multimedia course(s)

A multimedia course can take a variety of forms, dependent on the level of the students. An outline of a possible course that might be given with the book is presented below.

theoretical part

The theoretical part focuses around the themes of

theory

- digital convergence
- broadband communication
- multimedia information retrieval

These themes allow for paying attention to a variety of subjects, popular trends in digital entertainment, but also standards in development such as MPEG-4, compression and multimedia information retrieval. (For more background on why I developed this material, look at the book proposal.)

In addition to the material in the book, there are additional lectures, that present material from a variety of sources. One exampleof such a lecture is a close reading of the MPEG-4 standard proposal, that has been modified for presentation using the slides format.

An example schedule is presented below.

schedule

1. introduction, practical assignment
2. multimedia authoring – flash/flex
3. virtual environments
4. digital convergence, information spaces
5. codecs, MPEG-4, standards
6. information retrieval
7. questions
8. examination

Additional material is referred to in the various *research directions* sections, as well as the resources that are given for each chapter.

presentation(s)

Powerpoint presentations are available for a selected number of lectures. In addition, each section may be presented as described in the readme using dynamic HTML or VRML slides.

practical assignment

As a practical assignment, the students have to develop an *Annotated Tour of Amsterdam*, as described (in skeleton form) below.

Concluding the practical assignment is there will be a so-called Roundup, a public meeting where the most beautiful/funny/extreme presentations are shown.

assignment

title
Annotated Tour in Amsterdam

duration
1-2 minutes

effort
1 or 2 week(s)

format
flash

description Take a map of Amsterdam and select a particular route. Make a presentation that offers information about a number of locations on that route. The choice of locations is free. The information given must be relevant from some chose perspective. For example, looking at the buldings on the route you may take an historic perspective (and skectch the development in time) or an architectural perspective (and analyse and compare various styles of building). Alternatively, you may take a cultural perspective, and show fragments of the ife and working of living or dead artists. The presentation must be entertaining, not to say compelling. The user/viewer must be ableto enjoy the presentation without being obliged to make any choice or giving directives.

phases

1. determination of the concept – that is route, perspective and (global) contents
2. detailed scenario – choice of images and other material, description of scenes and transitions
3. technical realisation – elaboration of scenes and (visual) effects
4. finalproduction – finishing touch and conversion to shockwave format
5. justification – a brief description of the presentation, and an explanation of what 'meta-information'is needed to make you presentation accessible for search

deliverables The groeps must maintain a web-site where all the deliverables of the project are available for inspection:

1. week 1: determination of concept – 1 'page'
2. week 3: detailed scenarion – max 10 pages, with timeline, schetches, photo material, and a brief description per scene
3. week 6: technical realisation – keep a record of the work done
4. week 8: final production – movie in flash format
5. week 8: justification – one or two pages

procedure The deliverables must be available in the web directory of your account. Take care to make the site attractive and sufficiently informative.

deadlines

See your online information

guidance For each phase there is a deliverable. The deliverables must be approved before you may continue with the next phase.

remarks Learning flex and/or flash takes time. In week 1, 2 and 3, when you work on the concept and scenario, you must get familiar with your tool of choice and do experiments by realizing fragments of your presentation and exploring the features of your tool.

1 – getting to know eachother

The introductory meeting is meant to give an overview of the course. It introduces the *themes and variations* of the course and explains the assignment.

contents

- overview
- assignment

The overview is included to allow for immediate presentation, using the *talk* or*show* icon on the top right of the page.

2 – authoring and stuff

Since the students must get on their way with their practical assignment quickly, this lecture is primarily devoted to a discussion of what authoring involves (based on the *collected wisdom* of section 2.3), a brief introduction to writing a scenario or presentation script, and a general introduction to flex or flash.

topics

- *collected wisdom*
- multimedia for the web
- a short intro to flex/flash

You may skip the part on flex/flash and replace it by material concerning your authoring tool(s) of choice.

3 – virtual environments

Although this material forms the last chapter of part iii of the book, I thought it worthwhile to start with the *conclusion*, so to speak, and present *virtual environments* as the ultimate interface to multimedia information systems.

virtual environments

- 8.1 virtual context

- 8.2 navigation by query
- 8.3 intelligent agents

The material includes online demos of the prototypes discussed. See also the resources.

4 – digital convergence

Digital convergence is one of the central themes of the course. This theme is also related to, for example, the MPEG-4 standard that is treated later. Also, the notion of *information spaces* is introduced here, which is needed for a better understanding of the information retrieval issues that will be dealt with later.

digital convergence

- 1.2 entertainment
- 1.2 digital convergence
- 2.1 information spaces
- 2.2 hypermedia
- 1.3 commercial impact

It is not likely that you can treat all the material that is listed here in one session. You'll have to make a choice. You may, for example, leave out hypermedia.

5 – codecs and standards

Compression and decompression need to be understood by the student of multimedia. Perhaps not in all detail, but in any case to the extent that the consequences of the choice for a particular media type may be evaluated with respect to (network) resource demands.

The MPEG-4 standard is most directly related to the issue of compression and decompression. The other standards, nevertheless, are sufficiently related with MPEG-4 to justify a combined treatment.

Before looking at codecs and standards, we briefly discuss the current state of multimedia on the web.

codecs and standards

- (multimedia) semantic web?
- codecs
- standards (MPEG-4, SMIL, RM3D)
- game engine(s)

The MPEG-4 standard proposal is an adapted version of the original proposal, that is suited for presentation using either one of the slide formats.

6 – information retrieval

Based on the *Amsterdam Drugport* scenario we can identify the need for information retrieval capabilities for images, text documents, audio and video. Before we delve intoany of the media types, we discuss an information system architecture that allows for a uniform approach to a variety of media types. After discussing information retrieval and content annotation for the various media types, we elaborate on the uniform approach by defining the notion of *media abstraction*.

information retrieval

- multimedia scenarios
- information system architecture
- images
- documents
- audio, video
- media abstractions

You may leave out the audio media type, since the material discussed there is not immediately relavant to the scenario sketched. The section on feature extraction might be discussed seperately or in combination with the audio 'research directions'.

7 – questions

Basically, preparation for the exam.

see checklist

QUESTIONS?

If no questions are asked, take some selected items from the exam, and show how you require these to be answered.

Exam Questions Introduction Multimedia

1. (*) Give a short description of the contents and structure of your presentation. Indicate how the information contained in your presentation can be made accessible (for example in search).

(1)

2. (*) Sketch the developments in *multimedia*. What do you expect to be the commercial impact of multimedia in the (near) future?
3. Explain what is meant by *digital convergence*.
4. Which kinds of *(digital) convergence* do we have?
5. Discuss the relation between the *medium* and the *message.*.
6. Give a brief sketch of the development of *digital entertainment*.
7. Characterize: HDTV, SDTV, ITV.
8. Discuss convergence with respect to *platforms*.
9. Discuss convergence with respect to *delivery*.

(2)

10. (*) What factors play a role in the development of *multimedia information systems*? What research issues are there? When do you expect the major problems to be solved?
11. Define the notion of *information spaces*?
12. Indicate how multimedia objects may be placed (an queried for) in an *information (hyper) space*?
13. Characterize the notion of *hypermedia*.
14. Discuss which developments make a large scale application of multimedia information systems possible.
15. Give a characterization of an object, a query and a clue in an *information space*.
16. Describe the *Dexter Hypertext Reference Model*.
17. Give a description of the *Amsterdam Hypermedia Model*.

(3)

18. (*) What role do standards play in *multimedia*? Why are standards necessary for compression and delivery. Discuss the MPEG-4 standard and indicate how it is related to other (possible) standards.
19. What is a *codec*?
20. Give a brief overview of current multimedia standards.
21. What criteria must a *(multimedia) semantic web* satisfy?
22. What is the *data rate* for respectively (*compressed*) voice, audio and video?
23. Explain how a *codec* functions.
24. Which considerations can you mention for choosing a compression method?

25. Give a brief description of: XML, MPEG-4, SMIL, RM3D

(5)

26. (*) What is meant by the *complementarity of authoring and retrieval*? Sketch a possible scenario of (multimedia) information retrieval and indicate how this may be implemented. Discuss the issues that arise in accessing multimedia information and how content annotation may be deployed.
27. How would you approach *content-based description* of *images*?
28. What is the difference between a *metric* approach and the *transformational* approach to establishing similarity between images?
29. What problems may occur when searching in text or document databases?
30. Give a definition of: *shape descriptor* and *property descriptor*. Give an example of each.
31. How would you define *edit distance*?
32. Characterize the notions *precision* and *recall*.
33. Give an example (with explanation) of a *frequency table*.

(6)

34. (*) How can video information be made accessible? Discuss the requirements for supporting video queries.
35. What are the ingredients of an *audio data model*
36. What information must be stored to enable search for video content?
37. What is *feature extraction*? Indicate how feature extraction can be deployed for arbitrary media formats.
38. What are the parameters for *signal-based (audio) content*?
39. Give an example of the representation of *frame-dependent* en *frame-independent* properties of a video fragment.
40. What are the elements of a query language for searching in video libraries.
41. Give an example (with explanation) of the use of *VideoSQL*.

(7)

42. (*) What are the issues in designing a *(multimedia) information system architecture*. Discuss the tradeoffs involved.
43. What considerations would you have when designing an architecture for a multimedia information system.
44. Characterize the notion of *media abstraction*.
45. What are the issues in *networked multimedia*.
46. Describe (the structure of) a video database, using *media abstractions*.
47. Give a definition of the notion of a *structured multimedia database*.
48. Give an example (with explanation) of querying a *hybrid multimedia database*.
49. Define (and explain) the notion of *virtual objects* in *networked multimedia*.

(8)

50. (*) Discuss how *virtual environments* may be used for giving access to *(multimedia) information*. Give a brief characterization of *virtual environments*, and indicate how *information (hyper) spaces* may be projected in a virtual environment.

51. What is meant by *virtual context*?
52. Give an example of *navigation by query*, and indicate its possible advantages.
53. Discuss the deployment of *(intelligente) navigation agents*.
54. Give a brief characterization of: VRML.
55. What is a *viewpoint transformatie*?
56. What kinds of navigation can you think of?
57. How may intelligent avatars be realized? Give an example.

SELECTED SECTIONS FROM TOPICAL MEDIA & GAME DEVELOPMENT

section(s)

A. Eliëns (10/9/07)

1.1 entertainment and experience

The question of *what is multimedia* is rather elusive. We may, nevertheless, look at how the phrase *multimedia* is used, and how the concept *multimedia* is related to other concepts. as in the concept graphs that may be obtained with the Visual Thesaurus[1], providing as input *multimedia*.

[1] www.visualthesaurus.com

We then see that the notion of multimedia is related to *systems*, in particular interactive and hypermedia systems, and indirectly also to the notion of *transmission*, which will even become more apparent when we inspect the graph for the concept of *medium*, depicted in figure X below.

However, although this gives us some indication of how to position *multimedia* in the larger area of computer applications, in particular when exploring the *systems* node, it does not so much tell us what multimedia is all about.

From the perspective of human cognition, we may look at how multimedia contributes to our understanding of ourselves and the world around us. Traditionally, three levels of cognitive functioning are distinguished, Education, corresponding with three levels of meaning:

levels of meaning

- actionary level – action and movements
- sensory/iconic level – images and impressions
- symbolic level – language and mathematics

Multimedia is clearly (most strongly) related to the sensory/iconic level, although for games one could say there is also a strong relation with the actionary level, and to some extent (for both multimedia and games) with the symbolic level.

For a more serious and deep understanding of how multimedia artefacts provide meaning and what role they play in our daily life, or how that meaning is affected by social contexts, we need to take recourse to *semiotic theory*, which is now one step too far, buth which we will look at in chapter 12.

Antother perspective from which to understand the meaning of *multimedia*, is to look at the function of media in our society, or, in other words, how *multimedia* is situated in our cultural institutions.

Consider this quote from the preface of all of all MIT books in the *Leonardo* series:

cultural convergence

> *The cultural convergence of art, science, and technology provides ample opportunity for artists to challenge the very notion of how art is produced and to call into question its subject matter and its function in society.*

Although the quote is about *art*, it is essentially related to *multimedia*, to the extent that the quote refers to *media art*. The MIT Media Lab[2] is one of the worlds most famous institutes in the field of multimedia. The *Leonardo* series is a collection of authoritive books on multimedia and related topics, which includes DeepTime, VirtualArt, InformationArts.

To understand the position of (computer supported) media in our society, we may observe following DeepTime: there are two forces, political and technological, and there is, currently, a trend towards standardization and uniformity

standardization and uniformity

1. Telematic media were incorporated very quickly in the globalization strategies of transnational corporations and their political administrators and they became increasingly dependent on existing power structures.
2. At the other end of the scale, there were individuals, or comparatively small groups, who projected great hopes onto these networks as a testing ground for cultural, artistic and political models that would give greater prominence and weight to divergence and plurality.

[2] www.media.mit.edu/

This reflects what DeepTime calls the *advanced media paradox*, facilitating heterogeneity and immersion on the one hand, and striving for universalisation on the other hand, as demanded by the centers of technological and political power.

Leaving the socio-political arena, we may in some sense predict the tension between *convergence* and *divergence*, by looking at the meaning context of the concept of *convergence*, again using the Visual Thesaurus, where we find that not only notions such as *overlap* and *occurrence* are related to it, but also the complementary concept of *divergence*. However, instead of speculating on the meaning of words, it might be more worthwhile to look at what we may consider to be the recent history of multimedia, entertainment.

entertainment

In november 2000, a theme issue of the Scientific American appeared, featuring a number of articles discussing (digital) entertainment in the era of digital convergence. Let's start with a quote:

Scientific American (november 2000)

> *The barriers between TV, movies, music, videogames and the Internet are crumbling. Audiences are fetting new creative options. Here is what entertainment could become if the technological and legal hurdles can be cleared ...*

Moreover, the editors made some wildly speculative claims, such as *digitizing everything audio and video will disrupt the entertainment industry's social order*, and *the whole concept of holding a CD or movie in your hand will disappear once d-entertainment is widely available.* To some extent this seems already to be true, as for example the music industry can painfully testify to.

