User-system models & believability

Institute for Perception Research
PO Box 513, 5600 MB Eindhoven

JW/jw 94/27
17.11.1994

Rapport no. 1014

User-system models & Believability: discussions in the multimedia SIG

J.H.D.M. Westerink
and E.J.A. Verheijen

User-system models & Believability: discussions in the multimedia SIG

J.H.D.M. Westerink
E.J.A. Verheijen

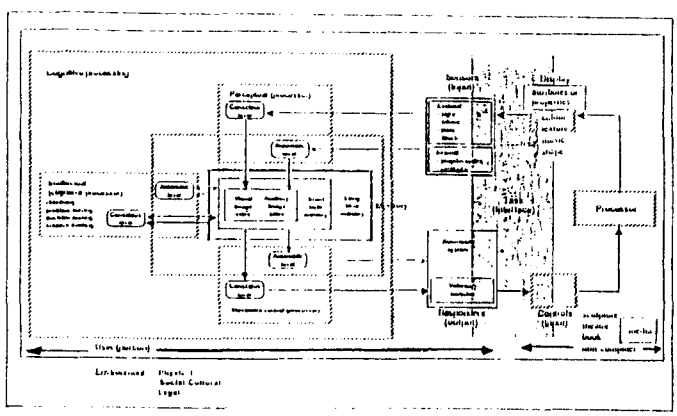
Summary:

At IPO a number of special interest groups (SIGs) exist around various topics which try to be a forum for discussion within the institute. One of those groups is the SIG on 'multi-media and multi-modality'. Within this group, focus has been on two distinct topics:

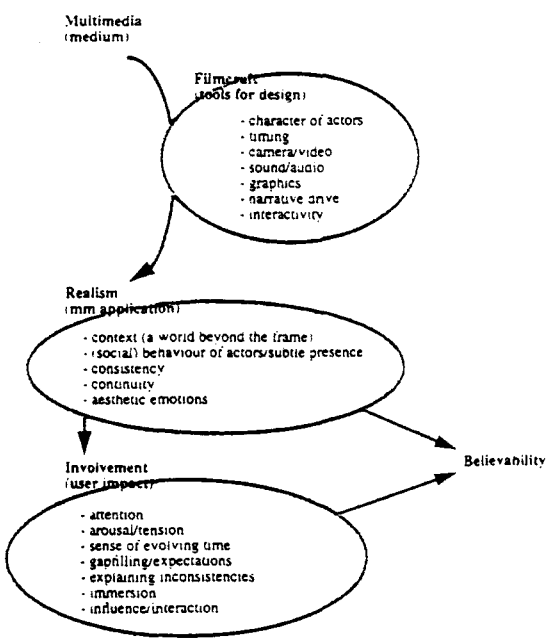
- user-system models for interaction with multi-media systems,
- believability of multi-media products.

This report summarizes the MM-SIG's discussions and activities around these topics, and presents some tentative conclusions:

- For the discussions on the multimedia user-system interaction model, the conclusions take the form of a proposed model (see figure 1a).
- For the discussions on believability, the conclusions take the form of a list and clustering of important aspects of believability and related terms (see figure 1b), and some ideas for measuring believability.



1a



1b

Fig. 1: Impression of resulting models
Figure 1a gives the multimedia user-system model; figure 1b organizes organizes our list of believability aspects. Readable versions of these figures appear on pages 7 and 13, respectively.

Contents

1 Introduction

At IPO a number of special interest groups (SIGs) exist around various topics which try to be a forum for discussion within the institute. One of those groups is the SIG on 'multi-media and multi-modality'. Regular members of this group in the past year were: David Connah, Anita Cremers, Cynthia Grover, Berry Eggen, David Keyson, Han Kohar, Judith Masthoff, Alp Tiritoglu, Ellen Verheijen, Joyce Westerink and Remko Westrik.

At the start of the MM-SIG, we made an inventory of what we thought were interesting topics within the wide field of multi-media and multi-modality. This led to a focus on two more or less separate topics:

- user-system models for interaction with multi-media systems (september - november 1994),
- believability of multi-media products (december 1993 - may 1994).

This report is mainly intended to help the members in consolidating the things we've learned and touched upon in our discussions. Therefore, they are reflected more or less chronologically. The report captures and summarizes the MM-SIG's discussions and activities around these topics, and presents some tentative conclusions.

2 User-system models for interaction with multi-media systems

At the start of our SIG on multi-media & multi-modality, we though it would be relevant to come as close as possible to a definition of the concepts media and modality. We have attempted to do this in the form of a model for interaction with multimedia systems, that accomodates both. The model was slowly built up and discussed in a number of sessions in september till november 1993. They are described below.

2.1 Models in literature

We first set out to see which models for interaction with multi-media systmes exist in literature. To this end we read the introduction to a number of books about multi-media (Mayes, 1992, Von Wodtke & Mark, 1993, Taylor, Néel & Bouwhuis, 1989, Clarck & Graig, 1992, Barker, 1989, Laurel, 1991, Blattner & Dannenberg, 1992, Waterworth, 1992).

