures for distributed and collaborative systems

To combine multimodal systems with collaborative interactions and remote and local multi-user support, new architectures have to be developed. These systems have to provide mechanisms which guarantee persistence, synchronicity and low latency. To achieve this, shortcomings of network technology also have to be identified and resolved.

3.3.2 Data integration in external systems or processes

Automatic exchange/transformation of data with VEs

It is crucial for the integration of VR/AR-technology in existing workflows that data from various sources can be automatically prepared and adapted to the requirements imposed by the VR-technology. Also the conversion back to the original data sources and their relation has to be carried out. These systems are mainly CAD systems, CAE systems, game/edutainment authoring systems or databases as PDM/PLM systems. Automated techniques have to be researched and tools have to be developed.

3.3.3 Interoperability

Standardization on data, scene description and functionalities

With a well defined ontology a formal data standard can be established where scene content, system configuration and interactions can be described in an unambiguous way and therefore can be publicly available.

VR/AR interface integration in existing applications

A common standardized interface has to be developed, such that new applications can utilize VR/AR-interface technologies in their application development comparable to the process of using WIMP based interfaces.

3.4 Interaction

3.4.1 Generic Interactions

Multimodal Interaction concepts for VEs

Interface components can be bundled to interaction concepts where input and output with different modalities can be combined to an interaction concept, e.g. navigating through a virtual scenery can be realized with haptic and visual rendering as output and haptic input and tracking as input modalities. Only arbitrary concepts exist up to date (Ruddle & Payne 1997).

3.4.2 Application/activity-specific human interaction/interface paradigm

Application specific interactions, which are strongly linked to specific applications, are also a research and development topic. These interaction techniques are described in detail in the respective application oriented research roadmaps. Here only the generic interactions which can be exploited in many application fields are mentioned.

Natural interaction for creation and composing of geometry/objects/worlds in immersive environments (design)

This finds great use in the automotive industry, where through an immersive environment the engineer and designer can virtually shape the automotive structure based on design preferences and safety features. The virtual model replies instantly and the changes take effect immediately providing information to the designers about possible reactions to given situations.

Process Simulator

The research under this area ought to be focusing on VR simulation of simple and complex manufacturing processes.

Education (teacher/learner interaction, abstract concepts)

VR/VE environments are also used for education, training and learning purposes (Bloom et al. 1984). A very good example of such a case is when practical knowledge needs to be passed from a senior to a junior employee for instance. A VR system stores the data of a task executed by a senior employee of an industry, and the data could afterwards be provided to the junior trainee to assess and learn from the process. This is particularly useful when the task involves the use of machinery and dangerous equipment (Regian et al. 1992).

3.5 Human Factors

3.5.1 Basic Research

Human Factors are an important field in human machine interface research because the results not only justify the benefits of a new technology, but also shape the tech-

nical design towards a more humane realization and application of this technology. Many questions are raised, which cannot be answered without exploring the technology itself. Many conclusions can be drawn only from existing prototypes, which are built without the gained experience of its application. So, it is an iterative process where the outcome of this research shapes the design of the system and vice versa (Stoaklei et al. 1995).

Presence Research

The question of presence in computer mediated environments, how to measure it and the benefit for the user, has not been adequately explained yet. It may be important to design some VE applications to encourage feelings of presence, e.g. in a training application to train astronauts on maintenance tasks in space in zero gravity conditions. In this example it is important to afford the feeling of zero gravity when conducting these maintenance tasks. Another example is VEs designed to treat people with phobias – if the user feels like he/she is really in the situation which scares her, then there is likely to be a good transfer of any progress into the real world. In the case of design applications, a feeling of presence may refer to an acceptance that the VE represents the real working environment and 'being able to think oneself into an environment in order to design and evaluate it' (Barfield et al. 1995).

Perception/Cognition Research

From cognition research the mapping of basic understanding of immersive interface technology has to be derived. What is space and how it will be perceived? How accurate and how well synchronized should the sensory stimulation be? (Wang 1996, Wickens & Baker 1995)

Understanding of collaboration in immersive environments

The way people collaborate in immersive environments effectively and efficiently should be explored and evaluated, when true collaborative hard- and software systems have been setup.

Evaluation Technologies

Evaluation methods are needed to compare various technologies, systems and concepts, regarding their usefulness, ergonomics and health conformity.