Underlying the importance of entertainment in the era of digital convergence is the premisse governing an entertainment economy, which may be stated as

there is no business without show business

Additionally, the authors of the introduction to the theme issue speculate that *the creation of content will be democratized*, due to the availability of low cost digital movie cameras and PC video editors. Producing a video movie is now possible for just a few thousend euro or dollars. However, given the aesthetic ignorance of the average individual making video movies, it seems doubtful that this will hold true for entertainment in general.

In that same issue of the Scientific American, Gloria Davenport, a pioneer in the field of multimedia, presents list of applications characterizing the evolution of digital entertainment, Entertainment:

evolution of digital entertainment

- 1953: Winky Dink (CBS) – interactive television, drawing exercise
- 1972: Pong (Atari) – ping-pong on computer screen
- 1977: Adventure – text-based interactive fiction
- 1983: Dragon's Liar – laser-disc technology 3D game
- 1989: SimCity – interactive simulation game
- 1989: Back to the Future – the Ride
- 1993: Doom – 3D action game

- 1995: The Spot – interactive web-based soap opera (Webisodic)
- 1999: IMAX3D – back to Atlantis (Las Vegas)
- 2000: Big Brother – TV + around the clock Web watch + voting
- 2001: FE Sites – fun enhanced web sites

It is interesting to note that *Big Brother*, which was originally created by a Dutch team, has become a huge success in many countries. Although the integration with the web was limited, it may be seen as the start of a number of television programs with web-based interaction facilities.

digital experience

The list compiled by Gloria Davenport suggests, a convergence towards an 'ultimate digital experience', Now, what does *digital experience* mean?

In a special issue of the Communications of the ACM, about the next 1000 years of computing, Ramesh Jain makes the following observation, Experience:

> *The desire to share experiences will be the motivating factor in the development of exciting multimedia technology in the foreseeable future.*

Considering the variety of means we have at our disposal to communicate, as reflected in the list below, we may wonder whether our current technology really stands out as something special.

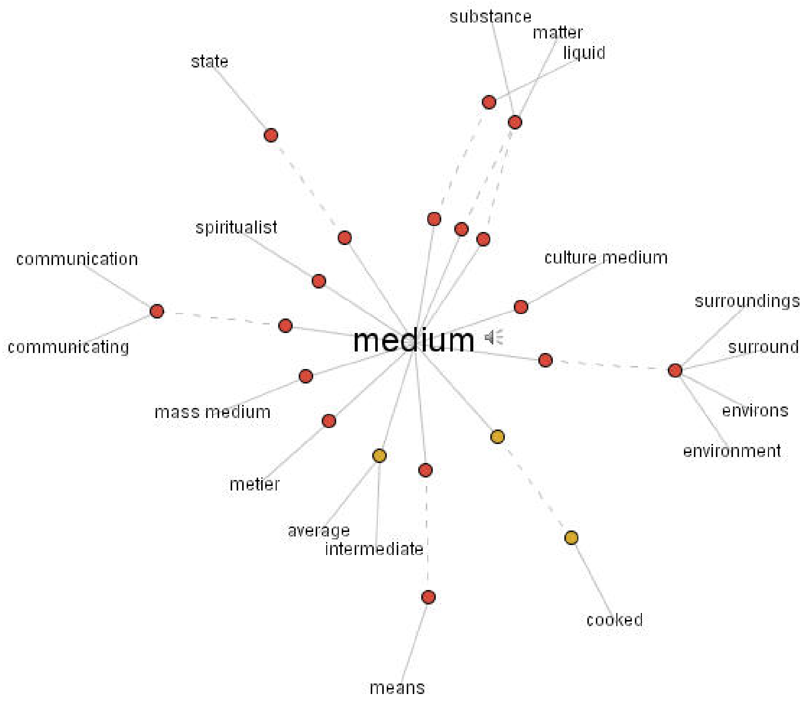
communication technology

- *oral* – communicate symbolic experiences
- *writing* – record symbolic experiences
- *paper* – portability
- *print* – mass distribution
- *telegraph* – remote narrow communication
- *telephone* – remote analog communication
- *radio* – analog broadcasting of sound
- *television* – analog A/V broadcasting
- *recording media* – analog recording
- *digital processing* – machine enhancement
- *internet* – multimedia communication

According to Ramesh Jam, internet-based multimedia communication differs from earlier communication technology in that it somehow frees the message from the medium. Reflecting on Marshall McLuhan phrase – *the medium is the message* – he observes that:

the medium was the message when only one medium could be used to communicate messages.

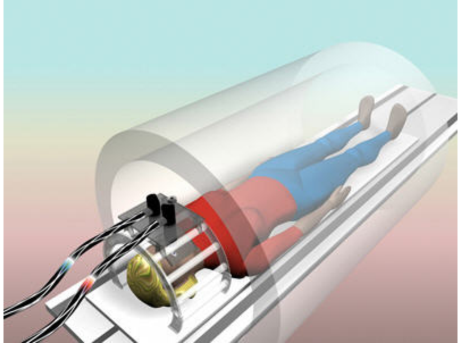
Now, that the Internet allows the synthesis and rendering of information and experiences using whatever is the most appropriate media to convey the message, the message is, as Jain phrases it, just the message, and the medium is just the medium. In other words, the medium itself does not seem to constrain what message can be conveyed. Looking at the documentary *Fahrenheit 9/11* though, we may seriously doubt whether this is true. Although it is possible to gain knowledge about the alliances that underly politics, even in the age of the internet, the television campaigns seem to be more dominant in affecting the general publics opionion about global politics than anything else, due to the conventional formats of presentation and editing.



13

Let's once more look at a graph, above, indicating the concept relations for the notion of *medium*. What strikes me as important are the relations with the distinct concepts of *substance*, *communication*, *environment*, and *intermediate*. In some respects the notion of *medium*, underlying the plural use of it in *multimedia* is comparable to the notion of *ether*, which was once seen as a vehicle for the transport of broadcasted information. But I also like to stress the 'substantial' aspect of multimedia, as a material for design and creation, similar to paint.

The basic issue here is what is a medium and how does it affect, or even shape our experience(s). Following Ramesh Jain, we may speculate that the range of sensory information offered by multimedia applications may become much richer than is currently the case, and we may then predict that there will be a tremendous progress in presentation technology, multisensory presentation technology! Clearly, from a technological perspective there seems to be no limit, except those imposed by our own phantasy. However, it should be equally obvious that compelling experiences rely on carefully staged presentations, and as such require an entirely new discipline of design.



VR for pain relief

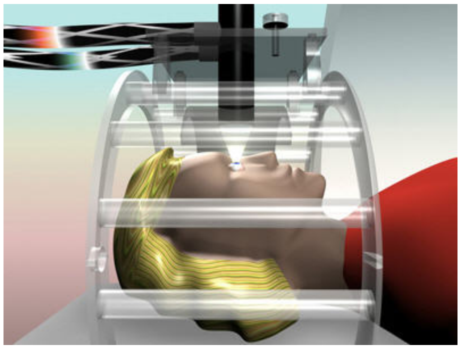


image delivery system

14

example(s) – *VR for pain relief*

The research project fMRI Research on Virtual Reality Analgesia[3] at the Human Interaction Laboratory (Washington) has explored the use of VR to reduce the agonay of taking MRI scans. The U.W Radiology Digital Imaging Science Centers wide field of view magnet-friendly virtual reality image delivery system makes it possible for volunteers and patients to have the illusion of going into virtual reality during fMRI brain scans. As explained on the website, the image on the left above, shows a woman in virtual reality during an fMRI brain scan, looking into a custom magnet-friendly virtual reality goggles. VR images from projectors in another room are carried to the participant in the form of light (photons, no electrons) via optic fiber image guides. The participant has the illusion of going inside the virtual world, allowing researchers to measure what happens to her brain when she reports reductions in pain during VR. The white cage-like structure around the womans head, in the image on the right, shows fMRI receiver coils used by the fMRI brain scanner to collect the information about changing patterns of brain activity.

Another project investigating the use of VR techniques for pain distraction can be found at the site of the Virtual Environments[4] of the Georgia Institute of Technology, Atlanta.

research directions– *the face of cyberspace*

The notion of *cyberspace* was introduced in William Gibson's novel *Neuromancer*, that appeared in the early 1980's, signifying a vast amount of (digital) data that could be accessed only through a virtual reality interface that was controlled by neuro-sensors. Accessing data in *cyberspace* was not altogether without danger, since data protection mechanisms (including firewalls, as we call them nowadays) were implemented using neuro-feedback. Although the vision expressed in *Neuromancer* is (in our days) still futuristic, we are confronted with a vast amount of information and we need powerful search engines and visualisation techniques not to get lost. So what is the reality of *cyberspace* today?

[3] www.hitl.washington.edu/research/magnet

[4] www.gvu.gatech.edu/virtual

... cyberspace is a construct in terms of an electronic system.

as observed by Vivian Sobschack, 1996, quoted from History, p. 321. On reflection, our (electronic) world of today might be more horrendous than the world depicted in *Neuromancer*. In effect,

cyberspace

> *television, video cassettes, video tape-recorder/players, video games, and personal computers all form an encompassing electronic system whose various forms interface to constitute an alternative and absolute world that uniquely incorporates the spectator/user in a spatially decentered, weakly temporalized and quasi-disembodied state.*

All these gadgets make us dizzy, stoned with information and fried by electro-magnetic radiation. However, the reality of everyday computer use is (fortunately?) less exciting than the images in *Neuromancer* suggest. User interfaces are usually tiresome and not at all appealing. So except for the fanatic, the average user does easily get bored. Would this change when virtual reality techniques are applied pervasively? What is virtual reality?

virtual reality

> *virtual reality (is) when and where the computer disappears and you become the 'ghost in the machine' ...*

In other words, virtual reality is a technology that provokes immersion, sensuous immersion, supported by rich media and powerful 3D graphics. In our age of information, we may wonder how all that information should be presented. Rephrasing the question, we may ask what are the limits of the digital experience, or more importantly, what should be the norm: 3D virtual environments, plain text, or some form of XP?

1.2 technological developments

Let's see if we are able to give a more precise characterization of *digital convergence*. In their introduction to the theme issue of the Scientific American, Forman and SaintJohn locate the beginning of digital convergence, historically, at the 1939 New York World Fair, and more in particular the RCA Pavillion, which should be considered as the formal debut of television broadcast. They observe that

history

> *the receiver at the RCA Pavillon was way ahead of its time, it was a combination of television - radio - recorder - playback - facsimile - projector ...*

Moreover, they remark that this *in hindsight suggests that we humans have a fundamental desire to merge all media in one entity.*
By way of definition we may state, following Forman and SaintJohn, that digital convergence is:

digital convergence

> the union of audio, video and data communication into a single source, received on a single device, delivered by a single connection

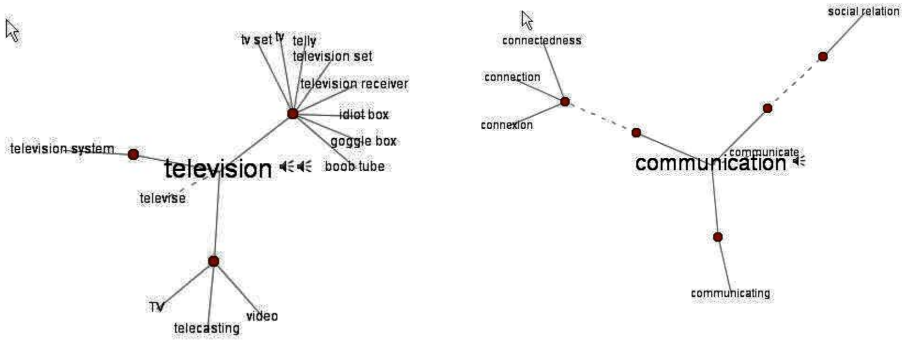
And, as they say, *predicted for decades, convergence is finally emerging, albeit in a haphazard fashion.*

Taking a somewhat closer look, we may discern subsidiary convergences with respect to content, platform and distribution:

subsidiary convergences

- *content* – audio, video, data
- *platform* – PC, TV, internet, game machine
- *distribution* – how it gets to your platform

Here, Forman and SaintJohn continue by speculating that if compatibility standards and data protection schemas can be worked out, all d-entertainment will converge into a single source *that can shine into your life on any screen, whereever you are ...* However, observe that the number of competing standards and architectures is enormous, and that apart from the technical issues involved it is not entirely clear what business model should underly such convergence. In computer shops, there PCs with TV receivers are sold in the range of 1000-2000 euro. This does not include the screen. They come with either the XP Home or Windows Media Center. One of the first in this line of machines, in the higher prices range, was the Sony W1.



15

TV or PC

It is fair to say that no device has changed the way we live so dramatically as television. Television, for one, has altered the way we furnish our living rooms, not to speak about the time we waste watching the thing. Comparing the graphs for *television* and *communication*, we immediately see that their underlying concepts are very different. And more specifically, the association of television with a phrase such as idiot box may raise doubt whether the promise of convergence, which does include communication as an essential feature, will ever become a reality.

Now, we may wonder what interactive television and enhanced televison have to offer us. Looking back, we may observe that it takes some time for the new possibilities to catch on. For example, interactive television was introduced in 1970, but apparently people did not want to communicate with the broadcaster. As another example of enhanced televison, take Big Brother. Although many people watched Big Brother when it first appeared on television, the willingness of the audience to react other than by phone was (apparently) somewhat disappointing. Perhaps, in the Netherlands this was due to the fact that only a fraction of the PC owners was, at that time, permanently online.