We wanted to see which frameworks were presented for understanding multi-media and multi-modal interactions and to which extent those could be understood in terms of the most common model of user-system interaction (see figure 2). It turned out, however, that only few of the books expressed a view on multi-media that lended itself to be summarized in terms of this model. Instead, various books highlighted various aspects as important:

- Action and interaction are the most important elements of an interface, as they allow the user 'do' something. They have the power to engage people as a drama on stage does.
- Multimedia applications are hard to capture in one generalized framework, because content, as well as structure as well as affordances are critically dependent on the task the user will perform with it.
- Dual Coding Theory describes that verbal and visual information are organized in different cognitive systems. Based on this it is sometimes claimed that multi-media would be particularly suited for learning as it addresses both cognitive systems. Research, however, rather

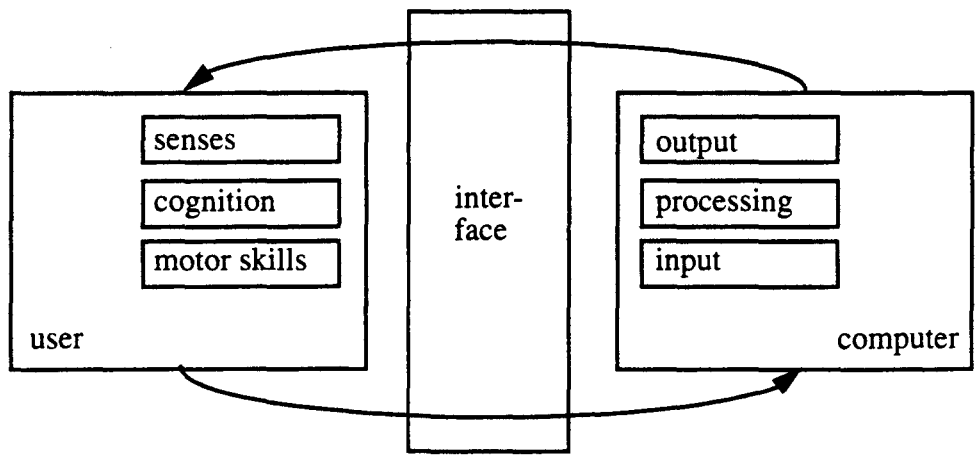


Fig. 2: General model of user-system interaction (simplified version)

suggests that it is the **interactivity** that is responsible for learning gains measured in multi-media environments. This is because interactivity affords learner control, and enhances participation adn motivation.

- Important is the way the computer is (presented to and) perceived by the user. Two basic scenarios are the computer as dialogue partner, or the computer as tool. Of course, a combination is possible as well.
- The terms medium, modality and sensory modality are often used, but no general consensus about their meaning seems to exist. Thus they can give rise to considerable confusion.

2.2 Our 'single medium' model

We started a discussion to see which concepts and words we would be needing in our SIG to avoid confusion about the terms medium, multi-media, modality and sensory modality. We came up with a simple model (see figure 3), that we usually call the sculpture-model, as it is based on an analogy with a sculpture made by an artist and viewed by a person. The most important characteristic of this model is that it treats a multi-media computer as *one single medium* for expression, and considers things like text and sound (that are traditionally viewed as separate media) as attributes of the application.