3.5.2 Ergonomics and Design

Design Guidelines and Methodologies

From the above described research fields design guidelines and methodologies can be concluded. These help to standardize the development of VE's and ensure a controlled quality of these environments regarding human factors.

4. Conclusions

A general issue to be taken into account is the support concerning underlying themes for breaking the current barriers in deploying VR technologies. These barriers must be overcome, principally related to the lack of visibility on real benefits of VR as an enabler to fast, cost-effective and valid product management, and to still difficult

implementation of technologies and systems, which present narrow application envelopes. Drivers encourage investing into research activities, which are related to the opportunity to extend the virtual approach to the whole product lifecycle management, increase the coverage of performances both by multi-domain simulation and interactive experience of users and engineers during the development, and contribute to the process agility and fast reaction, through extended co-locate and remote collaboration among engineers, decision makers and users.

Concerning the implementation and use of VR technologies in the European industrial sector, the strategic challenges are mainly related to:

- Faster design, validation and production start-up of better products, targeted at the highest personalization, and customer and society perceived added value.

- Shorter reaction time to market dynamics, including anticipating and fostering unexpressed or emerging needs.

These challenges force, in the short-medium term, to improve and extend the coverage of the virtual approach in the current development processes, and in the long term to envisage and implement a new product life cycle management process, truly built around the virtual approach. This is expected to fully exploit a vision consisting of a new product lifecycle management process, in which users (both of product and processes) are thoroughly and consistently involved in all stages: concept definition, design, development, evaluation and refinement, manufacturing design, logistics, training, marketing and presentation, operation and maintenance.

Acknowledgements

References

Badler, N.; Bindganavale, R.; Bourne, J.; Palmer, M.; Shi, J. & Schuler, W. (1998). A Parameterized Action Representation for Virtual Human Agents. In: *Proceedings of the First Workshop on Embodied Conversational Characters* (pp. 1-8).

Barfield, W.; Zeltzer, D.; Sheridan, T. & Slater, M. (1995). Presence and performance with virtual environments. In: W. Barfield & T.A. Furness (eds), *Virtual environments and advanced interface design* (pp. 473-513). New York: Oxford Press.

Blach, R.; Dangelmaier, M.; Bimpas, M.; Frangakis, N.; Walzack, K.; Gravez, P. & Bourdot, P. (2006). *ERA Structuring Plan and Future Research Activities*. Technical Report- Deliverable D1.E_1, Location: www.intuition.org.

Bloom, B.S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13 (6), 4-16.

Chryssolouris, G.; Mavrikios, D.; Fragos, D.; Karabatsou, V. & Alexopoulos K. (2004). A hybrid approach to the verification and analysis of assembly and maintenance processes using Virtual Reality and Digital Mannequin technologies. In: A.Y.C. Nee & S.K. Ong (eds.), *Virtual Reality and Augmented Reality Applications in Manufacturing*. London: Springer.

Experts Group on Collaboration@Work (2006). *New Collaborative Working Environments 2020*. Report on industry-led FP7 consultations and 3rd Report of the Experts Group on Collaboration@Work.

Kennedy, R.S.; Lane, N.E.; Berbaum, K.S. & Lilienthal, M.G. (1993). Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology, 3*, (3), 203-220.

Kennedy, R.S.; Lanham, D.S.; Drexler, J.M.; Mssey, C.J. & Lilienthal, M.G. (1997). A comparison of cybersickness incidences, symptom profiles, measurement techniques, and suggestions for further research. *Presence-Teleoperators and Virtual Environments, 6*, 638-644.

Musse, S.R. & Thalmann, D. (2001). Hierarchical Model for Real Time Simulation of Virtual Human Crowds, *IEEE Transactions on Visualization and Computer Graphics, 7,* (2), 152-164.

Regian, J. W.; Shebilske, W. & Monk, J. (1992). A preliminary empirical evaluation of virtual reality as an instructional medium for visual-spatial tasks. *Journal of Communication, 42,* (4), 136-149.

Ruddle, R.A. & Payne, S.J. (1997). Navigating buildings in "desk-top" virtual environments: Experimental investigations using extended navigational experience. *Journal of Experimental Psychology, 3*, 143-159.