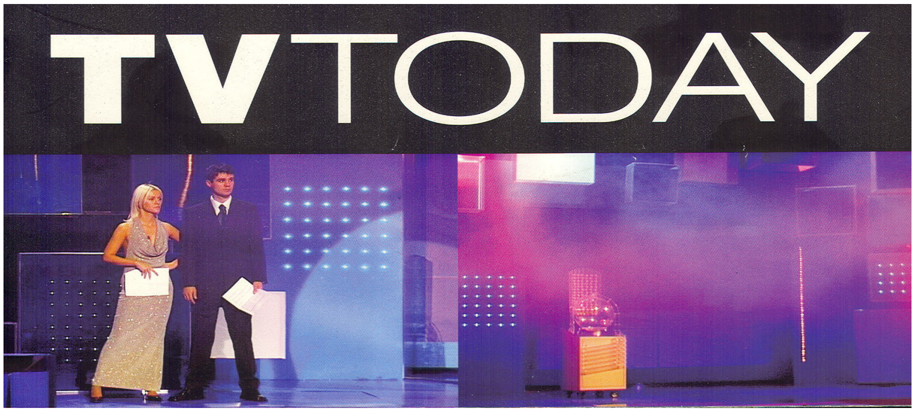
In spite of the failed experiments, Forman and SaintJohn state, somewhat optimistically, that *the convergence of digital content, broadcast distribution and display platforms create the* big *convergence of d-entertainment and information with feedback supporting human interactivity.*

Before looking at *digital television* more closely, let's summarize what digital convergence involves:

convergence

- *content* – 2D/3D graphics, data, video, audio
- *distribution* – broadcast, wireless, DVD, internet, satelite, cable
- *platform* – PC, television, game machine, wireless data pad, mobile phone

This summary indicates the technical opportunities, and the possible functinal extensions that may enhance our use of television, computer, game console and mobile phone. As concerns digital television, we may come up with some immediate advantages, such as enhanced resolution, a multiplication of channels, and (more relevant to the issue of convergence) interactive television.



exposition on the history of TV in Institute for Time-based Arts/Montevideo[5]

16

To get you failiar with some common acronyms, when speaking about (digital) television, we must make a further distinction between:

- HDTV – high definition television
- SDTV – standard definition television
- ITV – interactive television

As further discussed in chapter 3, we have (standard) codecs for d-TV, in particular MPEG-2, for recording digital video, and MPEG-4, for high-quality streaming video on the internet, both from the Motion Picture Expert Group, that enable the effective delivery of digital video, possibly in combination with other content.

Unfortunately, experts disagree on what might become the most suitable appliance or platform to consume all those digital goodies. Here are some possible choices:

a *killer* d-TV appliance ...

- personal television – TiVo, Replay-TV (MPEG-2 cache)
- game machine – Sony PS 2/3, X-Box

Will we prefer to watch stored video, instead of live televison broadcasts? Will the Internet be able to compete with traditional television broadcasting. Will DelayTV or Replay-TV, which allows you to watch previous broadcasts at a time that suits you become popular? Will an extended game machine or PC replace your television? Currently,we must observe that streaming media (still) have rather poor resolution.

Leaving game machines aside, will it then be the TV or PC that will become our platform of choice? Forman and SaintJohn observe:

TV or PC

> *The roadblock to the Entertainment PC could be the PC itself. Even a cheap TV doesn't crash or freeze. The best computers still do.*

However, they conclude that it might make sense to adopt a programmable PC that can support competing TV standards, rather than construct a stack of TV peripherals. Nevertheless, there are a number of problems that occur when we (collectively) choose for the PC as our platform for d-entertainment. Should we have thin clients, for example based on the Sun/Java platform or so-called fat clients based on some version of Microsoft windows> How do we handle the fact that the current internet protocols are not robust, and how can we provide what is known as *quality of service*? Should we adopt any of the proprietary architectures and codecs, such as RealVideo, QuickTime, Windows media, or should we adhere to an open standard such as MPEG-4?

Evidently, the situation becomes even more complex when we just consider the range of alternatives for connectivity, that is for possible ways of distributing contents:

distribution

- *telephone network* – from 0.5 - 2 Mbps to 60 Mpbs (2.5km)
- *broadcast TV* – 6 MHz / 19 Mbps (4 channels MPEG HDTV)
- *cable TV* – hybrid fiber-optic coaxial cable 6 Mbps
- *fixed wireless* – 2 Mbps (radiotowers + rooftop antenna), phones/handhelds
- *satellite* – downloads to 100kbps, modem for uploads ...

Most probably, convergence with respect to distribution will not result in one single way of being connected, but rather a range of options from which one will be selected transparently, dependent on content and availability.

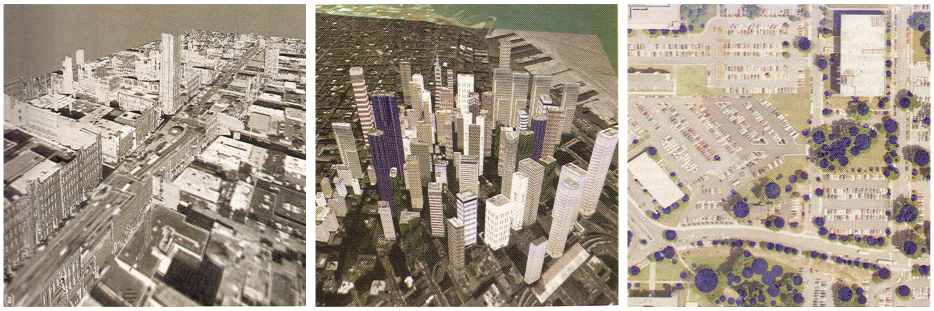
Let's stay optimistic, and ask ourselves the following question:

what will we do with convergence once we have it?

One possible scenario, not too unlikely after all, is to deploy it for installing computing devices everywhere, to allow for, to name a few, smart houses, smart clothes, or, in other words, to create a smart world. I wonder what a smart world will look like. In the end we will have to wait and see, but whatever will emerge

we will watch

That is to say, it is not likely that we will have a world without television. Television as we are used to it seems to be the dominant paradigm for d-entertainment, for both the near and distant future.



Berkeley mesh San Francisco view augmented terrain map

17

example(s) – *visible world*

Just imagine that every visible place on earth would be accessible in a virtual world. Researchers of the Georgia Institute of Technology[6], Atlanta, have developed software for the semi-automated construction of detailed interactive urban environments, that takes data from multiple sources, including geo-corrected imagery from aerial photography and satelites and ground-based close-ups, World.

The challenge here is to collect data from multiple sources and convert this into models, and perhaps even more difficult, to make the models visible so that they can be navigated in an interactive fashion. Recently, the Georgia group teamed up with a group from Berkeley to develop more complex models (images on the left), and together they are working on automating the extraction of information from aerial pictures (image on the right), in particular the detection of groups of trees, and height estimation.

There are many applications for such technology, including urban planning, emergency response, tourism and entertainment, military operations, traffic management, construction and maintenance, mobile services, citizen-government relations, and (not in the least) games.

The next step might be to connect the cameras, that are already there in many of these places, to the model, to observe what happens there in real life. But, somehow, this vision becomes frightening.

However, if you want to give it a try yourself, and populate the virtual globe with your own creations, go download the viewer and editing tool from *Google Earth*:

Google Earth

- Earth – earth.google.com
- SketchUp – sketchup.google.com/download.html

and read the tutorials!

research directions– *technological determinism*

Although there are many technical issues involved in (digital) multimedia, as exemplified in the issues that play a role in digital convergence, a technical perspective alone does not suffice. Each technological innovation has its consequences on our social life. Conversely,

[6] www.gvu.gatech.edu/datavis/research

each trend in society might result in the adoption or development of new technology. Looking at the history of the media, we may observe that media become *materials* in our social processes. Or, as phrased in History:

media as materials

> *each medium of communication tended to create a dangerous monopoly of knowledge*

For example (History, p. 8) for Christians, images where both a means of conveying information and a means of persuasion, that is part of the rethorics of institutionalized religion.

Looking at our age, and the media that have come into existence in the previous century (radio, television, ...), History observe that:

technological determinism

> *technological determinism was not the answer, ... more attempts were to be made to provide answers about the social consequences of television than had ever been asked about radio.*

In effect, underlying all developments in the media (including the computer) we may assume a basic need for information. A rather problematic need, for that matter:

information

> *Information became a major concern anywhere during the late 1960 and 1970s where there was simultaneous talk both of 'lack of information' and 'information saturation'.* History, p. 555

Nowadays, we regard information as a commodity. Train schedules, movies, roadmaps, touristic information, stock prices, we expect it all to be there, preferably online, at no cost. No information, no life. Information drives the economy. Upwards and downwards!

3.1 codecs

Back to the everyday reality of the technology that surrounds us. What can we expect to become of networked multimedia? Let one thing be clear

compression is the key to effective delivery

There can be no misunderstanding about this, although you may wonder why you need to bother with compression (and decompression). The answer is simple. You need to be aware of the size of what you put on the web and the demands that imposes on the network. Consider the table, taken from Codecs, below.

media	uncompressed	compressed
voice 8k samples/sec, 8 bits/sample	64 kbps	2-4 kbps
slow motion video 10fps 176x120 8 bits	5.07 Mbps	8-16 kbps
audio conference 8k samples/sec 8bits	64 kbps	16-64 kbps
video conference 15 fps 352x240 8bits	30.4 Mbps	64-768 kbps
audio (stereo) 44.1 k samples/s 16 bits	1.5 Mbps	128k-1.5Mbps
video 15 fps 352x240 15 fps 8 bits	30.4 Mbps	384 kbps
video (CDROM) 30 fps 352x240 8 bits	60.8 Mbps	1.5-4 Mbps
video (broadcast) 30 fps 720x480 8 bits	248.8 Mbps	3-8 Mbps
HDTV 59.9 fps 1280x720 8 bits	1.3 Gbps	20 Mbps

You'll see that, taking the various types of connection in mind

(phone: 56 Kb/s, ISDN: 64-128 Kb/s, cable: 0.5-1 Mb/s, DSL: 0.5-2 Mb/s)

you must be careful to select a media type that is suitable for your target audience. And then again, choosing the right compression scheme might make the difference between being able to deliver or not being able to do so. Fortunately,

images, video and audio are amenable to compression

Why this is so is explained in Codecs. Compression is feasible because of, on the one hand, the statistical redundancy in the signal, and the irrelevance of particular information from a perceptual perspective on the other hand. Redundancy comes about by both spatial correlation, between neighboring pixels, and temporal correlation, between successive frames.

The actual process of encoding and decoding may be depicted as follows:

codec = (en)coder + decoder

signal → source coder → channel coder (encoding)

signal ← source decoder ← channel decoder (decoding)

Of course, the coded signal must be transmitted accross some channel, but this is outside the scope of the coding and decoding issue. With this diagram in mind we can specify the *codec design problem*:

> *From a systems design viewpoint, one can restate the codec design problem as a bit rate minimization problem, meeting (among others) constraints concerning: specified levels of signal quality, implementation complexity, and communication delay (start coding – end decoding).*

18

compression methods

As explained in Codecs, there is a large variety of compression (and corresponding decompression) methods, including model-based methods, as for example the object-based MPEG-4 method that will be discussed later, and waveform-based methods, for which we generally make a distinction between lossless and lossy methods. Hufmann coding is an example of a lossless method, and methods based on Fourier transforms are generally lossy. Lossy means that actual data is lost, so that after decompression there may be a loss of (perceptual) quality.

Leaving a more detailed description of compression methods to the diligent students' own research, it should come as no surprise that when selecting a compression method, there are a number of tradeoffs, with respect to, for example, coding efficiency, the complexity of the coder and decoder, and the signal quality. In summary, the follwoing issues should be considered:

tradeoffs

- *resilience to transmission errors*
- *degradations in decoder output – lossless or lossy*
- *data representation – browsing & inspection*
- *data modalities – audio & video.*
- *transcoding to other formats – interoperability*
- *coding efficiency – compression ratio*
- *coder complexity – processor and memory requirements*
- *signal quality – bit error probability, signal/noise ratio*

For example, when we select a particular coder-decoder scheme we must consider whether we can guarantee resilience to transmission errors and how these will affect the users' experience. And to what extent we are willing to accept degradations in decoder output, that is lossy output. Another issue in selecting a method of compression is whether the (compressed) data representation allows for browsing & inspection. And, for particular applications, such as conferencing, we should be worried about the interplay of data modalities,in particular, audio & video. With regard to the many existing codecs and the variety of platforms we may desire the possibility of transcoding to other formats to achieve, for example, exchange of media objects between tools, as is already common for image processing tools.

compression standards

Given the importance of codecs it should come as no surprise that much effort has been put in developing standards, such as JPEG for images and MPEG for audio and video.

Most of you have heard of MP3 (the audio format), and at least some of you should be familiar with MPEG-2 video encoding (which is used for DVDs).

Now, from a somewhat more abstract perspective, we can, again following Codecs, make a distinction between a *pixel-based approach* (coding the raw signal so to speak) and an *object-based approach*, that uses segmentation and a more advanced scheme of description.

- *pixel-based* – MPEG-1, MPEG-2, H3.20, H3.24
- *object-based* – MPEG-4

As will be explained in more detail when discussing the MPEG-4 standard in section 3.2, there are a number of advantages with an object-based approach. There is, however, also a price to pay. Usually (object) segmentation does not come for free, but requires additional effort in the phase of authoring and coding.

MPEG-1 To conclude this section on codecs, let's look in somewhat more detail at what is involved in coding and decoding a video signal according to the MPEG-1 standard.

MPEG-1 video compression uses both *intra-frame analysis*, for the compression of individual frames (which are like images), as well as. *inter-frame analysis*, to detect redundant blocks or invariants between frames.

The MPEG-1 encoded signal itself is a sequence of so-called I, P and B frames.

frames

- I: intra-frames – independent images
- P: computed from closest frame using DCT (or from P frame)
- B: computed from two closest P or I frames

Decoding takes place by first selecting I-frames, then P-frames, and finally B-frames. When an error occurs, a safeguard is provided by the I-frames, which stand on themselves.

