

The original paper is published as:

Coomans, M.K.D, H.J.P Timmermans (1997). Towards a Taxonomy of Virtual Reality User Interfaces, Proceedings of the International Conference on Information Visualisation (IV97) 27-29 August 1997, London.

The paper is available in [postscript](#) (1.957 Kb), and [zipped postscript](#) (96 Kb).

Towards a Taxonomy of Virtual Reality User Interfaces

[M.K.D. Coomans](#) and H.J.P. Timmermans
[Eindhoven University of Technology](#), Faculty of Architecture, Building and Planning
P.O. BOX 513, mail station 20, 5600MB Eindhoven, The Netherlands

Abstract

Virtual Reality-based user interfaces (VRUIs) are expected to bring about a revolution in computing. VR can potentially communicate large amounts of data in an easily understandable format. VR looks very promising, but it is still a very new interface technology for which very little application-oriented knowledge is available. As a basis for such a future VRUI-design theory, a taxonomy of VRUIs is required. In this paper, a general model of human-computer communication is formulated. This model constitutes a frame for the integration of partial taxonomies of human computer interaction that are found in the literature. The whole constitutes a general user interface taxonomy. The field of VRUIs is described and delimited with respect to this taxonomy.

1. Introduction

Virtual Reality (VR) is getting an increasing amount of attention in both the scientific and business community. While at first only the entertainment value of VR was widely accepted, more and more people are getting now convinced of the potential of VR as a medium for easier communication between humans and computers. Our society is increasingly generating ever larger amounts of information. This information is also increasingly available in digital format, and world-wide accessible through on-line databases. This trend towards the availability of larger amounts of world-wide information raises the problem of how to use this data efficiently. VR is increasingly regarded as the most promising solution for this problem, as VR can potentially communicate large amounts of data in an easily understandable format. Virtual environments are expected to have an immense impact on our ability to deal with information [1],[2],[3].

The keywords to VR are visualization and natural interaction. The goal of visualization is to present data in ways that make them perceptible. Information is experienced instead of clumsily retrieved, by drawing on our vast experience in organising and interpreting massive sensory input, rather than purely on our linguistic thinking power.

The second key-characteristic, natural interaction, allows an easier response to the computer. Through drawing on our experience with real world interaction, users should no longer have to shift their attention from what they want to do, to how they have to do it. When the way in which we can perform certain actions is as trivial as grabbing a cup or indicating the size of a box, we can purely reserve our mental resources to the core of the task at hand. The combination of visualization and natural interaction suggests that VR-based user interfaces (VRUIs) may bring about a revolution in computing that is as important as the one caused by graphical user interfaces in the mid 80's.

VR looks very promising, but it is still a very new interface technology. Very little VR-specific scientific knowledge on how and when it can and should be applied is available. Several authors ascertain that the success of a VR application stands or falls with the quality of the chosen visualization and interaction paradigm [4],[1]. However, it is much harder to find information on how these paradigms must be designed and how candidates can be evaluated. Some considerable knowledge exists in disciplines that are relevant to VR : conventional UI design, cognitive psychology, ergonomics, etc. The major problem of VR research today seems to be the lack of general insight in the design space of VRUIs.

In this paper we propose a first version of a taxonomy of the VRUI design space. The purpose of this taxonomy is to serve, as a base for a theory on VRUI design. Such a theory would fit in the long tradition of HCI modality analysis. Clearly, the VRUI taxonomy should also be embedded in a general classification of user interfaces (UIs). This has brought us to the following sequence of research actions:

1. Find or formulate a general taxonomy of human-computer interfaces.
2. Delimit the field of VRUIs in this taxonomy
3. When necessary, refine the taxonomy for the specific subspace of VRUIs.

In this paper, we will discuss the results of steps one and two. Step three has not been addressed yet.

2. User interfaces classification

HCI (human computer interaction) modality analysis has a long history in specific media such as static graphics. More recently, the growing interest in multimedia has generated many studies on the usability of these new media. Unfortunately, most of the studies are directed at one particular kind of medium or interface component. They lack a general framework through which the many results can be linked and integrated. Taxonomy work on UIs is still in its infancy. Recently, a number of interesting publications on UI taxonomy have emerged, in particular by participants of the ESPRIT funded AMODEUS-II project [5],[6],[7], although they still do not address all parts of human-computer interfaces.