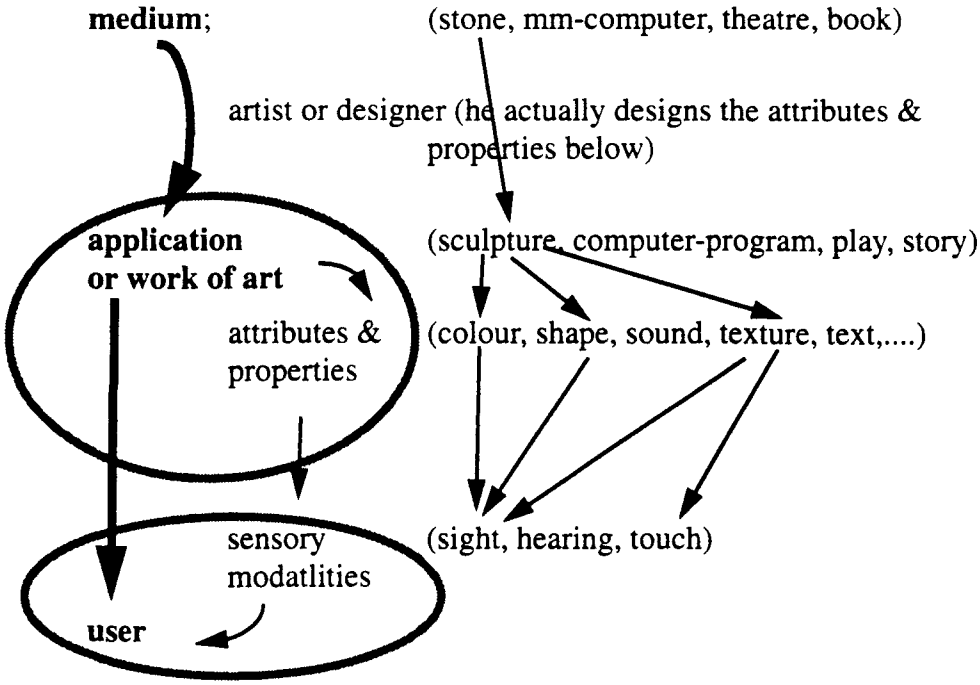


Fig. 3: The 'sculpture' model

In a discussion of the sculpture model we came up with several suggestions for imporvements. We agreed that an improved model should have a closed loop allowing user input into the system, that it should represent the sensory modalities as part of the user, and that cognitive models of the user are very interesting. Also we were interested to see whether it was possible to merge the sculpture model with the general user-system interaction model of figure 2. The combined result is presented in figure 4.

Reviewing the final model of figure 4, our discussion quickly revealed that there were still an number of areas that were not represented, yet very important for the understanding of multimedia. We came to a list of 5 issues, and for some of them attempts (in subgroups) were made to try to incorporate them:

- **Contents & form**
  Because content and form are determined by the application, it seemed appropriate to elaborate the system-side of the model. Also, an influence of content on form (presentation) seemed more important than the reverse. Appendix A presents an elaboration of this.
- **Believability & illusion**
  We did not succeed in incorporating elements of believability into the model. This indicates that this aspect is new in thinking about user-computer systems. For this reason it was decided to spend a series of MM-SIG meetings on this topic (see section 3).
- **Combination of sensory modalities & of processing activities**
  Several elements were suggested to elaborate the user-side of the model. The most important ones are, attention, parallel or sequential use of different modalities, and their consequences for coherence and expectations. More detail is provided in appendix B.
- **Emotions**
  Positive and negative emotion can both be relevant in any application. How to design for them?
- **Goals & environment**
  They determin any multi-media interface. Maybe it is possible to classify them.