Subsequent standards were developed to accomodate for more complex signals and greater functionality. These include MPEG-2, for higher pixel resolution and data rate, MPEG-3, to support HDTV, MPEG-4, to allow for object-based compression, and MPEG-7, which supports content description. We will elaborate on MPEG-4 in the next section, and briefly discuss MPEG-7 at the end of this chapter.

example(s) – *gigaport*

GigaPort[7] is a project focussing on the development and use of advanced and innovative Internet technology. The project, as can be read on the website, *focuses on research on next-generation networks and the implementation of a next-generation network for the research community.*

Topics for research include:

GigaPort

- optical network technologies - models for network architecture, optical network components and light path provisioning.
- high performance routing and switching - new routing technologies and transport protocols, with a focus on scalability and stability robustness when using data-intensive applications with a high bandwidth demand.
- management and monitoring - incident response in hybrid networks (IP and optical combined) and technologies for network performance monitoring, measuring and reporting.
- grids and access - models, interfaces and protocols for user access to network and grid facilities.
- test methodology - effective testing methods and designing tests for new technologies and network components.

As one of the contributions, internationally, the development of optical technology is claimed, in particular *lambda* networking, networking on a specific wavelength. Locally, the projects has contributed to the introduction of fibre-optic networks in some major cities in the Netherlands.

research directions– *digital video formats*

In the online version you will find a brief overview of *digital video technology*, written by Andy Tanenbaum, as well as some examples of videos of our university, encoded at various bitrates for different viewers.

What is the situation? For traditional television, there are three standards. The american (US) standard, NTSC, is adopted in North-America, South-America and Japan. The european standard, PAL, whuch seems to be technically superior, is adopted

[7] www.gigaport.nl/info/en/about/home.jsp

by the rest of the world, except France and the eastern-european countries, which have adopted the other european standard, SECAM. An overview of the technical properties of these standards, with permission taken from Tanenbaum's account, is given below.

system	**spatial resolution**	**frame rate**	**mbps**
NTSC	704 x 480	30	243 mbps
PAL/SECAM	720 x 576	25	249 mbps

Obviously real-time distribution of a more than 200 mbps signal is not possible, using the nowadays available internet connections. Even with compression on the fly, the signal would require 25 mbps, or 36 mbps with audio. Storing the signal on disk is hardly an alternative, considering that one hour would require 12 gigabytes.

When looking at the differences between streaming video (that is transmitted real-time) and storing video on disk, we may observe the following tradeoffs:

item	**streaming**	**downloaded**
bandwidth	equal to the display rate	may be arbitrarily small
disk storage	none	the entire file must be stored
startup delay	almost none	equal to the download time
resolution	depends on available bandwidth	depends on available disk storage

So, what are our options? Apart from the quite successful MPEG encodings, which have found their way in the DVD, there are a number of proprietary formats used for transmitting video over the internet: Quicktime, introduced by Apple, early 1990s, for local viewing; RealVideo, streaming video from RealNetworks; and Windows Media, a proprietary encoding scheme fromMicrosoft. Examples of these formats, encoded for various bitrates are available at Video at VU.

Apparently, there is some need for digital video on the internet, for example as propaganda for attracting students, for looking at news items at a time that suits you, and (now that digital video cameras become affordable) for sharing details of your family life.

Is digital video all there is? Certainly not! In the next section, we will deal with standards that allow for incorporating (streaming) digital video as an element in a compound multimedia presentation, possibly synchronized with other items, including synthetic graphics. Online, you will find some examples of digital video that are used as texture maps in 3D space. These examples are based on the technology presented in section **??**, and use the streaming video codec from Real Networks that is integrated as a rich media extension in the *blaxxun* Contact 3D VRML plugin.

comparison of codecs A review of codecs[8], including Envivio MPEG-4, QuickTime 6, RealNetworks 9 en Windows Media 9 was published januari 2005 by the European Broadcast Union[9]. It appeared that The Real Networks codecs came out best, closely followed by the Windows Media 9 result. Ckeck it out!

[8] www.ebu.ch/trev_301-samviq.pdf

[9] www.ebu.ch/trev_home.html

19

3.2 standards

Imagine what it would be like to live in a world without standards. You may get the experience when you travel around and find that there is a totally different socket for electricity in every place that you visit.

Now before we continue, you must realize that there are two types of standards: *de facto* market standards (enforced by sales politics) and committee standards (that are approved by some official organization). For the latter type of standards to become effective, they need consent of the majority of market players.

For multimedia on the web, we will discuss three standards and RM3D which was once proposed as a standard and is now only of historical significance.

standards

- XML – eXtensible Markup Language (SGML)
- MPEG-4 – coding audio-visual information
- SMIL – Synchronized Multimedia Integration Language
- RM3D – (Web3D) Rich Media 3D (extensions of X3D/VRML)

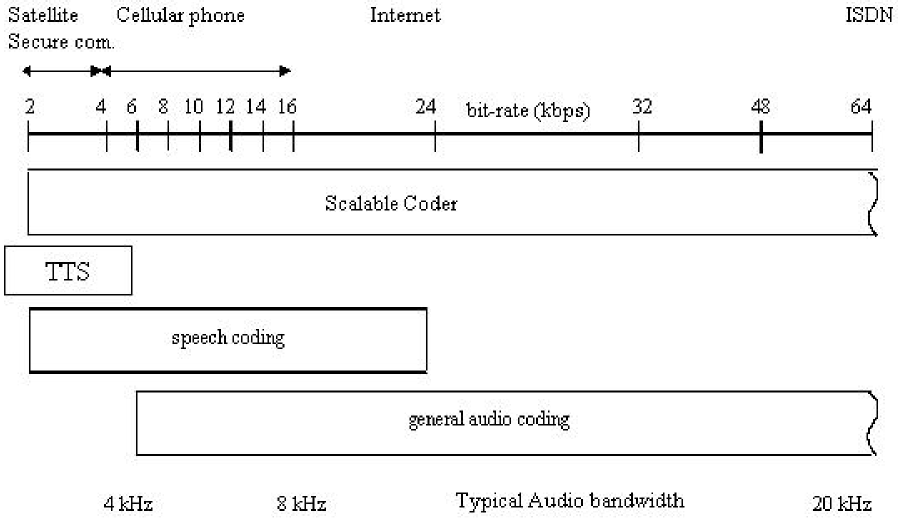
XML, the *eXtensible Markup Language*, is becoming widely accepted. It is being used to replace HTML, as well as a data exchange format for, for example, business-to-business transactions. XML is derived from SGML (Structured Generalized Markup Language) that has found many applications in document processing. As SGML, XML is a generic language, in that it allows for the specification of actual markup languages. Each of the other three standards mentioned allows for a syntactic encoding using XML.

MPEG-4 aims at providing "the standardized technological elements enabling the integration of production, distribution and content access paradigms of digital television, interactive graphics and multimedia", MPEG-4. A preliminary version of the standard has been approved in 1999. Extensions in specific domains are still in progress.

SMIL, the *Synchronized Multimedia Integration Language*, has been proposed by the W3C "to enable the authoring of TV-like multimedia presentations, on the Web". The SMIL language is an easy to learn HTML-like language. SMIL presentations can be composed of streaming audio, streaming video, images, text or any other media type, SMIL. SMIL-1 has become a W3C recommendation in 1998. SMIL-2 is at the moment of writing still in a draft stage.

RM3D, *Rich Media 3D*, is not a standard as MPEG-4 and SMIL, since it does currently not have any formal status. The RM3D working group arose out of the X3D working group, that addressed the encoding of VRML97 in XML. Since there were many disagreements on what should be the core of X3D and how extensions accomodating VRML97 and more should be dealt with, the RM3D working group was founded in 2000 to address the topics of extensibility and the integration with rich media, in particular video and digital television.

remarks Now, from this description it may seem as if these groups work in total isolation from eachother. Fortunately, that is not true. MPEG-4, which is the most encompassing of these standards, allows for an encoding both in SMIL and X3D. The X3D and RM3D working groups, moreover, have advised the MPEG-4 commitee on how to integrate 3D scene description and human avatar animation in MPEG-4. And finally, there have been rather intense discussions between the SMIL and RM3D working groups on the timing model needed to control animation and dynamic properties of media objects.



20

MPEG-4

The MPEG standards (in particular 1,2 and 3) have been a great success, as testified by the popularity of mp3 and DVD video.

Now, what can we expect from MPEG-4? Will MPEG-4 provide *multimedia for our time*, as claimed in Time. The author, Rob Koenen, is senior consultant at the dutch KPN telecom research lab, active member of the MPEG-4 working group and editor of the MPEG-4 standard document.

> *"Perhaps the most immediate need for MPEG-4 is defensive. It supplies tools with which to create uniform (and top-quality) audio and video encoders on the Internet, preempting what may become an unmanageable tangle of proprietary formats."*

Indeed, if we are looking for a general characterization it would be that MPEG-4 is primarily

MPEG-4

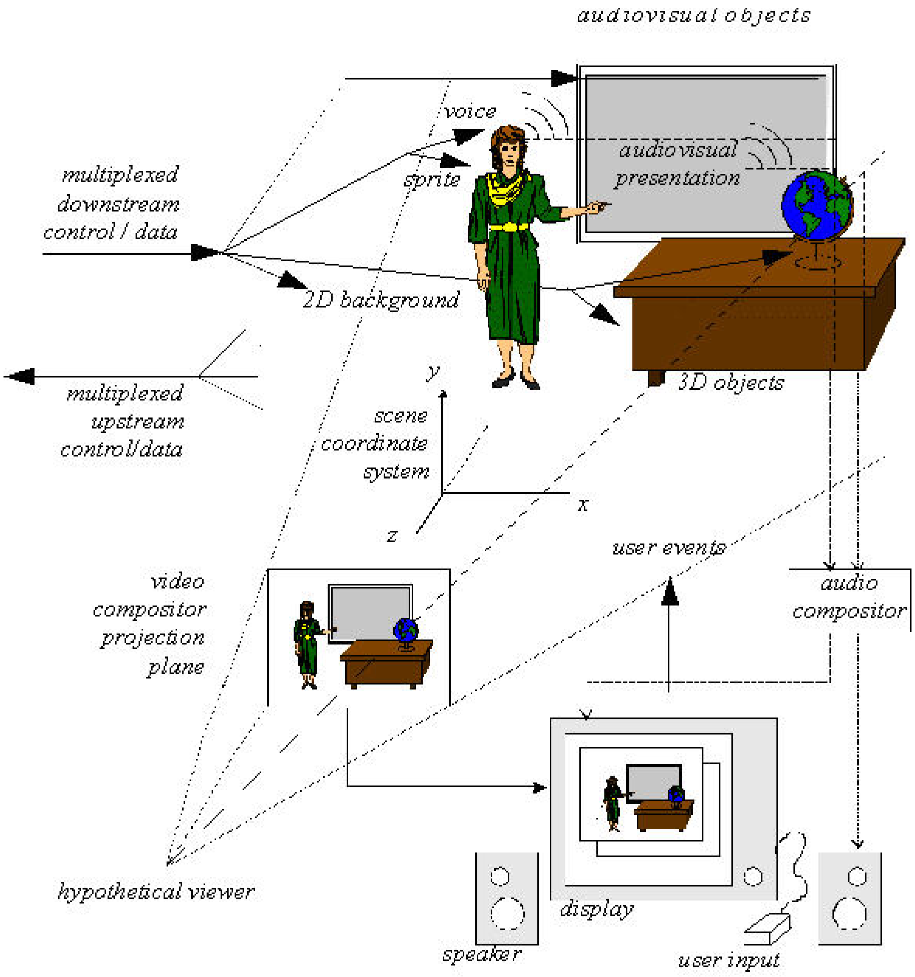
a toolbox of advanced compression algorithms for audiovisual information

and, moreover, one that is suitable for a variety of display devices and networks, including low bitrate mobile networks. MPEG-4 supports scalability on a variety of levels:

scalability

- *bitrate* – switching to lower bitrates
- *bandwidth* – dynamically discard data
- *encoder and decoder complexity* – signal quality

Dependent on network resources and platform capabilities, the 'right' level of signal quality can be determined by selecting the optimal codec, dynamically.



media objects It is fair to say that MPEG-4 is a rather ambitious standard. It aims at offering support for a great variety of audiovisual information, including still images, video, audio, text, (synthetic) talking heads and synthesized speech, synthetic graphics and 3D scenes, streamed data applied to media objects, and user interaction – e.g. changes of viewpoint.

Let's give an example, taken from the MPEG-4 standard document.

example

> *Imagine, a talking figure standing next to a desk and a projection screen, explaining the contents of a video that is being projected on the screen, pointing at a globe that stands on the desk. The user that is watching that scene decides to change from viewpoint to get a better look at the globe ...*

How would you describe such a scene? How would you encode it? And how would you approach decoding and user interaction?

The solution lies in defining *media objects* and a suitable notion of composition of media objects.

media objects

- *media objects* – units of aural, visual or audiovisual content
- *composition* – to create compound media objects (audiovisual scene)
- *transport* – multiplex and synchronize data associated with media objects
- *interaction* – feedback from users' interaction with audiovisual scene

For 3D-scene description, MPEG-4 builds on concepts taken from VRML (Virtual Reality Modeling Language, discussed in chapter 7).

Composition, basically, amounts to building a *scene graph*, that is a tree-like structure that specifies the relationship between the various simple and compound media objects. Composition allows for placing media objects anywhere in a given coordinate system, applying transforms to change the appearance of a media object, applying streamed data to media objects, and modifying the users viewpoint.