There are two possible approaches to a general UI classification. One is based on styles of UI design and describes their properties, strengths and weaknesses. Such an approach is excellent for understanding the diversity of UIs that are around. Backer and Buxton have presented an example of such a style overview [8]. The "styles" that are described in such overviews are always fuzzy and are partially overlapping because of the many possible UI components and the seemingly endless configurations in which they can be combined.

The second approach to UI classification is formulating a true taxonomy featuring a complete set of non overlapping categories. To establish such a taxonomy, a decompositional approach is necessary. UIs are decomposed into their generic components. Orthogonal and complete characteristics for each of these generic components are formulated, forming taxonomies for these UI components. The combination of all these taxonomies forms a new taxonomy for UIs as a whole.

In the next section, we will present a general model for HCI. Then, in section 4, this decompositional model will be used as a frame for the integration of partly taxonomies of HCI that are found in the literature. A proposal will be developed for some parts of the general model not covered by existing taxonomies.

3. HCI-model

In a human-computer co-operative process, both the user and the computer possess information which needs to be exchanged. The user either wants to import his knowledge to the computer, or he wants to tell the computer what data he wants to retrieve from it. The computer is supposed to deliver the requested information to the user, it can send status messages and ask the user for more information.

This information exchange in both directions has to go through a physical medium. In the physical medium, the information is represented as a specific energy state. For example, the user of a conventional computer typically retrieves information through looking at the screen or through listening at the sounds that the computer produces. These are respectively the visual and the acoustic medium (light energy and noise energy). In the other direction, the user sends information to the computer through hitting the keyboard or manipulating the mouse. Both rely on the physical medium of bodily movement (movement energy).

The computer is "connected" with these physical media through input and output devices. I/O Devices form the bridge between the physical media and the digital information carried inside the computer. Typical output devices are the computer monitor for graphics and speakers for sounds. A VR system might also have a glove for haptic output, etc. Common input devices are keyboard and mouse. Multimedia systems have in addition a microphone and a camera.

The physical signals received through the input devices are highly fragmented and contain no explicit semantics. For example, suppose a user wants to commit some numbers and words to a (conventionally equipped) computer. He might choose to do this through hitting the appropriate character-keys on the keyboard. In this way, he communicates the original word through its textual representation, using the physical means of button pushes. The input device (the keyboard) detects these physical signals, and describes them in a digital form, resulting in a list of separate characters. After that, an abstraction function is necessary to recognise the larger symbolic forms in this fragmented list: the words and numbers. The result of the abstraction function is still merely a representation and not the concept behind the representation. To retrieve the concept, the computer must use an interpretation function. The retrieved concepts can then be further processed, or stored in some internal computer representation.

An opposite two-step transformation process is necessary at the output side. For example, when a user wants to retrieve a list of numbers from the computer database, the information must be represented in some form. This could be a table with the numbers in textual format, or it could be a bar-chart with the numbers represented as bar lengths. The chosen representation must subsequently be specified or rendered into a raw description of the representation that the output devices can handle.