Stoaklei R.; Conway, M.J. & Pausch, R. (1995). Virtual reality on a WIM: interactive worlds in miniature. In: *Proceedings of ACM CHI'95 Conference on Human Factors in Computing Systems* (pp.265-272).

Wang, X. (1996). Learning Planning Operators by Observation and Practice. *Ph.D. Thesis, School of Computer Science, Carnegie Mellon University.*

Wickens, C.D. & Baker, P. (1995). Cognitive issues in virtual reality. In: W. Barfield & T.A. Furness (eds), *Virtual environments and advanced interface design* (pp. 514-541). New York: Oxford Press.

Trends and Developments from the VR2007 Conference

JOHANN HABAKUK ISRAEL

Fraunhofer Institute for Production Systems and Design Technology (FhG IPK), Berlin, Germany

Virtual Reality (VR) is not a niche but key of a larger trend in next generation Reality Based Interfaces (RBI) according to Robert J. K. Jacob from Tufts University in his keynote address at this year's 3D User Interfaces (3DUI) Symposium, which took place in conjunction with the IEEE Conference on Virtual Reality. "Use reality wherever you can, for the rest, use something different." Jacob (figure 1) encouraged researchers to continue and broaden their work towards the development of interfaces which in his opinion should be based on natural users' skills and incorporate new technologies whenever what we gain by the computers' capabilities is more than we lose by giving up familiarity in handling objects and information.

The annual IEEE Conference on Virtual Reality brings together researchers and companies involved in the field. The 2007 conference took place in Charlotte, North Carolina. It covered the whole range of VR related themes ranging from perceptual and human factors issues, technical questions and interaction research, to modeling and rendering of virtual humans, to name just a few.

Beside much in-depth research work presented during the conference, it was also the place to share opinions on future trends and developments in VR. Jim Foley from Georgia Tech opened the discussion in his keynote by claiming three big T's which will dominate Virtual Environments in future: Training, Therapy, and Theatre. As for many other technologies, Foley sees games and entertainment as the main driving factors in making VR technology affordable for consumers. This in turn could open up new markets and bring VR applications into daily life. He also drew attention to the point that complete immersion is not necessary for all tasks. While considerable immersion is achievable at relatively low costs, a full immersion system might easily inflate the budget, so it is crucial for the economic success of Virtual Environments (VEs) to keep the relationship between costs and effectiveness in mind when planning them. When asked for what comes after Virtual Reality, he speculated that

Figure 1: Rob Jacob (right), keynote speaker, and Wolfgang Stürzlinger (left), organizer of the 3DUI Symposium.

in a future evolutionary step we might lose our ability to walk and live our entire life in virtual environments, similar to the popular Second Life.

In a statement addressed to those planning an academic career and shaping the future of VR, Frederick P. Brooks from University of North Carolina, Chapel Hill, one of the fathers of VR, gave the advice to "do what makes you crazy and do it well, find your passions, and find a way you make your passion useful." Many agreed that the enthusiasm and innovations which drive the computer games industry could also be a model for developing business and industrial VR systems. Applications could be structured as serious games in order to make them more lively, usable and less tiring.

The program of the conference included several panels and sketch sessions which covered:

- Perception & Human Factors

- 3DUI & VR/AR Systems

- Scene Complexity Management

- Modeling & Simulation

- Distributed & Networked VR

- Display

- Multi-sensory Interaction

- Modeling & Rendering

- Modeling, Rendering & Virtual Humans (Sketches)

Figure 2: Infrared based optical tracking system iotracker from Vienna University of Technology.

- VR Systems & Applications (Sketches)

- Augmented & Mixed Reality (Sketches)

The talks covered the whole spectrum from basic physiological/perceptual research to core VR technology. Issues related solely to interaction were covered in the Symposium on 3D User Interfaces (3DUI), which took place two days before the conference. There were a broad range of contributions. Niklas Elmqvist from Chalmers University of Technology, Göteborg, presenting a taxonomy and design patterns for 3D occlusion management techniques, a vital step in order to help users to discover and access occluded virtual objects and find relationships between them (Elmqvist & Tsigas 2007).

Virtual mirrors were suggested as new interaction techniques for Augmented Reality (AR) systems by Nassir Navab from Technische Universität München. They are of great use when moving the head is too complicated or even impossible. Navab demonstrated the mirror in the context of laparoscopic surgery, showing impressive pictures of precisely superimposed views of real patient videos and CT images (Navab, Feuerstein, & Bichlmeier 2007). Examples like these re-emphasized what medicine can gain from using VR. "Medicine is the immediate future of VR" as participants noted during the discussion.