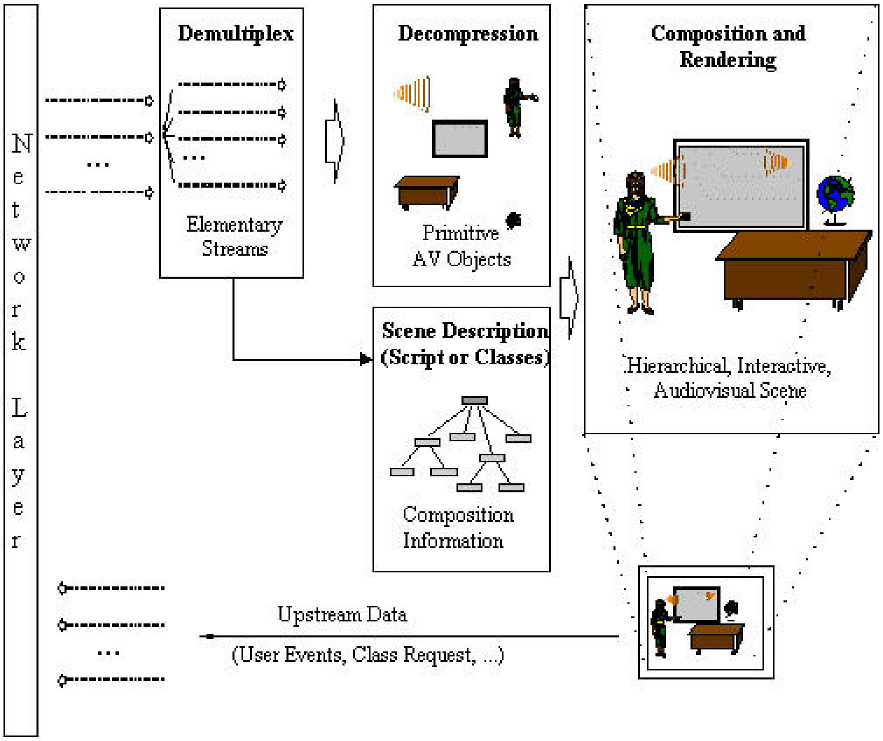
So, when we have a multimedia presentation or audiovisual scene, we need to get it accross some network and deliver it to the end-user, or as phrased in MPEG-4:

transport

> *The data stream (Elementary Streams) that result from the coding process can be transmitted or stored separately and need to be composed so as to create the actual multimedia presentation at the receivers side.*

At a system level, MPEG-4 offers the following functionalities to achieve this:

- BIFS (Binary Format for Scenes) – describes spatio-temporal arrangements of (media) objects in the scene
- OD (Object Descriptor) – defines the relationship between the elementary streams associated with an object
- *event routing* – to handle user interaction

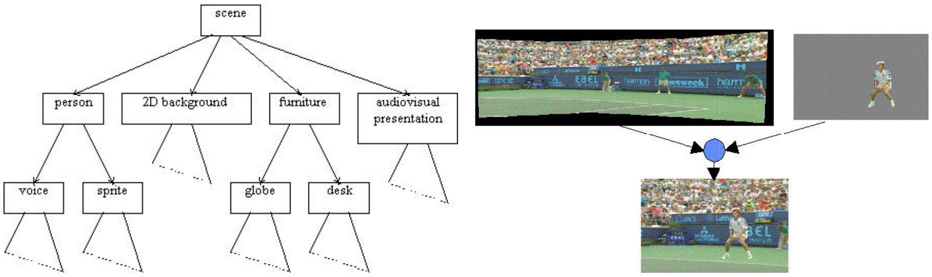

Network Layer
Demultiplex
Elementary Streams
Decompression
Primitive AV Objects
Composition and Rendering
Hierarchical, Interactive, Audiovisual Scene
Scene Description (Script or Classes)
Composition Information
Upstream Data
(User Events, Class Request, ...)

22

In addition, MPEG-4 defines a set of functionalities For the delivery of streamed data, DMIF, which stands for

Delivery Multimedia Integration Framework

that allows for transparent interaction with resources, irrespective of whether these are available from local storage, come from broadcast, or must be obtained from some remote site. Also transparency with respect to network type is supported. *Quality of Service* is only supoorted to the extent that it ispossible to indicate needs for bandwidth and transmission rate. It is however the responsability of the network provider to realize any of this.



(a) scene graph (b) sprites

23

authoring What MPEG-4 offers may be summarized as follows

benefits

- *end-users* – interactive media accross all platforms and networks
- *providers* – transparent information for transport optimization
- *authors* – reusable content, protection and flexibility

In effect, although MPEG-4 is primarily concerned with efficient encoding and scalable transport and delivery, the *object-based* approach has also clear advantages from an authoring perspective.

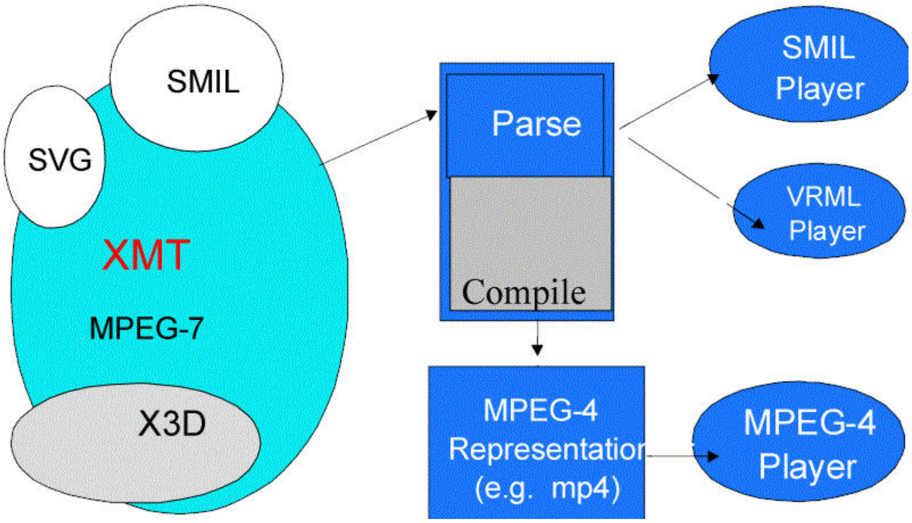
One advantage is the possibility of reuse. For example, one and the same background can be reused for multiplepresentations or plays, so you could imagine that even an amateur game might be 'located' at the centre-court of Roland Garros or Wimbledon.

Another, perhaps not so obvious, advantage is that provisions have been made for

managing intellectual property

of media objects.

And finally, media objects may potentially be annotated with meta-information to facilitate information retrieval.



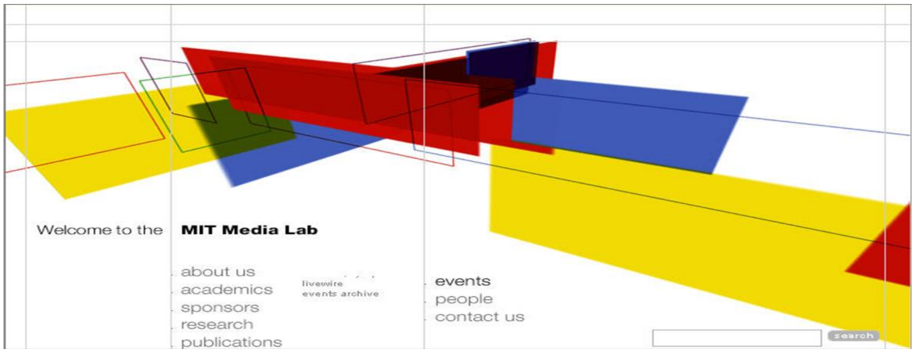
24

syntax In addition to the binary formats, MPEG-4 also specifies a syntactical format, called XMT, which stands for *eXtensible MPEG-4 Textual format.*

XMT

- XMT contains a subset of X3D
- SMIL is mapped (incompletely) to XMT

when discussing RM3D which is of interest from a historic perspective, we will further establish what the relations between, respectively MPEG-4, SMIL and RM3D are, and in particular where there is disagreement, for example with respect to the timing model underlying animations and the temporal control of media objects.



25

example(s) – *structured audio*

The Machine Listening Group[10] of the MIT Media Lab[11] is developing a suite of tools for *structered audio*, which means *transmitting sound by describing it rather than compressing it.* It is claimed that tools based on the MPEG-4 standard will be the future platform for computer music, audio for gaming, streaming Internet radio, and other multimedia applications.

The structured audio project is part of a more encompassing research effort of the Music, Mind and Machine Group[12] of the MIT Media Lab, which *envisages a new future of audio technologies and interactive applications that will change the way music is conceived, created, transmitted and experienced*,

SMIL

SMIL is pronounced as *smile*. SMIL, the Synchronized Multimedia Integration Language, has been inspired by the Amsterdam Hypermedia Model (AHM). In fact, the dutch research group at CWI that developed the AHM actively participated in the SMIL 1.0 committee. Moreover, they have started a commercial spinoff to create an editor for SMIL, based on the editor they developed for CMIF. The name of the editor is GRINS. Get it?

As indicated before SMIL is intended to be used for

TV-like multimedia presentations

The SMIL language is an XML application, resembling HTML. SMIL presentations can be written using a simple text-editor or any of the more advanced tools, such as GRINS. There is a variety of SMIL players. The most wellknown perhaps is the RealNetworks G8 players, that allows for incorporating RealAudio and RealVideo in SMIL presentations.

[10] sound.media.mit.edu/mpeg4

[11] www.media.mit.edu

[12] sound.media.mit.edu

parallel and sequential

> *Authoring a SMIL presentation comes down, basically, to name media components for text, images,audio and video with URLs, and to schedule their presentation either in parallel or in sequence.*

Quoting the SMIL 2.0 working draft, we can characterize the SMIL presentation characteristics as follows:

presentation characteristics

- The presentation is composed from several components that are accessible via URL's, e.g. files stored on a Web server.
- The components have different media types, such as audio, video, image or text. The begin and end times of different components are specified relative to events in other media components. For example, in a slide show, a particular slide is displayed when the narrator in the audio starts talking about it.
- Familiar looking control buttons such as stop, fast-forward and rewind allow the user to interrupt the presentation and to move forwards or backwards to another point in the presentation.
- Additional functions are "random access", i.e. the presentation can be started anywhere, and "slow motion", i.e. the presentation is played slower than at its original speed.
- The user can follow hyperlinks embedded in the presentation.

Where HTML has become successful as a means to write simple hypertext content, the SMIL language is meant to become a vehicle of choice for writing *synchronized hypermedia*. The working draft mentions a number of possible applications, for example a photoalbun with spoken comments, multimedia training courses, product demos with explanatory text, timed slide presentations, onlime music with controls.

As an example, let's consider an interactive news bulletin, where you have a choice between viewing a weather report or listening to some story about, for example, the decline of another technology stock. Here is how that could be written in SMIL:

example

```
<par>
   <a href=" #Story"> <img src="button1.jpg"/> </a>
   <a href=" #Weather"> <img src="button2.jpg"/></a>
    <excl>
        <par id="Story" begin="0s">
          <video src="video1.mpg"/>
          <text src="captions.html"/>
        </par>

        <par id="Weather">
          <img src="weather.jpg"/>
          <audio src="weather-rpt.mp3"/>
        </par>
    </excl>
</par>
```

Notice that there are two *parallel* (PAR) tags, and one *exclusive* (EXCL) tag. The *exclusive* tag has been introduced in SMIL 2.0 to allow for making an exclusive choice,so that only one of the items can be selected at a particular time. The SMIL 2.0 working draft defines a number of elements and attributes to control presentation, synchronization and interactivity, extending the functionality of SMIL 1.0.

Before discussing how the functionality proposed in the SMIL 2.0working draft may be realized, we might reflect on how to position SMIL with respect to the many other approaches to provide multimedia on the web. As other approaches we may think of *flash*, dynamic HTML (using javascript), or java applets. In the SMIL 2.0 working draft we read the following comment:

history

> *Experience from both the CD-ROM community and from the Web multimedia community suggested that it would be beneficial to adopt a declarative format for expressing media synchronization on the Web as an alternative and complementary approach to scripting languages.*
>
> *Following a workshop in October 1996, W3C established a first working group on synchronized multimedia in March 1997. This group focused on the design of a declarative language and the work gave rise to SMIL 1.0 becoming a W3C Recommendation in June 1998.*

In summary, SMIL 2.0 proposes a *declarative format* to describe the temporal behavior of a multimedia presentation, associate hyperlinks with media objects, describe the form of the presentation on a screen, and specify interactivity in multimedia presentations. Now,why such a fuzz about "declarative format"? Isn't scripting more exciting? And aren't the tools more powerful? Ok, ok. I don't want to go into that right now. Let's just consider a *declarative format* to be more elegant. Ok?

To support the functionality proposed for SMIL 2.0 the working draft lists a number of modules that specify the interfaces for accessing the attributes of the various elements. SMIL 2.0 offers modules for animation, content control, layout, linking, media objects, meta information, timing and synchronization, and transition effects.

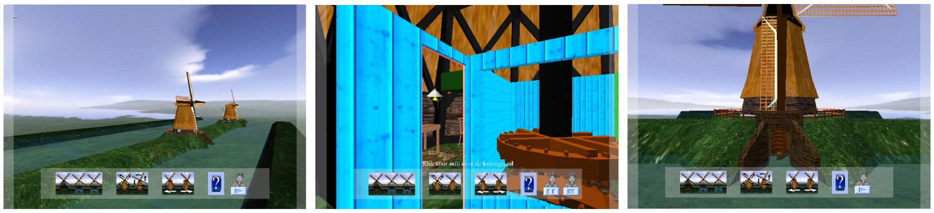
This modular approach allows to reuse SMIL syntax and semantics in other XML-based languages, in particular those that need to represent timing and synchronization. For example:

module-based reuse

- SMIL modules could be used to provide lightweight multimedia functionality on mobile phones, and to integrate timing into profiles such as the WAP forum's WML language, or XHTML Basic.
- SMIL timing, content control, and media objects could be used to coordinate broadcast and Web content in an enhanced-TV application.
- SMIL Animation is being used to integrate animation into W3C's Scalable Vector Graphics language (SVG).
- Several SMIL modules are being considered as part of a textual representation for MPEG4.

The SMIL 2.0 working draft is at the moment of writing being finalized. It specifies a number of language profiles topromote the reuse of SMIL modules. It also improves on the accessibility features of SMIL 1.0, which allows for, for example,, replacing captions by audio descriptions.