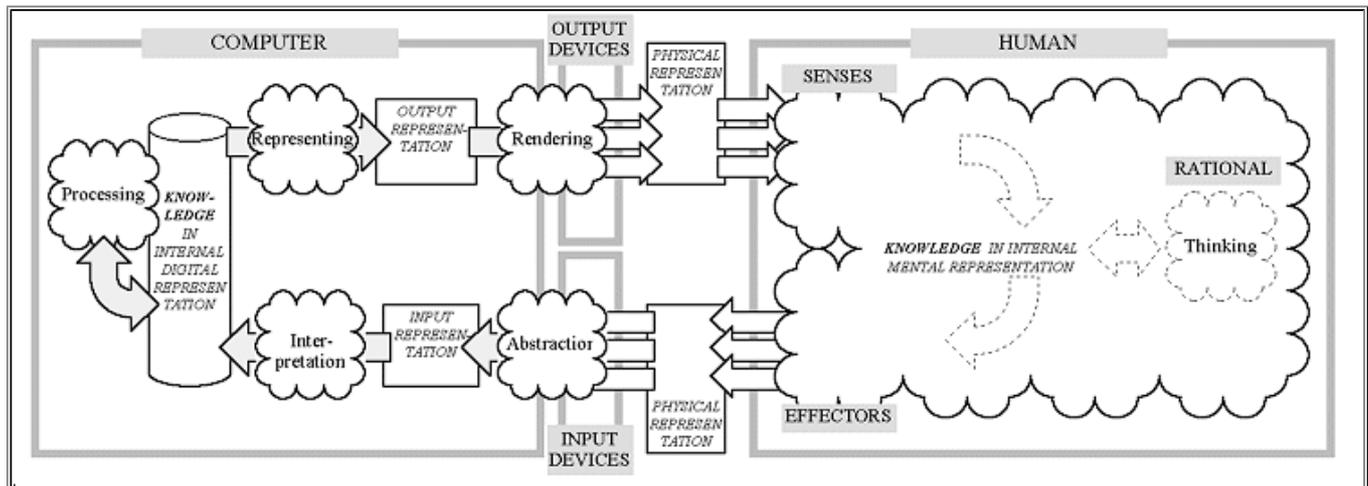


Figure 1—A decompositional model of human-computer interaction

In principle, all computer input and output require this two-step transformation: abstraction and interpretation for input, representation and rendering for output. The representations between two transformations are called representational modalities [5]. These are modes or ways of representing information to humans or computers, realisable in a physical inter-subjective form.

Figure 1 shows the described process of information transformation that the computer performs during computer-human dialogue. As suggested in [6], humans might perform similar symmetrical manipulation of information (knowledge). In this HCI-model however, we stick to the more basic scheme. Information is perceived through the senses, exported through effectors, and is internally processed and stored in distinct mental representations.

4. UI Taxonomy

The presented model defines the different information states, the transformation processes, and the devices involved in human-computer interaction. In the next sections, we will present a taxonomy for each of these UI components.

4.1 Taxonomy of physical media

Bernsen (1995,1996) proposes to discriminate between the following physical media :

Output media:

1. Graphics
2. Acoustics
3. Haptics

Input media:

1. Kinaesthetics (body movement of the user as input to the computer)
2. Acoustics (voice input)
3. Graphics (video capturing with possibly gesture recognition, etc.)

As Bernsen points out, these are the media through which all existing human computer interaction takes place, including VR systems. More media (that are purely theoretically possible) could be added but that would not serve the practical usability.

4.2 Taxonomy of representational modalities

Bernsen (1995,1996) has presented taxonomies for unimodal representational modalities for input and output. A unimodal modality is an un-combined way of representing information. Representations that are composed of two or more unimodal representation forms are called multimodal. For example, charts are usually multimodal as they are composed of a graph with textual annotations. The pure graph and the annotations are both unimodal.

Unimodal modalities for both input and output are characterised by a physical medium and a profile. The physical medium is the medium in which the modality can be realised (output), or from which the modality is abstracted (input). The profile consists in a set of characteristics selected from the following list of binary opposites : linguistic/non-linguistic, analogue/non-analogue, arbitrary/non-arbitrary, static/dynamic.

On the basis of these characteristics, Bernsen defines 20 categories of representational output modalities. These can be grouped to the following 4 "super-level"-categories :

1. Linguistic modalities. Linguistic representations are based on existing syntactic-semantic systems of meaning. Linguistic representations can secondarily be analogue or non-analogue. Examples of analogue linguistic modalities are (some) hieroglyphs (analogue graphics), and the gestural language that people use to show someone the way (analogue graphics). Examples of non-analogue linguistic modalities are written words and numerals (non-analogue graphic), spoken words and numerals (non-analogue acoustic), touch words and numerals such as Braille (non-analogue haptic)
2. Analogue modalities. Analogue representations represent through aspects of similarity between the representation and what it represents. Examples are maps, cartoons, photographs (analogue graphics), and Geiger counters (analogue acoustic).
3. Arbitrary modalities. These are non-linguistic and non-analogue representations that do NOT rely on an already existing system of meaning and representations. Examples are diagrams consisting of geo-metrical elements (arbitrary graphics), the "biceps" that computers produce when an error occurs (arbitrary acoustics).
4. Explicit structures. Non-linguistic and non-analogue representations that do rely on an already existing system of meaning and representations (other than linguistic and appearance based). Examples are form fields, frames, table-grids, line separations, windows,...(graphic structures).

The preliminary version of Bernsen's taxonomy of representational input modalities contains four similar super-level categories, and 36 categories at the "generic" level.