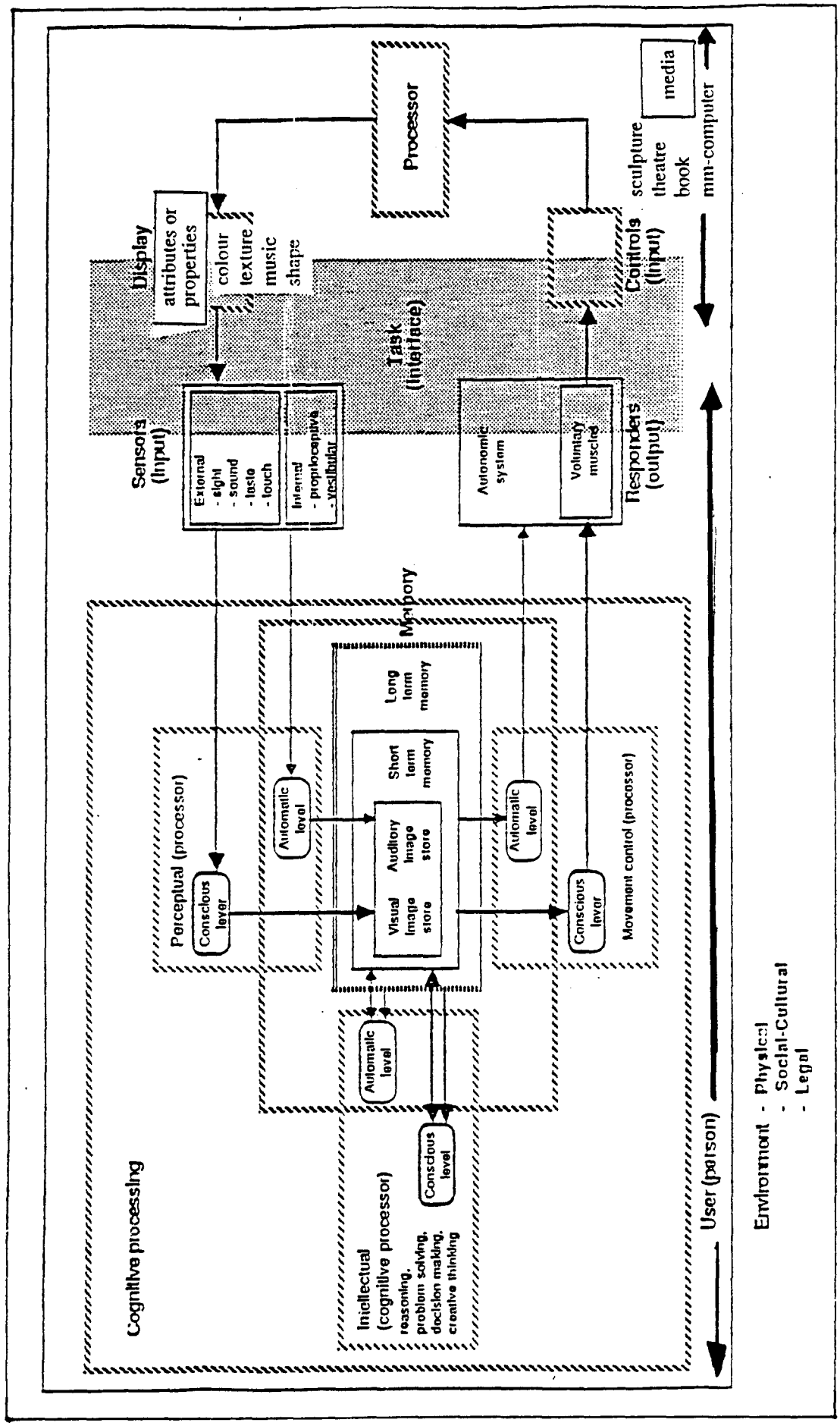


Fig. 4: Final, merged, model of multi-media user-system interaction

3 Believability of multi-media products

One of the issues that we found difficult to represent in any model for user-system communication was believability of the system, yet we agreed it would be very important for a good communication. This was, combined with curiosity for new elements, for us a reason to investigate the topic of believability in a few sessions. The sessions took place about monthly from december 1993 till may 1994. They are described below.

3.1 Elements of believability

David Connah presented his view on believability as an introduction to the topic. His interests mainly lie in the creation of believable worlds, and his main points were:

- Believability has always been important in any production (e.g. a book), but has now gained in importance because in interactive products, the designer has no full control over his characters any more: they have become autonomous (agents) and the user is allowed interaction.
- It seems that **gap filling** is an important element in believability; it is needed to evoke the right imaginations and suggestions. If something is unbelievable, the user/viewer tries to fill in the gaps in order to arrive at a coherent story (Eagleton, 1983). If one possibility exists to fill in the gap, the story is believable and a certain suggestion is brought across. If no possibilities for filling the gap can be found, believability breaks down. If more that one possibility exist, there is an arc of tension (e.g. in a detective story). Here the matter of perspective is important.

A compilation of videos was presented that touched upon elements of believability:

- Walt Disney's Junglebook has believable **characters**, even the snake who takes the form of a staircase.
- Andre van Duin's voice-overs on a clip with monkeys turn an average zoo scene into a running contest for monkeys.
- The time-perception during productions (e.g. a commercial for Aviko french fries) can be very dependent on whether or not there is **sound** that builds up tension.
- **Audio** can also be used to portray an environment that is not visible (e.g. the chameleon commercial for Trouw).
- The rendering of the 'Alice' computer game is definitely unbelievable at some points. It is curious to see that this can be regarded as funny, adds some sort of **humour**.
- The 'Voyeur' CD-i has an **evolving time-scale** (time goes on). Although, or rather because, this means you can miss things, it brings a feeling of believability in that it mimics the real world.

In the discussion we found that concepts as **involvement** and **immersion** are close to believability. We wondered whether theatre critics have criteria for believability, and thought of the concept of 'willing suspension of disbelieve' (Coleridge, 1907). Drama was thought to be different in that it did not allow interactivity. We wondered whether the addition of interactivity would disrupt the viewers involvement because conscious decisions are required on an other level of abstraction, or whether it would help to achieve a higher involvement through personalization of interests.

3.2 Interactive drama

In this meeting an article about interactive drama was discussed, entitled 'Dramatic Presence' (Kelso, Weyhrauch & Bates, 1992). Some points emerged from the article and the discussion:

- The authors state that one of the main problems of interactive drama will be the lack of direction in the plot: will it have a build-up of **tension** toward a final destiny? There has to be a person or program module that specifically monitors and manipulates the emotions evoked in the interactor, even if he can interact in the plot himself. This function is the interactive equivalent of the director's function in film craft.
- It turns out to be important to be able to explain inconsistent behaviour in the plot, because these will be inevitable. This is somewhat linked to the issue of filling up gaps. According to the article, subjects (interactors) had only few problems in finding possible explanations in seemingly **inconsistent situations**.
- The perception of **time** seems to be very different for the interactor than for people in the audience. Though the story progressed only very slowly in the eyes of non-involved viewers, the interactors did not seem to be aware of the many large parts in which nothing happened.

3.3 Film craft

In a cluster of 2 MM-SIG sessions we touched the topic of film craft. This was done using the CHI tutorial 'Film craft in user interface design' (Young & Clanton, 1993). The tutorial presents a number of techniques that are used in the world of film making, and give examples of how these techniques can also be used to make and improve user interfaces. We specifically had a closer look at 3 of the 6 chapters, namely 'camera and composition', 'editing', and 'animation'. In the second session, we discussed (after preparation in subsgroups) to what extent these techniques had been used in the CD-i 'Treasures of the Smithsonian'.