In a talk on Distributed Virtual Environments (DVEs), Silvia Rueda from the Universidad de Valencia presented studies underlining the benefits of peer-to-peer (P2P) networks in terms of scalability, response times and system saturation (Rueda, Morillo, Orduña & Duato 2007).

Non-isomorphic manipulation, that is non one-to-one mapping of manipulations of the input device onto the controlled virtual objects, is a common way of interacting within virtual environments. Joseph J. LaViola from Brown University presented work showing that a possibly generalizable amplification factor of three is best for non-isomorphic 3D rotation in VEs (LaViola & Katzourin 2007).

Figure 3: Tactile feedback device by VW Group Research and Bauhaus-University Weimar.

One panel was dedicated to spatial perception in immersive virtual environments. Most of the contributions were dealing with distance compression, the fact, still not entirely understood, that users in virtual environments systematically underestimate distances when acting on the space. Various related aspects were discussed, e.g. the fact that recent research in psychology has shown that visual perception is influenced by a person's purposes, psychological state, emotions, and recent activities.

Several research groups and companies presented their latest developments at a small exhibition. Among them was the low cost, four camera, infrared based optical

tracking system *iotracker* from Vienna University of Technology (see figure 2) which also astonished industrial competitors with its high accuracy and large tracking volume of up to 40 m^3. A lightweight tactile feedback device for the fingertips which could replace cumbersome data-gloves was presented by Volkswagen Group Research and Bauhaus-University Weimar (see figure 3). It enables users to feel and interact with virtual objects using two fingers and the thumbs of both hands (Scheibe, Moehring & Froehlich 2007).

Industrial exhibitors tried to attract visitors with effective demonstrations of their latest VR technology, mostly Head-Mounted-Displays (HMDs), tracking systems, and software toolkits (see figure 4). The HMDs presented featured fields of view (FOV) ranging from 40° to 80°, with varying image quality and resolutions up to 1280x1024 per channel. The prices for professional VR systems, including HMD, tracking system, visualization software, and rendering hardware, are varying among the suppliers, starting from US\$ 50,000 to 80,000.

The VR community will meet again at this year's ACM Symposium on Virtual Reality Software and Technology (VRST) from November 12–14 in Newport Beach, California and ACM SIGGRAPH from 5–9 August in San Diego, California.

Figure 4: High resolution HMD and tracking technology presented at the booth of WorldViz.

References

Elmqvist, N. & Tsigas, P. (2007). Taxonomy of 3D Occlusion Management Techniques. In: *Proc. IEEE Virtual Reality 2007, Charlotte, North Carolina, 10–14 March 2007* (pp. 51-58). Los Alamitos, California: IEEE Computer Society.

LaViola, J. & Katzourin, M. (2007). An Exploration of Non-Isomorphic 3D Rotation in Surround Screen Virtual Environments. In: *Proc. IEEE Symposium on 3D User Interfaces 2007, Charlotte, North Carolina, 10–11 March 2007* (pp. 49–54). Los Alamitos, California: IEEE Computer Society.

Navab, N.; Feuerstein, M. & Bichlmeier, C. (2007). Laparoscopic Virtual Mirror - New Interaction Paradigm for Monitor Based Augmented Reality. In: *Proc. IEEE Virtual Reality 2007, Charlotte, North Carolina, 10–14 March 2007* (pp. 43–50). Los Alamitos, California: IEEE Computer Society.

Rueda, S.; Morillo, P.; Orduña, J. M. & Duato, J. (2007). On the Characterization of Peer-To-Peer Distributed Virtual Environments. In: *Proc. IEEE Virtual Reality 2007, Charlotte, North Carolina, 10–14 March 2007* (pp. 107–114). Los Alamitos, California: IEEE Computer Society.

Scheibe, R.; Moehring, M. & Froehlich, B. (2007). Tactile Feedback at the Finger Tips for Improved Direct Interaction in Immersive Environments. In: *Proc. IEEE Symposium on 3D User Interfaces 2007, Charlotte, North Carolina, 10–14 March 2007* (pp. 123–130). Los Alamitos, California: IEEE Computer Society.