4.3 Taxonomy of human-computer dialogue semantics

The previous taxonomy addressed the representations of exchanged information. In a human-computer dialogue, the way in which things are said will be closely related to what is said. Therefore, a taxonomy of the content of human-computer interactions seems equally important. Unfortunately, little work can be found on this topic. The following list constitutes a basis for a taxonomy of human-computer dialogue contents:

1. Dialogue control information. The part of the human-computer dialogue that is not concerned with the actual task at hand but with the dialogue itself. Sub-types with examples :
 - 1.1 Dialogue flow. Signing on and signing off, "I'm waiting", "I'm active",...
 - 1.2 Dialogue channel. Create new view, select view style, select input mode
 - 1.3 Task history and context. Undo, save, retrieve particular state,...
2. Task information. The part of the dialogue that addresses the specific task at hand. This part can be expected to be the largest part. A general classification of task-related information seems not feasible as large differences must be expected between different application domains.
3. Domain knowledge. Some part of the dialogue might be about domain-related information itself. This is the case in certain programmable applications in which the user can add and adapt new domain-related knowledge. For example, he might define new domain-specific object-types, new procedures for dealing with them, etc. As is the case for task information, a general classification of domain information seems not feasible.

4.4 Rendering and abstraction functions

Currently, no taxonomy of rendering and abstraction functions can be presented.

Some classification could be made on the basis of the properties (1) the degree of abstraction/rendering, and (2) fusion/fission. Both are derived from the Multi-Sensory-Motor framework discussed in (Coutaz,1993).

4.5 Representation and interpretation functions

For representation and interpretation functions, no taxonomy is currently available either. For these functions, some classification could be made on the basis of the fusion/fission property as well. Further, representation functions can be divided in the following four distinct groups :

1. Straightforward demonstration. Pure analogy based representation (through picture, sound recording, ...). Can only be used for information with a native perceptible character.
2. Metaphorical demonstration. Representation on the basis of an metaphor.
3. Linguistic description. Describing characteristics in a text.
4. Word representation. Representing the linguisting term.

5. VR User interfaces

What is VR and what are VR user interfaces ? Many different definitions of VR can be found in the literature. The following list resembles most aspects found in VR definitions (see also [9] and [3]):

1. Interaction. Natural user interaction .
2. Immersion. Sensory immersion in a virtual environment. Full body immersion or partial immersion. Also tele-presence.
3. Simulation. Visual, acoustic, and haptic simulation; also scientific simulations and virtual prototyping.
4. Visualization. Making native non-visual information visual.
5. Real space. On-line communication and infor-mation retrieval, from anywhere, without time delay.
6. Autonomous agents. The computer starts behaving autonomously, takes own initiative.

Of these characteristics are 1,2, a part of 3, and 4 related to the UI. All other have no relation to the UI : real space (5) relates to the networking capabilities; autonomous agents (6) are related to a kind of artificial intelligence; scientific simulations and virtual proto-typing (2nd part of 3) are purely relate to information processing capabilities (e.g. Finite element analysis.).

In the following, we will discuss the UI related topics in the framework of the presented taxonomy.

5.1 Interaction

By interaction, we mean the mutual response of computer and user on each other's actions. Humans and computers can interact in many ways. Typical for VR-interaction is, that it mimics the way in which we control our real world environment. This has the obvious advantage that we all know how we can perform this type of control thanks to our everyday experience.

In order to appear natural , interactivity must have the following characteristics :

1. Sensitivity to natural commands : the system must understand actions that appear natural to the user. Of course, adequate physical input devices are essential for this purpose.
2. Real-time feedback : a user's action must immediately be reacted upon. This requires a lot of processing power : abstraction, interpretation, semantical processing, representation, and rendering must be performed in fractions of a second.
3. Natural feedback : the system must respond in the way that appears natural to the user. Again, the appropriate output devices must be available.

As examples of interactions that meet the three characteristics, we can think of navigating through a space and manipulating an object with our hands.

5.2 Immersion

By virtue of its natural interactivity, VR can produce a feeling of immersion. Immersion is the feeling of being deeply engaged. Participants enter a make-believe world as if it is real.