The lessons learned from the combined sessions are:

- The rules and guidelines in the tutorial can be seen as the grammar of a sort of film language. As in spoken language, they can be broken for the purpose of creating an enhanced effect. Here the intended effect on the viewer is the starting point, and the most suited film craft technique is then selected to achieve it.
- Because both film craft and multimedia have visual components, most of the film craft guidelines for camera and composition are directly applicable to multimedia products as well. Especially techniques to suggest **a world beyond the frame** and to give you the impression you are there, are important to create believability. These are used occasionally in the 'Smithsonian CD-i', but not often enough to support the metaphor of 'visiting a museum' as you cannot more or less free walk around in space. The result is that believability breaks down: this is no museum, it's an encyclopedia.
- Editing is usually needed for manipulating time while keeping **continuity**. Through editing the film director manipulates both the **running time** (acutal film length) of the video as well as the **experienced time**. In a true multimedia application the running time is in the hands of the user through interactivity, as is the case in the Smithsonian CD-i (e.g.the user of skip-buttons as well as the Bell X-1 demo). It is not necessarily a problem if the experienced (story) time is less than running time, but care has to be taken their ratio doesn't drop under a

minimum. For example, the loading of certain image sequence takes so long (running time) while almost nothing happens (experienced story time near zero) it disrupts the involvement of the user, and thus believability breaks down.

- Walt disney type animation has to do with giving character and live (animate) to artificial objects, and how to do this in a believable way. In the smithsonian CD-i, animation techniques are used to portray objects in motion, mostly inanimate objects. Here the added sound helps to convey a feeling of reality (one could think of sqeaking sounds to accompany the closing drawers). The objects, however, are animated only in a technical sense: this serves to show the importance of **character** for true animation.

3.4 Film and semiotics in film

Hans van Driel, working at the KUB university in Tilburg, gave a presentation about 'semiotics in film' (Van Driel, 1989). He describes the interpretation of film in the framework of Peircean semiotics: when is something (including film material)a *sign*?; how do signs *refer*?; and eventually, how do they get *meaning*? Applied to the making of a film, Van Driel presented the following concepts:

- In the world of film, there is a difference between *how* the film is made (camera) and *what* is filmed (the ante-filmic): **form versus content**. This distinction immediately leads to two types of films. Firstly, there is the Hollywood film, which more or less has its own film language that was developed until roughly the 1940-ies, and in which the story (the what) is the important force. All technical aspects (the how's) are supposed not to **attract attention**, but to support the believability of the world created. The (European) art-film, on the other hand, focuses on breaking with the standard conventions through putting emphasis on the techniques. Looking at 'the how' immediately brings about a greater distance to and less involvement with the subject of the film.
- Some effects that are used as conventions in film language are verified experimentally. There is for instance the effect of **context**: a neutral close-up of an actor was juxtaposed to a shot of a child, a plate of soup, and a woman's corpse. Subjects reported the face of the actor expressed happiness, hunger and grieve, respectively.
  As for movement, motion from left to right is perceived to portray forward motion, motion from right to left indicates backward motion (e.g. a withdrawing army).
- Van Driel did not foresee a great future for interactive film, because he basically believes that people want one-way street entertainment. Again the distinction between what and how is relevant, because the interactivity can be used to influence the story (what) of the point of view (how), or both. Van Driel only expected that maybe people would like to select the information portrayed (e.g. select whether you pay attention to the host or to the guest in a talk show), but they would like to keep the story intact.

3.5 Believability of synthetic actors

Guy Dugdale, who has considerable experience as a film director, and is now involved in the interactive-film project at PRL (Rankin, 1994), gave a presentation about 'the believability of synthetic actors' (Dugdale, 1994a). In this presentation he also interweaved some of his views

about Bates' work on interactive drama (see section 3.2 and Dugdale, 1994b). His main points are:

- An interactive story has two, more or less orthogonal, driving forces: the **narrative drive**, that builds up tension and works toward some sort of conclusion, and the **interactivity**, that allows personalization of the story and thus some degree of uncertainty. As a result of these two forces, the story develops within the boundaries of some sort of 'narrative corridor' (see figure 5), beyond the edges of which the users are not allowed to go. One of the problems is how to present these edges to the users in a natural way.

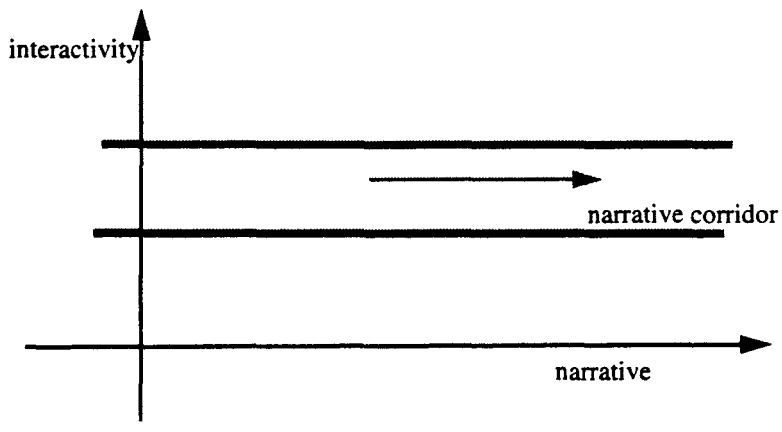


Fig. 5: The narrative corridor

- Important elements of the believability of actors are the **aesthetic emotions** they portray (explanations of our real world), and their '**subtle presence**', which is evoked by various seemingly irrelevant details (like smell, body language, gestures, etc.). These will also be important elements for the believability of synthetic actors, and to give the audience the conviction that there are dramatic characters there. In this, it is not necessarily the optimum to be maximally similar to real humans; the 'actor' in the answering machine for instance can be really threatening.
- The viability of interactive film or drama is not yet proven. If it does come to existence on a somewhat broader scale, Dugdale expects its major impact to be that then over lunch, people will discuss the various versions of a story, and the differences between them. Also fiction might get networked, as well as games.