In conclusion, SMIL 2.0 is an interesting standard, for a number of reasons. For one, SMIL 2.0 has solid theoretical underpinnings in a well-understood, partly formalized, hypermedia model (AHM). Secondly, it proposes interesting functionality, with which authors can make nice applications. In the third place, it specifies a high level declarative format, which is both expressive and flexible. And finally, it is an open standard (as opposed to proprietary standard). So everybody can join in and produce players for it!



26

RM3D – not a standard

The web started with simple HTML hypertext pages. After some time static images were allowed. Now, there is support for all kinds of user interaction, embedded multimedia and even synchronized hypermedia. But despite all the graphics and fancy animations, everything remains flat. Perhaps surprisingly, the need for a 3D web standard arose in the early days of the web. In 1994, the acronym VRML was coined by Tim Berners-Lee, to stand for *Virtual Reality Markup Language*. But, since 3D on the web is not about text but more about worlds, VRML came to stand for *Virtual Reality Modeling Language*. Since 1994, a lot of progress has been made.

www.web3d.org

- VRML 1.0 – *static 3D worlds*
- VRML 2.0 or VRML97 – *dynamic behaviors*
- VRML200x – *extensions*
- X3D – *XML syntax*
- RM3D – *Rich Media in 3D*

In 1997, VRML2 was accepted as a standard, offering rich means to create 3D worlds with dynamic behavior and user interaction. VRML97 (which is the same as VRML2) was, however, not the success it was expected to be, due to (among others) incompatibility between browsers, incomplete implementations of the standards, and high performance requirements.

As a consequence, the Web3D Consortium (formerly the VRML Consortium) broadened its focus, and started thinking about extensions or modifications of VRML97 and an XML version of VRML (X3D). Some among the X3D working group felt the need to rethink the premisses underlying VRML and started the Rich Media Working Group:

groups.yahoo.com/group/rm3d/

> *The Web3D Rich Media Working Group was formed to develop a Rich Media standard format (RM3D) for use in next-generation media devices. It is a highly active group with participants from a broad range of companies including 3Dlabs, ATI, Eyematic, OpenWorlds, Out of the Blue Design, Shout Interactive, Sony, Uma, and others.*

In particular:

RM3D

> *The Web3D Consortium initiative is fueled by a clear need for a standard high performance Rich Media format. Bringing together content creators with successful graphics hardware and software experts to define RM3D will ensure that the new standard addresses authoring and delivery of a new breed of interactive applications.*

The working group is active in a number of areas including, for example, multitexturing and the integration of video and other streaming media in 3D worlds.

Among the driving forces in the RM3D group are Chris Marrin and Richter Rafey, both from Sony, that proposed *Blendo*, a rich media extension of VRML. Blendo has a strongly typed object model, which is much more strictly defined than the VRML object model, to support both declarative and programmatic extensions. It is interesting to note that the premisse underlying the Blendo proposal confirms (again) the primacy of the TV metaphor. That is to say, what Blendo intends to support are TV-like presentations which allow for user interaction such as the selection of items or playing a game. Target platforms for Blendo include graphic PCs, set-top boxes, and the Sony Playstation!

27

requirements The focus of the RM3D working group is not *syntax* (as it is primarily for the X3D working group) but *semantics*, that is to enhance the VRML97 standard to effectively incorporate rich media. Let's look in more detail at the requirements as specified in the RM3Ddraft proposal.

requirements

- *rich media* – audio, video, images, 2D & 3D graphics (with support for temporal behavior, streaming and synchronisation)
- *applicability* – specific application areas, as determined by commercial needs and experience of working group members

The RM3D group aims at interoperability with other standards.

- *interoperability* – VRML97, X3D, MPEG-4, XML (DOM access)

In particular, an XML syntax is being defined in parallel (including interfaces for the DOM). And, there is mutual interest and exchange of ideas between the MPEG-4 and RM3D working group.

As mentioned before, the RM3D working group has a strong focus on defining an object model (that acts as a common model for the representation of objects and their capabilities) and suitable mechanisms for extensibility (allowing for the integration of new objects defined in Java or C++, and associated scripting primitives and declarative constructs).

Notice that extensibility also requires the definition of a declarative format, so that the content author need not bother with programmatic issues.

The RM3D proposal should result in effective 3D media presentations. So as additional requirements we may, following the working draft, mention: high-quality realtime rendering, for realtime interactive media experiences; platform adaptability, with query functions for programmatic behavior selection; predictable behavior, that is a well-defined order of execution; a high precision number systems, greater than single-precision IEEE floating point numbers; and minimal size, that is both download size and memory footprint.

Now, one may be tempted to ask how the RM3D proposals is related to the other standard proposals such as MPEG-4 and SMIL, discussed previously. Briefly put, paraphrased from one of Chris Marrin's messages on the RM3D mailing list

> *SMIL is closer to the author* and *RM3D is closer to the implementer*.

MPEG-4, in this respect is even further away from the author since its chief focus is on compression and delivery across a network.

RM3D takes 3D scene description as a starting point and looks at pragmatic ways to integrate rich media. Since 3D is itself already computationally intensive, there are many issues thatarise in finding efficient implementations for the proposed solutions.

28

timing model RM3D provides a declarative format formany interesting features, such as for example texturing objects with video. In comparison to VRML, RM3D is meant to provide more temporal control over time-based media objects and animations. However, there is strong disagreement among the working group members as to what time model the dynamic capabilities of RM3D should be based on. As we read in the working draft:

working draft

> *Since there are three vastly different proposals for this section (time model), the original <RM3D> 97 text is kept. Once the issues concerning time-dependent nodes are resolved, this section can be modified appropriately.*

Now, what are the options? Each of the standards discussed to far provides us with a particular solution to timing. Summarizing, we have a time model based on a spring metaphor in MPEG-4, the notion of cascading time in SMIL (inspired by cascading stylesheets for HTML) and timing based on the routing of events in RM3D/VRML.

The MPEG-4 standard introduces the *spring metaphor* for dealing with temporal layout.

MPEG-4 – spring metaphor

- duration – minimal, maximal, optimal

The spring metaphor amounts to the ability to shrink or stretch a media object within given bounds (minimum, maximum) to cope with, for example, network delays.

The SMIL standard is based on a model that allows for propagating durations and time manipulations in a hierarchy of media elements. Therefore it may be referred to as a cascading modelof time.

SMIL – cascading time

- time container – speed, accelerate, decelerate, reverse, synchronize

Media objects, in SMIL, are stored in some sort of container of which the timing properties can be manipulated.

```
<seq speed="2.0">
   <video src="movie1.mpg" dur="10s"/>
   <video src="movie2.mpg" dur="10s"/>
   <img src="img1.jpg" begin="2s" dur="10s">
               <animateMotion from="-100,0" to="0,0" dur="10s"/>
   </img>
   <video src="movie4.mpg" dur="10s"/>
</seq>
```

In the example above,we see that the speed is set to *2.0*, which will affect the pacing of each of the individual media elements belonging to that (sequential) group. The duration of each of the elements is specified in relation to the parent container. In addition, SMIL offers the possibility to synchronize media objects to control, for example, the end time of parallel media objects.

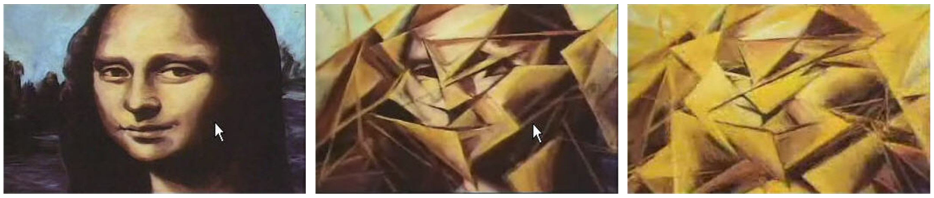
VRML97's capabilities for timing rely primarily on the existence of a *TimeSensor* thatsends out time events that may be routed to other objects.

RM3D/VRML – event routing

- *TimeSensor* – isActive, start, end, cycleTime, fraction, loop

When a *TimeSensor* starts to emit time events, it also sends out an event notifying other objects that it has become active. Dependent on itsso-called *cycleTime*, it sends out the fraction it covered since it started. This fraction may be send to one of the standard interpolators or a script so that some value can be set, such as for example the orientation, dependent on the fraction of the time intercal that has passed. When the *TimeSensor* is made to loop, this is done repeatedly. Although time in VRML is absolute, the frequency with which fraction events are emitted depends on the implementation and processor speed.

Lacking consensus about a better model, this model has provisionally been adopted, with some modifications, for RM3D. Nevertheless, the SMIL cascading time model has raised an interest in the RM3D working group, to the extent that Chris Marrin remarked (in the mailing list) *"we could go to school here"*. One possibility for RM3D would be to introduce *time containers* that allow for a temporal transform of their children nodes, in a similar way as grouping containers allow for spatial transforms of their children nodes. However, that would amount to a dual hierarchy, one to control (spatial) rendering and one to control temporal characteristics. Merging the two hierarchies, as is (implicitly) the case in SMIL, might not be such a good idea, since the rendering and timing semantics of the objects involved might be radically different. An interesting problem, indeed, but there seems to be no easy solution.



29

example(s) – *rich internet applications*

In a seminar held by *Lost Boys*, which is a dutch subdivison if Icon Media Lab[13], *rich internet applications* (RIA), were presented as the new solutions to present applications on the web. As indicated by Macromedia[14], who is one of the leading companies in this fiwld, *experience matters*, and so plain html pages pages do not suffice since they require the user to move from one page to another in a quite unintuitive fashion. Macromedia presents its new line of *flash*-based products to create such *rich internet applications*. An alternative solution, based on general W3C recommendations, is proposed by BackBase[15]. Interestingly enough, using either technology, many of the paricipants of the seminar indicated a strong preference for a backbuuton, having similar functionality as the often used backbutton in general internet browsers.

research directions– *meta standards*

All these standards! Wouldn't it be nice to have one single standard that encompasses them all? No, it would not! Simply, because such a standard is inconceivable, unless you take some proprietary standard or a particular platform as the defacto standard (which is the way some people look at the Microsoft win32 platform, ignoring the differences between 95/98/NT/2000/XP/...). In fact, there is a standard that acts as a glue between the various standards for multimedia, namely XML. XML allows for the interchange of data between various multimedia applications, that is the transformation of one encoding into another one. But this is only syntax. What about the semantics?

Both with regard to delivery and presentation the MPEG-4 proposal makes an attempt to delineate chunks of core fuctionality that may be shared between applications. With regard to presentation, SMIL may serve as an example. SMIL applications themselves already (re)use functionality from the basic set of XML-related technologies, for example to access the document structure through the DOM (Document Object Model). In addition, SMIL defines components that it may potentially share with other applications. For example, SMIL shares its animation facilities with SVG (the Scalable Vector Graphics format recommended by the Web Consortium).

The issue in sharing is, obviously, how to relate constructs in the syntax to their operational support. When it is possible to define a common base of operational support

[13] www.iconmedialab.com

[14] www.macromedia.com/resources/business/rich_internet_apps/whitepapers.html

[15] www.backbase.com

for a variety of multimedia applications we would approach our desired meta standard, it seems. A partial solution to this problem has been proposed in the now almost forgotten HyTime standard for time-based hypermedia. HyTime introduces the notion of *architectural forms* as a means to express the operational support needed for the interpretation of particular encodings, such as for example synchronization or navigation over bi-directional links. Apart from a base module, HyTime compliant architectures may include a units measurement module, a module for dealing with location addresses, a module to support hyperlinks, a scheduling module and a rendition module.

To conclude, wouldn't it be wonderful if, for example, animation support could be shared between rich media X3D and SMIL? Yes, it would! But as you may remember from the discussion on the timing models used by the various standards, there is still to much divergence to make this a realoistic option.

4.1 developments in hardware and software

Following Moore's law (predicting the doubling of computing power every eighteen months), computer hardware has significantly improved. But perhaps more spectacular is the growth in computing power of dedicated multimedia hardware, in particular what is nowadays called the GPU (graphics processing unit). In Cg, he NVIDIA GeForce FX GPU is said to have 125 million of transistors, whereas the Intel 2.4GHz Pentium 4 contains only 55 million of transistors. Now, given the fact that the CPU (central processing unit) is a general purpose, or as some may like to call it, *universal* device, why is it necessary or desirable to have such specialized hardware, GPUs for graphics and, to be complete DSPs (digital signal processors) for audio?

a little bit of history

Almost everyone knows the stunning animation and effects in movies made possible by computer graphics, as for example the latest production of Pixar, *The Incredibles.* Such animation and effects are only possible by offline rendering, using factories of thousands of CPUs, crunching day and night to render all the frames.

At the basis of rendering lies traditional computer graphics technology. That is, the transformation of vertices (points in 3D space), rasterization (that is determining the pixel locations and pixel properties corresponding to the vertices), and finally the so-called raster operations (determining whether and how the pixels are written to the framebuffer). OpenGL, developed by SGI was the first commonly available software API (application programmers interface) to control the process of rendering. Later, Microsoft introduced Direct3D as an alternative for game programming on the PC platform.

The process outlined above is called the *graphics pipeline.* You put models, that is collections of vertices, in and you get (frames of) pixels out. This is indeed a simplification in that it does not explain how, for example, animation and lighting effects are obtained. To gain control over the computation done in the graphics pipeline, Pixar developed Renderman, which allows for specifying transformations on the models (vertices) as well as operations on the pixels (or fragments as they are called in Cg) in a high level language. As vertex operations you may think of for example distortions of shape due to a force such as an explosion. As pixel operations, the coloring of pixels using textures (images) or special lighting and material properties. The languages for specifying such vertex or pixel operations are collectively called *shader* languages. Using offline rendering, almost anything is possible, as long as you specify it mathematically in a computationally feasible way.

The breakthrough in computer graphics hardware was to make such shading languages available for real-time computer graphics, in a way that allows, as Cg phrase it, 3D game and application programmers and real-time 3D artists to use it in an effective way.