In order to realise this illusion, VR systems imitate the natural environment using real-life like stimuli for all the human senses: images, sounds, as well as synthetic tactile stimuli, and possibly smells. They also monitor as detailed as possible how the user responds on that environment with bodily movement; voice input, or object manipulation. Specific devices have been developed to facilitate VR-immersion: head-mounted displays, 'caves', 6D mice, touch simulators, etc.,

Immersion is not an exclusive characteristic of VR. It can also be realised through more conventional media like film and literature. The differences between these media can be reduced to differences between representational modalities. VR establishes a virtual environment through a coherent visual, acoustic and haptic representation, responding to the user's actions in a way that appears 'natural' to the user. Film is limited to the establishment of a visual and acoustic representation of the virtual environment; it lacks the interactivity of VR. Literature provides a purely linguistic description of the virtual environment. Such a linguistic representation not only lacks the interactivity but also the analogue character of VR. The virtual environment is described, not shown, nor offered to participate in.

5.3 Simulation

Mathematical simulation processes deliver the realistic representations of the virtual environments. In the terms of the presented HCI-model, simulation processes are rendering functions.

The realistic presentations are derived from abstract representational descriptions that present the information that the user is interested in. These abstract descriptions are turned into detailed physical representations according to the physical laws of light, sound, gravity, tactile feedback, and so on. Of course, realistic simulation can only be experienced by the user if the output device can deliver the result of the simulation effectively.

5.4 Visualization

Visualization stands for representing information in a visible format. This can be either through straight-forward demonstration or through metaphorical representation. Visualization eases in many cases the interpretation of the information by the user. In the case of straightforward demonstration, this is due to the analogue character. In the case of metaphorical visualization, the advantage flows from the special capabilities of our sensory system, such as pattern recognition.

In VR, the term Visualization is usually broadened to the representation of information in an integrated visible, audible, and touchable virtual scene. Perceptualization might actually be a more apropos term in VR [1].

6. Concluding discussion

In the previous section we have discussed the main aspects of VRUIs. Typical for VR are analogue representational output modalities that integrate visual, acoustics, and haptics. VR uses specific input and output devices that can stimulate all senses and capture all behaviour of the user. Everything a user does can represent valuable input for the system. System output must in no time respond to new input of the user.

It is important to notice that both interaction and immersion indicate an equal importance for natural input processing capabilities and natural output processing capabilities. This suggests that input-abstraction should get an equal amount of attention as simulations.

These ascertainments prove that an analysis of the capabilities of VRUIs requires an integral UI taxonomy. Choice of input and output devices, representational modalities, rendering and abstraction functions, interpretation and representation, seem all important for the quality of the interface. The general model and taxonomy presented in sections two and three, have successfully been used to give a general description of VRUIs. A more fine grained analysis of different types of VRUIs clearly requires more detailed framework with well worked-out taxonomies for all its components.

References

[1] T. Erickson, "Artificial Realities as data visualization environments: problems and prospects", in A. Wexelblat,

ed., "Virtual Reality - Applications and explorations", pp. 3-22, Academic Press, London, 1993

[2] F. Biocca, M.R.Levi, "Virtual Reality as a Communication Medium" in F. Biocca and M.R.Levi, eds., "Communication in the Age of Virtual Reality", pp. 15-32, Lawrence Erlbaum Ass., Hove, UK, 1995

[3] D.N. Chorafas, H. Steinmann, "Virtual Reality: Practical applications in business and industry" (sections 1 and 2), Prentice Hall, London, 1995

[4] S.Bryson, "Approaches to the successful design and implementation of VR applications", in R.A.Earnshaw, J.A.Vince, H.Jones, eds. "Virtual Reality Applications .", p.328, Academic Press Limited, London, 1995

[5] N.O.Bernsen, "A toolbox of output modalities", 1995, the Amodeus-II WWW-site, <http://www.mrc-apu.cam.ac.uk/amodeus/>

[6] J.Coutaz, L.Nigay, D. Salber, "Taxonomic Issues for multimodal and multimedia interactive systems", 1993, the Amodeus-II WWW-site, <http://www.mrc-apu.cam.ac.uk/amodeus/>

[7] N.O.Bernsen, "A taxonomy of input modalities", 1996, the Amodeus-II WWW-site, <http://www.mrc-apu.cam.ac.uk/amodeus/>

[8] R.M.Baecker, A.S. Buxton, "Interaction Styles and Techniques", in R.M.Baecker, A.S. Buxton, eds., "Readings in Human-Computer Interaction: A Multidisciplinary approach", pp. 427-434, Morgan Kaufmann Publishers, San Mateo, 1987

[9] M. Heim, "The Metaphysics of Virtual Reality", Oxford University Press, Oxford, 1993



Eindhoven University of Technology.