3.6 Technology and believability

We had a look at several multimedia productions with high-quality video and audio assets, to see what is their impact on believability. To this end Loek Sanders (IMS) gave a demonstration of some of the newest titles in this area. in our discussion, the following issues were discussed:

- **High quality graphics** (as in the '7th Guest' CD-i) or the inclusion of **video** ('Othello') certainly enhance the believability of the game. The same goes for **sound**. We wondered, however, whether these assets would continue to be so important once you're involved in the

game, and just want to play. Then the elements that added to believability at first encounter (like the typical character traits of the Othello players), might even start to annoy as they interfere with the flow of interaction.

- As the quality of video and audio assets gets better, the need for less-constrained **interaction** possibilities gets stronger. In the 'Super Mario' CD-i (the speedboat in the swamp), there is a continuum of paths possible for the user, rather than a set or sequence of selectable hotspots, that may or may not be visible. Indeed, if the video and audio assets go to great lengths to depict a world for immersion, the interactivity should not remind them of the technical constraints of the world 'outside'. Part of the unconstrained interaction is the input device that is used. In de 'Maddog mcCree' CD-i for instance, a gun-like input device (e.g. an Airmouse) would certainly less disrupt the user from his experiences in the Wild West, than a trackerball, which make him pay attention to technology.

3.7 Conclusions

In a final session, we tried to make an inventory of all elements identified so far, that are connected with believability. They are in boldface and underlined (like **this**) in the preceding sections. They are repeated in the lower part of figure 6. We agreed that they will have their impact on believability, as well as on involvement and realism. The way in which the latter three are related was not fully clear: some people judged believability to be one of the contributors to involvement, whereas others thought things were the other way round (involvement is part of believability). Also the importance of expectations for believability and involvement was discussed.

We did a little brainstorm on ways to measure 'believability' in an experiment with users. Many of the ideas generated depended to some extent on the triangular relation between believability, involvement and expectations:

- To assess the believability of the sound that conveys the footsteps of a person, ask the subjects to estimate this person's weight. To assess the believability of a story, let the subject estimate the 'real' time lapse or the 'real' distance in the story.
- Measure believability through the existence of expectations with the subject: stop him at a certain point and ask what comes next.
- Measure believability through the involvement of the subject: generate irregular beeps in the experiment room, and ask how many he has heard at the end. Or, ask him to estimate the amount of time that has elapsed since he started ('time flies', 'to forget time').
- Measure non-believability: ask what the subject thinks was the least believable in the multimedia application, or observe what disrupts him most or most frequently.
- It might be interesting to investigate the evolvement of believability over time in a longer session with a multimedia product.