Leading to the programmable computer graphics hardware that we know today, Cg distinguish between four generations of 3D accellerators.[16]

4 generations of GPU

- Before the introduction of the GPU, there only existed very expensive specialized hardware such as the machines from SGI.
- The first generation of GPU, including NVIDIA TNT2, ATI Rage and 3dfx Voodoo3, only supported rasterizing pre-transformed triangles and some limited texture operations.
- The second generation of GPUs, which were introduced around 1999, included the NVIDIA GeForce 2 and ATI Radeon 7500. They allowed for both 3D vertex transformations and some lighting, conformant with OpenGL and DirectX 7.
- The tird generation GPUs, including NVIDIA GeForce 3, Microsoft Xbox and ATI Radeon 8500, included both powerful vertex processing capabilities and some pixel-based configuration operations, exceeding those of OpenGL and DirectX 7.
- Finally, the fourth generation of GPUs, such as the NVIDIA GeForce FX and ATI Radeon 9700, allow for both complex vertex and pixel operations.

The capabilities of these latter generations GPUs motivated the development of high level shader languages, such as NVIDIA Cg and Microsoft HLSL. High level dedicated graphics hardware programming languages to control what may be called the programmable graphics pipeline.

the (programmable) graphics pipeline

Before discussing shading languages any further, let's look in some more detail at the graphics pipeline. But before that you must have an intuitive grasp of what is involved in rendering a scene.

Just imagine that you have created a model, say a teapot, in your favorite tool, for example Maya or 3D Studio Max. Such a model may be regarded as consisting of polygons, let's say triangles, and each vertex (point) of these triangles has apart from its position in (local) 3D space also a color. To render this model it must first be positioned in your scene, using so-called world coordinates. The *world transformation* is used to do this. The world transformation may change the position, rotation and scale of your object/model. Since your scene is looked at from one particular point of view we need to apply also a so-called *view transformation*, and to define how our view will be projected on a 2D plane, we must specify a *projection transformation.* Without going into the mathematical details, we may observe that these transformations can be expressed as 4x4 matrices and be combined in a single matrix, often referred to as the *world view projection matrix*, that can be applied to each of the vertices of our model. This combined transformation is the first stage in the process of rendering:

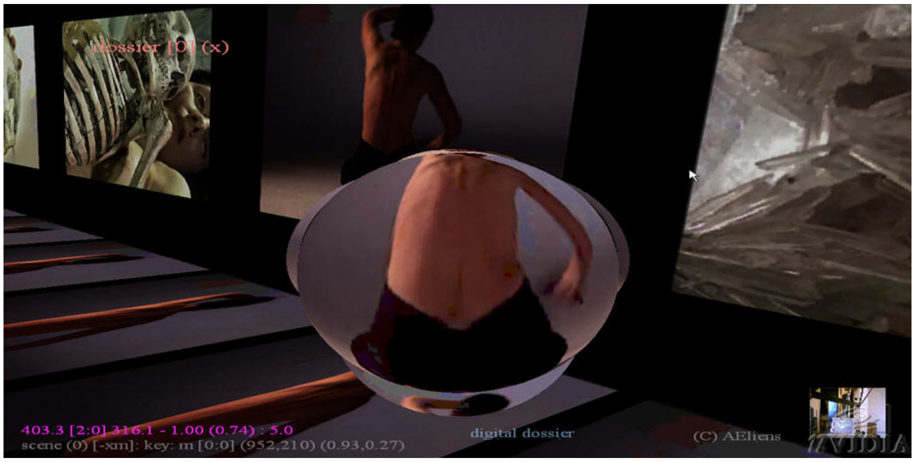
[16] The phrase GPU was introduced by NVIDIA to indicate that the capabilities of the GPU far exceed those of the VGA (video graphics array) originally introduced by IBM, which is nothing more than a dumb framebuffer, requiring updates from the CPU.

graphics pipeline

1. vertex transformation – apply world, view, projection transforms
2. assembly and rasterization – combine, clip and determine pixel locations
3. fragment texturing and coloring – determine pixel colors
4. raster operations – update pixel values

The second phase, roughly, consists of cleaning up the collection of (transformed) vertices and determining the pixel locations that correspond to the model. Then, in the third phase, using interpolation or some more advanced method, coloring and lighting is applied, and finally a sequence of per-fragment or pixel operations is applied. Both OpenGL and Direct3D support among others an alpha (or transparency) test, a depth test and blending. The above characterized the fixed function graphics pipeline. Both the OpenGL and Direct3D API support the fixed function pipeline, offering many ways to set relevant parameters for, for example, applying lights, depth and texturing operations.

To understand what the programmable graphics pipeline can do for you, you would best look at some simple shader programs. In essence, the programmable pipeline allows you to perform arbitrary vertex operations and (almost) arbitrary pixel operations. For example, you can apply a time dependent morphing operation to your model. Or you can apply an amplification to the colors of your scene. But perhaps more interestingly, you can also apply an advanced lighting model to increase realism.



403.3 [2:0] 316.1 - 1.00 (0.74) : 5.0
scene (0) [-xm]: key: m [0:0] (952,210) (0.93,0.27)

A simple morphing shader in ViP, see section 4.3.

30

a simple shader

When I began with programming shaders myself, I started with looking at examples from the DirectX SDK. Usually these examples were quite complex, and my attempt at modifying them often failed. Being raised in theoretical computer science, I changed strategy and developed my first shader program called *id*, which did nothing. Well, it just acted as the identity function. Then later I used this program as a starting point for writing more complex shader programs.

The *id* shader program is written in the DirectX 9 HLSL (high level shader language), and makes use of the DirectX Effects framework, which allows for specifying multiple vertex and pixel shaders, as well as multiple techniques and multiple passes in a single file.

The program starts with a declaration, specifying the global names for respectively the texture and the world/view/projection matrix. Also a texture sampler is declared, of which the function will become clear later.

HLSL declarations

```
texture tex;
float4x4 wvp;        // World * View * Projection matrix

sampler tex_sampler = sampler_state
{
   texture = /<tex/>;
};
```

It then defines, respectively, the vertex shader input and output, as structures. This declaration follows the standard C-like syntax for specifying elements in a structure, except for the identifiers in capitals, which indicate the semantics of the fields, corresponding to pre-defined registers in the GPU data flow.

vertex shader data flow

```
struct vsinput {
   float4 position : POSITION;
   float3 normal : NORMAL;
   float2 uv : TEXCOORD0;
};
struct vsoutput {
   float4 position   : POSITION;   // vertex position
   float4 color    : COLOR0;      // vertex diffuse color
   float2 uv  : TEXCOORD0;  // vertex texture coords
};
```

When the *vs_id* function, given below, is called, the input arguments are filled by the registers corresponding to the semantics pf the input structure. Similarly, the output results in setting the registers corresponding to the semantics of the output structure.

vertex shader

```
vsoutput vs_id( vsinput vx ) {
   vsoutput vs;

   vs.position = mul(vx.position, wvp);
   vs.color = color;
   vs.uv = vx.uv;

   return vs;
}
```

The *vs_id* function does exactly what the fixed graphics pipeline would do when transforming vertices. It applies the transformation to the vertex and passes both color and texture sampling coordinates to the pixel shader.

The pixel shader has a single color as output, which is obtained by sampling the texture, using the (interpolated) vertex color to modify the result.

pixel shader

```
struct psoutput
{
    float4 color : COLOR0;
};


psoutput ps_id( vsoutput vs )
{
    psoutput ps;

    ps.color = tex2D(tex_sampler, vs.uv) * vs.color;

    return ps;
}
```

Note that the *tex_sampler* comes from the global declaration above. The function *text2D* is a built-in for obtaining a color value from the sampler.

Finally, for each technique and each pass within a technique, in our case one technique with one pass, it must be indicated which function must be used for respectively the vertex shader and the pixel shader.

technique selection

```
technique render_id
{
    pass P0
    {
        VertexShader = compile vs_1_1 vs_id();
        PixelShader  = compile ps_2_0 ps_id();
    }
}
```

To make actual use of this program, the effect must be invoked from a DirectX or OpenGL program using the interface offered by the API.

Examples of Impasto, see examples – impasto

31

a morphing shader Slightly more complex is an example adapted from the morphing shader that can be found in ATI's Rendermonkey. To make a shader that morphs a cube into a ball and back, you must manipulate the vertices and the normals of the cube. For this to work your cube must have sufficient vertices, which is a property you can set in the tool that you use to make a cube.

morphing (vertex) shader

```
float3 spherePos = normalize(vx.position.xyz);
float3 cubePos = 0.9 * vx.position.xyz;

float t = frac(speed * time);
t = smoothstep(0, 0.5, t) - smoothstep(0.5, 1, t);

// find the interpolation factor
float lrp = lerpMin + (lerpMax - lerpMin) * t;

// linearly interpolate the position and normal
vx.position.xyz = lerp(spherePos, cubePos, lrp);
vx.normal = lerp(sphereNormal, cubeNormal, lrp);

// apply the transformations
vs.position = mul(wvp, vx.position);
```

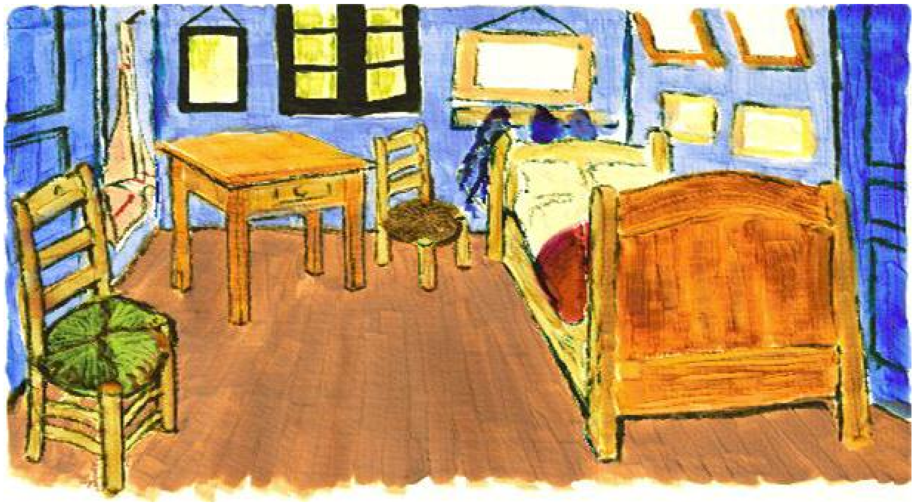
The example uses the built-in function *lerp*, that performs linear interpolation between two values using an interpolation factor between 0 and 1.

color amplification As an example of a pixel shader, look at the fragment defining an amplification of coloring below. It merely amplifies the RGB components of the colors when this exceeds a certain treshold.

coloring (pixel) shader

```
float4 x = tex2D(tex_sampler, vs.uv);
if (x.r > x.g && x.r > x.b) { x.r *= xi; x.g *= xd; x.b *= xd; }
else  if (x.g > x.r && x.g > x.b) { x.g *= xi; x.r *= xd; x.b *= xd; }
else  if (x.b > x.r && x.b > x.g) { x.b *= xi; x.r *= xd; x.g *= xd; }
ps.color = x;
```

When you develop shaders you must keep in mind that a pixel shader is generally invoked far more often than a vertex shader. For example a cube can be defined using 12 triangles of each tree vertices. However, the number of pixels generated by this might be up to a million. Therefore any complex computation in the pixel shader will be immediately noticable in the performance. For example, a slightly more complex pixel shader than the one above makes my NVIDIA GeForce FX 5200 accelerated 3 GHz machine drop to 5 frames per second!



rendering of van Gogh painting with Impasto

32

example(s) – *impasto*

IMPaSTo[17] is a realistic, interactive model for paint. It allows the user to create images in the style of famous painters as in the example above, which is after a painting of van Gogh. The *impasto* system implements a physical model of paint to simulate the effect of acrylic or oilpaint, using Cg shaders for real-time rendering, Impasto.

research directions – *the art of shader programming*

At first sight, shader programming seems to be an esoteric endeavor. However, as already indicated in this section, there are a number of high level languages for shader programming, including NVIDIA Cg and Microsoft HLSL. Cg is a platform independent language, suitable for both OpenGL and Direct3D. However, counter to what you might expect also Microsoft HLSL can be used for the OpenGL platform when you choose the proper runtime support.

To support the development of shaders there are, apart from a number of books, some powerful tools to write and test your shaders, in particular the already mentioned ATI Rendermonkey tool, the CgFx tool, which both produce HLSL code, as well as the Cg viewer and the effect tool that comes with the Microsoft DirectX 9 SDK.

Although I am only a beginning shader programmer myself, I find it truly amazing what shaders can do. For a good introducion I advice Cg. Futher you may consult Shader1, Shader2 and Shader3. Written from an artist's perspective is ShaderArt.

[17] gamma.cs.unc.edu/IMPaSTo

4.4 development(s) – gaming is a waste of time

The title of this section is borrowed from a lecture given for the VU computer science student association (STORM[18]), indeed, entitled *gaming is a waste of time*. This admittedly provocative title was on the one hand meant to emphasize the notion *waste of time*, since according to some of my collegue staff members my involvement in game development and multimedia technology was a mere waste of time, from some (from my point of view) obscure academic perspective. On the other hand, it (that is the title) also raised a more serious issue. Not being a game player myself, I do (in some sense) consider game playing a waste of time. Not that I deny the learning or entertainment aspects of games. On the contrary! Yes, as a passing of time, I prefer to keep myself busy with the construction of games, that is the creative and technological aspects of game development. And, likewise, I advise my students to do so.

When I was asked, in an alumni interview with the magazine of CWI[19], whether I believed in Second Life, my answr was simply: *I believe in nothing*! I take Second Life as an object of study, not in the last place because it has recently become so surprisingly popular. Yet, to be fair, Second Life has, after closer inspection, also technological merits of its own right.

In VUSL, we wrote: from a technical perspective, Second Life offers an advanced game engine that visitors and builders use (implicitly) in their activities. For essential components of game engine(s), we refer to section 11.1. In the following table, we give a brief comparative technical overview of, respectively, the Blaxxun Community Server (BlC), AlphaWorld (AW), the open source Delta3D engine (D3D), the Half Life 2 Source SDK (HL2), and Second Life (SL).