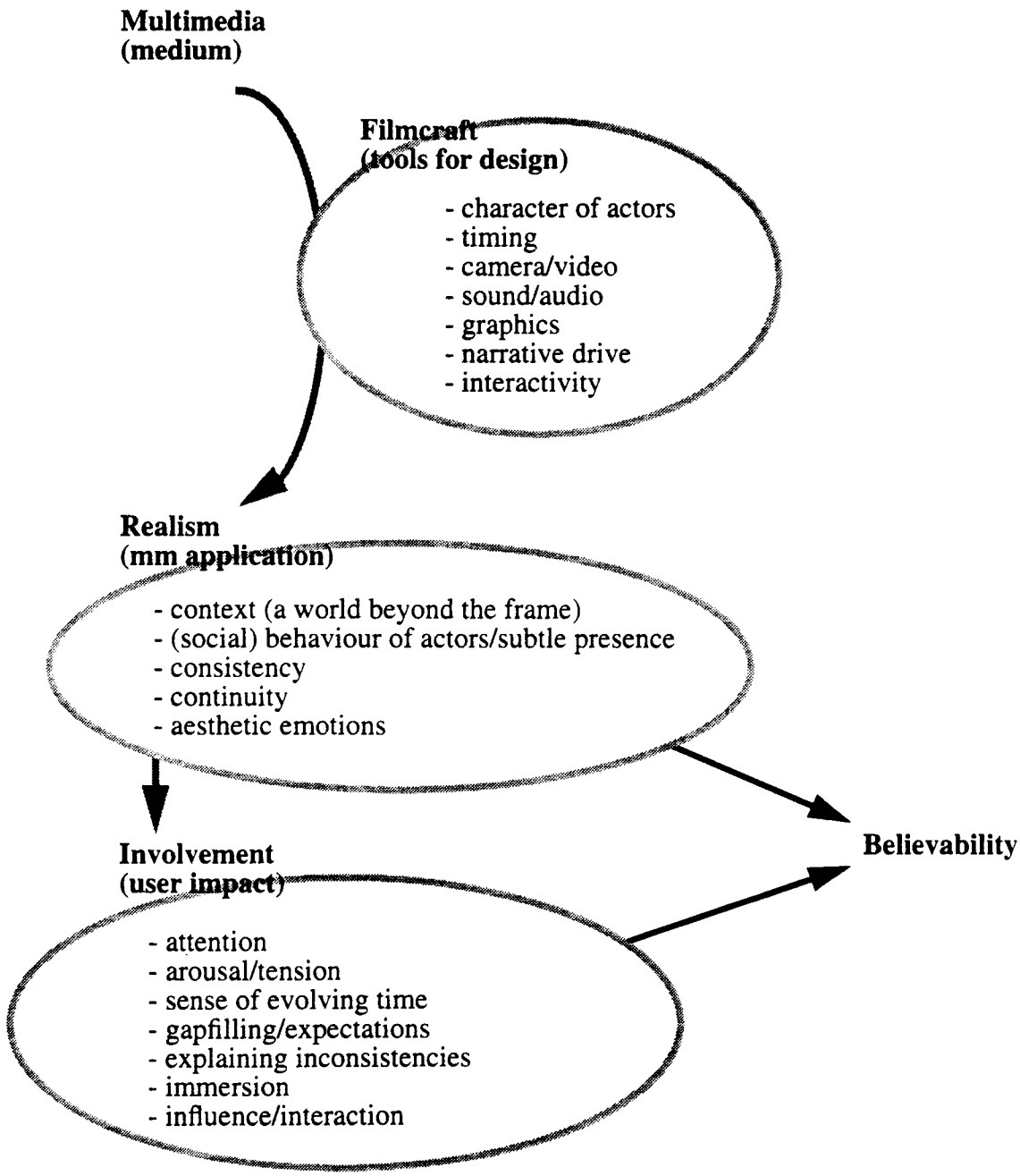


Fig. 6: Overview of important aspects in believability
The figure is modelled after the general multimedia model depicted in figures 3 and 4. The aspects mentioned emerged as important notions in sections 3.1 through 3.6.

References

Barker, P. (1989), *Multi-media computer assisted learning,* Kogan Page, London.

Blattner, M.M. & R.B. Dannenberg (1992), Multimedia interface design, Addison-Wesley, Reading.

Clarck, R.E. & T.G. Graig (1992), *Research and theory on multi-mdeia learning effects,* in: Interactive multimedia learning environments, ed. M. Giardina, NATO ASI series Vol. F93, Springer Verlag, Berlin.

Coleridge, S.T. (1907), Biographia Literaria, Oxford.

Dugdale, G. (1994a), *The believability of synthetic actors,* personal communication.

Dugdale, G. (1994b), *The view from Kansas, a consideration of Joe Bates' Oz project,* personal communication

Eagleton, T. (1983), *Literary theory*, Blackwell, pages 74-77.

Kelso, M.T., P. Weyhrauch & J. Bates (1992), *Dramatic Presence,* article submitted to presence, personal communication.

Laurel, B. (1991), *Computers as theatre*, Addison-Wesley, Amsterdam.

Mayes, J.T. (1992), *The M-word: multimedia interfaces and their role in interactive learning systems,* in: Multimedia interface design in education, eds. A.D.N. Edwards & S. Holland, NATO ASI series Vol. F76.

Rankin, P.J. (1994), *Hands-on movies: Techniques for interactive films,* Phlips Technical note no. 3206

Taylor. M.M., F. Néel & D.G. Bouwhuis, 1989, *Structure of multimodal dialogue including voice*, North-Holland, Amsterdam.

Van Driel, H. & M.C. Westermann (1989), *Semiotiek van de audiovisuele media*, Syllabus KUB (Tilburg university)

Von Wodtke, & Mark (1993), *Mind over media, creative tinking skills for electronic media,* McGraw-Hill, USA.

Young, E. & C. Clanton (1993), *Film craft in user interface design*, INTERCHI'93 Tutorial Notes, April 24-29, Amsterdam, The Netherlands.

Waterworth, J.A. (1992), *Multi-media interaction with computers; human factors issues,* Ellis Horwood, London.

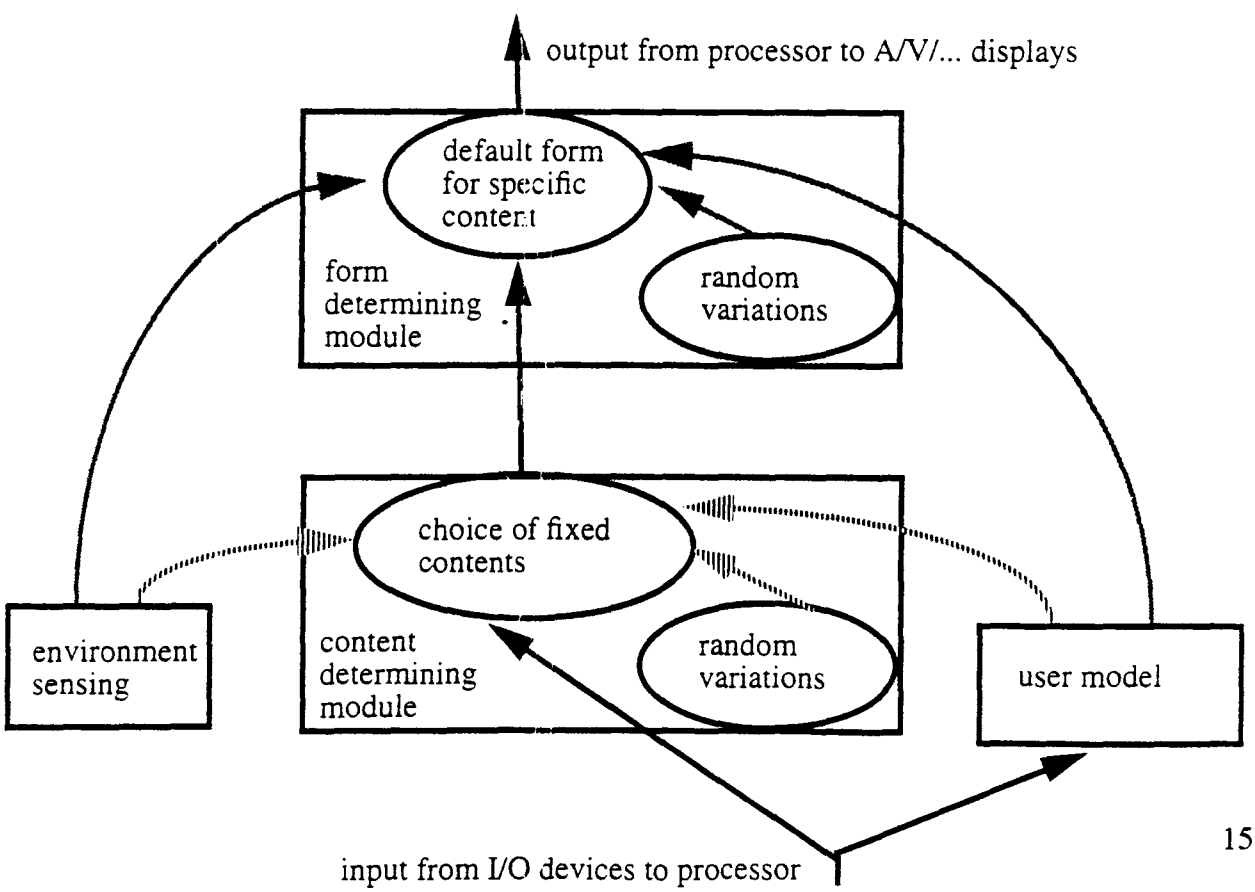
Appendices

Appendix A: Content and form

To be incorporated into the model:
The relation between contents & form, and the way they influence each other.

We made the following step in our reasoning:

- Content & form are determined by the system (application). Therefore, any elaboration of the model should take place on the system-side of the model.
- The contents can be classified in one of 3 major categories:
  - feedback (direct reaction to user actions)
  - guidance (initiated by (lack of) user actions, but looking forward in time)
  - system actions (initiated by the system, not as a consequence of any user actions: here input from the environment is important).
- Also the task is of influence: not only is it reflected (as in the model) in the user-interface area, but also it is present in the application running on the processor, at least to the extent that the application designer has tried to accomodate anticipated user tasks into the application. This can take the form of an implicit or explicit user model that will determine both content and form of what is produced by the application.
- Thus we propose to elaborate the 'processing' section as follows:



Appendix B: Sensory modalities & processing activities

Alterations to the model with respect to sensory modalities and processing

1. Display -> sensors

Parallel or sequential use of different modalities

In order to convey information it is possible to use more than one modality at the same time to carry over the same information, but to convey different information it is probably more efficent to use a sequence of different modalities, to avoid confusion or unnecessary fatigue for the user. However, it may be possible that a learning process is involved, and that people acquire the skill to process different information conveyed by different channels at the same time.

Coherence

Using one modality at the background suggests coherence, even if the foreground modality changes in time. This is used in movies: when the scene changes, the music stays the same and suggests coherence.

Expectations

Users have expectations about the modalities they will have to use in a certain application. They will probably have to get used to unexpected modalities.

2. Sensors -> Responders

There should be a direct feedback loop between sensors and responders. E.g. when you touch a key there is a direct muscular feedback.

3. Perceptual (processor)

Attention

Attention should be added here, it is a prerequisite for the intellectual (cognitive) processor. It should be directed at the modality that carries out the main task, e.g. in a flight simulator, the eyes should be on the track, but touch should be used for operating the gear; in medical equipment audio is used to draw the attention when something is wrong.

4. Task interface

There are different classes of tasks, so perhaps the task block should be subdivided in parts, and appropriate modalities should be attached to different tasks. There is probably a hierarchy of tasks.