	BlC	AW	D3D	HL2	SL
in-game building	-	+	+/-	-	++
avatar manipulation	+	++	+/-	+	++
artifical intelligence	+	-	+/-	+	-
server-side scripts	+	-	+/-	+	++
client-side scripts	++	-	+/-	+	-
extensibility	+	-	++	+	+/-
open source	-	-	++	-	+/-
open standards	-	-	+/-	-	+/-
interaction	+/-	+/-	++	++	+/-
graphics quality	+/-	+/-	++	++	+
built-in physics	-	-	+	++	+
object collision	-	-	++	++	+
content tool support	+/-	-	++	+	-

Obviously, open source engines allow for optimal extensibility, and in this respect the open source version of the SL client may offer many opportunities. Strong points of SL appear to be *in-game building*, *avatar manipulation*, and in comparison with BlC and AW *built-in physics* and *object collision detection*. Weak points appear to be *content development tool support*, and especially in comparison with D3D and HL2 *interaction*. For most types of action-game like interaction SL is simply too slow. This even holds for script-driven animations, as we will discuss in the next section. In comparison

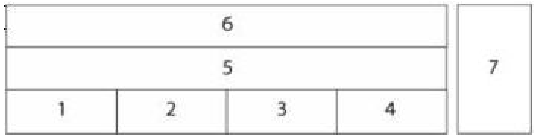
[18] www.storm.vu.nl

[19] www.cwi.nl

with a game as for example Age of Empires III[20], which offers in-game building and collaboration, Second Life distinguishes itself by providing a 3D immersive physics-driven environment, like the 'real' game engines.

Although we do not intend to realize Clima Futura in Second Life, we actually use *flash* to reach an audience as wide as possible, as a pilot parts of the game could fruitfully be realized in the VU virtual campus in Second Life, in particular the search for knowlegde, that is looking for an expert in a particular area of (climate-related) research. A similar quest was implemented in our Half Life 2 based game VULife, VULife, where the player had to visit nine information spots, which resulted in displaying in a HUD nine square matrix the location of a hidden treasure, which was then actually the power to use arms. Technical issues in realizing Clima Futura in Second Life are support for ranking, as well as meta-information with respect to locations where relevant information can be found, which may be realized with the techniques indicated in section 2.4.

In the beginning, we wrote in Climate, we envisioned the realization of our climate game as a first-person perspective role-playing game in a 3D immersive environment as for example supported by the Half Life 2 SDK, with which we gained experience in creating a *search the hidden treasure*[21] game in a detailed 3D virtual replica of our faculty. However, we soon realized that the use of such a development platform, would require far too much work, given the complexity of our design. So, instead of totally giving up on immersion, we decided to use *flash* video[22], indeed as a poor-man's substitute for real 3D immersion, which, using *flash*[23] interactive animations, has as an additional benefit that it can be used to play games online, in a web browser. Together with the Flex 2 SDK[24], which recently became open source, *flash* offers a rich internet application (RIA) toolkit, that is sufficiently versatile for creating (online) games, that require, in relation to console games or highly realistic narrative games like Half Life, a comparatively moderate development effort. To allow for component-wise development, we choose for a modular architecture, with four basic modules and three (variants) of integration modules, as indicated below.



Clima Futura architecture

module(s)

1. climate model(s) - action script module(s)
2. game play interaction - event-handler per game event
3. video content module - video fragment(s) and interaction overlays
4. minigame(s) - flash module(s) with actionscript interface
5. Clima Futura - integration of modules 1-4, plus server-side ranking
6. adapted versions – educational, commercial

[20] www.ageofempires3.com
[21] www.cs.vu.nl/∼eliens/game
[22] www.adobe.com/products/flash/video
[23] www.adobe.com/devnet/flash
[24] www.adobe.com/products/flex/sdk

7. multi-user version –with server-side support

In addition, we would like to develop a facility that allows players not only submit their own video material, but also to build or modify their own minigames, which might then be included in the collection of mini-games provided by Clima Futura. This, however, requires apart from a participatory (web 2.0) web-site, an appropriate game-description format, which we will discuss in section 11.4.

collada – gluing it all together

The wide variety of formats and tools for content production has been a stumbling block for many projects. How to find a unified format for digital content creation (DCC), so that content may be easily reused across projects and tools? A promising attempt in this direction is the *collada* initiative, Collada. The standard proposed in Collada is meant to serve as an intermediate or interchange format for interactive (multimedia) applications, such as games, which can be characterized as:

interactive application(s)

- realtime interaction – providing information
- content navigation – providing view(s)

Interactive multimedia applications have as a common property that, in contrast for example to computer graphics (CG) movies, everything must be available, that is computed, in real time. The intermediate format (*collada*), presented in Collada, is an XML-encoding of the various elements that may result from the content pipeline, that is the workflow of (digital) content creation, of a (multimedia) project, including:

collada[25]

- document(s) – *schema, asset, library, technique, ...*
- geometry – *array, accessor, meshes, vertices, polygons, ...*
- scene(s) – *material, light, optics, camera, imager, ...*
- effect(s) – *shader, profiles, techniques, pass, ...*
- animation(s) – *sampler, channel, controller, skin, morphing, ...*
- physics – *collision, equation, rigid body, constraints, force, ...*

The list above gives an indication of what (description) primitives the *collada* standard offers, to facilitate the exchange of (content) information and to promote re-use across tools and platforms.

5.1 scenarios

Multimedia is not only for entertainment. Many human activities, for example medical diagnosis or scientific research, make use of multimedia information. To get an idea about what is involved in multimedia information retrieval look at the following scenario, adapted from MMDBMS,

Amsterdam Drugport

[25] www.collada.org

> *Amsterdam is an international centre of traffic and trade. It is renowned for its culture and liberal attitude, and attracts tourists from various ages, including young tourists that are attracted by the availability of soft drugs. Soft drugs may be obtained at so-called coffeeshops, and the possession of limited amounts of soft drugs is being tolerated by the authories.*
>
> *The European Community, however, has expressed their concern that Amsterdam is the centre of an international criminal drug operation. Combining national and international police units, a team is formed to start an exhaustive investigation, under the code name* Amsterdam Drugport*.*

Now, without bothering ourselves with all the logistics of such an operation, we may establish what sorts of information will be gathered during the investigation, and what support for (multimedia) storage and (multimedia) information retrieval must be available.

Information can come from a variety of sources. Some types of information may be gathered continuously, for example by video cameras monitoring parking lots, or banks. Some information is already available, for example photographs in a (legacy database) police archive. Also of relevance may be information about financial transactions, as stored in the database of a bank, or geographic information, to get insight in possible dug traffic routes.

From a perspective of information storage our informatio (data) include the following media types: images, from photos; video, from surveillance; audio, from interviews and phone tracks; documents,from forensic research and reports; handwriting, from notes and sketches; and structured data, from for example bank transactions.

We have to find a way to store all these data by developing a suitable multimedia information system architecture, as discussed in chapter 6. More importantly, however, we must provide access to the data (or the information space, if you will) so that the actual police investigation is effectively supported. So, what kind of queries can we expect? For example, to find out more about a murder which seems to be related to the drugs operation.

retrieval

- *image query* – all images with this person
- *audio query* – identity of speaker
- *text query* – all transactions with BANK Inc.
- *video query* – all segments with victim
- *complex queries* – convicted murderers with BANK transactions
- *heterogeneous queries* – photograph + murderer + transaction
- *complex heterogeneous queries* – *in contact with* + murderer + transaction

Apparently, we might have simple queries on each of the media types, for example to detect the identity of a voice on a telephone wiretap. But we may also have more complex queries, establishing for example the likelihood that a murderer known by the police is involved, or even *heterogeneous* queries (as they are called in MMDBMS), that establish a relation between information coming from multiple information sources. An example of the latter could be, *did the person on this photo have any transactions with that bank in the last three months*, or nore complex, *give me all the persons that have been in contact with the victim (as recorded on audio phonetaps, photographs, and video surveillance tapes) that have had transactions with that particular bank.*

I believe you'll have the picture by now. So what we are about to do is to investigate how querying on this variety of media types, that is images, text, audio and video, might be realized.

33

research directions– *information retrieval models*

Information retrieval research has quite a long history, with a focus on indexing text and developing efficient search algorithms. Nowadays, partly due to the wide-spread use of the web, research in information retrieval includes modeling, classification and clustering, system architectures, user interfaces, information visualisation, filtering, descriptive languages, etcetera. See IR.

Information retrieval, according to IR, deals with the representation, storage, organisation of, and access to information items. To see what is involved, imagine that we have a (user) query like:

> *find me the pages containg information on ...*

Then the goal of the information retrieval system is to retrieve information that is useful or relevant to the user, in other words: *information that satisfies the user's information need.*

Given an information repository, which may consist of web pages but also multimedia objects, the information retrieval system must extract syntactic and semantic information from these (information) items and use this to match the user's information need.

Effective information retrieval is determined by, on the one hand, the *user task* and, on the other hand, the *logical view* of the documents or media objects that constitute the information repository. As user tasks, we may distinguish between *retrieval* (by query) and *browsing* (by navigation). To obtain the relevant information in retrieval we generally apply *filtering*, which may also be regarded as a ranking based on the attributes considered most relevant.

The logical view of text documents generally amounts to a set of index terms characterizing theb document. To find relevant index terms, we may apply operations to the document, such as the elimination of stop words or text stemming. As you may easily see, full text provides the most complete logical view, whereas a small set of categories provides the most concise logical view. Generally, the user task will determine whether semantic richness or efficiency of search will be considered as more important when deciding on the obvious tradeoffs involved.

information retrieval models In IR, a great variety of information retrieval models is described. For your understanding, an information retrieval model makes explicit how index terms are represented and how the index terms characterizing an information item are matched with a query.

When we limit ourselves to the classic models for search and filtering, we may distinguish between:

information retrieval models

- boolean or set-theoretic models
- vector or algebraic models
- probabilistic models

Boolean models typically allow for *yes/no* answers only. The have a set-theoretic basis, and include models based on fuzzy logic, which allow for somewhat more refined answers.

Vector models use algebraic operations on vectors of attribute terms to determine possible matches. The attributes that make up a vector must in principle be orthogonal. Attributes may be given a weight, or even be ignored. Much research has been done on how to find an optimal selection of attributes for a given information repository.

Probabilistic models include general inference networks, and belief networks based on Bayesan logic.

Although it is somewhat premature to compare these models with respect to their effectiveness in actual information etrieval tasks, there is, according to IR, a general consensus that vector models will outperform the probabilistic models on general collections of text documents. How they will perform for arbitrary collections of multimedia objects might be an altogether different question!

Nevertheless, in the sections to follow we will focus primarily on generalized vector representations of multimedia objects. So, let's conclude with listing the advantages of vector models.

vector models

- attribute term weighting scheme improves performance
- partial matching strategy allows retrieval of approximate material
- metric distance allows for sorting according to degree of similarity

Reading the following sections, you will come to understand how to adopt an attribute weighting scheme, how to apply partial matching and how to define a suitable distance metric.

So, let me finish with posing a research issue: *How can you improve a particular information retrieval model or matching scheme by using a suitable method of knowledge* *representation and reasoning?* To give you a point of departure, look at the logic-based multimedia information retrieval system proposed in Dolores.

5.2 images

An image may tell you more than 1000 words. Well, whether images are indeed a more powerful medium of expression is an issue I'd rather leave aside. The problem how to get information out of an image, or more generally how to query image databases is, in the context of our *Amsterdam Drugport* operation more relevant. There are two issues here

- obtaining descriptive information
- establishing similarity

These issues are quite distinct, although descriptive information may be used to establish similarity.

descriptive information

When we want to find, for example, all images that contain a person with say sunglasses, we need to have of the images in our database that includes this information one way or another. One way would be to annotate all images with (meta) information and describe the objects in the picture to some degree of detail. More challenging would be to extract image content by image analysis, and produce the description (semi) automatically.

According to MMDBMS, content-based description of images involves the identification of objects, as well as an indication of where these objects are located in the image, by using a *shape descriptor* and possibly *property descriptors* indicating the pictorial properties of a particular region of the object or image.

Shape and property descriptors may take a form as indicated below.

shape

- bounding box – (XLB,XUB,YLB,YUB)

property

- property – name=value

As an example of applying these descriptors.

example

shape descriptor: XLB=10; XUB=60; YLB=3; YUB=50
property descriptor: pixel(14,7): R=5; G=1; B=3

Now, instead of taking raw pixels as the unit of analysis, we may subdivide an image in a grid of cells and establish properties of cells, by some suitable algorithm.

definitions

- image grid: $(m * n)$ cells of equal size
- cell property: (Name, Value, Method)

As an example, we can define a property that indicates whether a particular cell if black or white.

example

property: (bwcolor,{b,w},bwalgo)

The actual algorithm used to establish such a property might be a matter of choice. So, in the example it is given as an explicit parameter.

From here to automatic content description is, admittedly, still a long way. We will indicate some research directions at the end of this section.

34

similarity-based retrieval

We need not necessarily know what an image (or segment of it) depicts to establish whether there are other images that contain that same thing, or something similar to it. We may, following MMDBMS, formulate the problem of similarity-based retrieval as follows:

> *How do we determine whether the content of a segment (of a segmented image) is similar to another image (or set of images)?*

Think of, for example, the problem of finding all photos that match a particular face.

According to MMDBMS, there are two solutions:

- *metric approach* – distance between two image objects
- *transformation approach* – relative to specification

As we will see later, the transformation approach in some way subsumes the metric approach, since we can formulate a distance measure for the transformation approach as well.

metric approach What does it mean when we say, the distance between two images is less than the distance between this image and that one. What we want to express is that the first two images (or faces) are more alike, or maybe even identical.

Abstractly, something is a distance measure if it satisfies certain criteria.

metric approach

distance $d : X \rightarrow [0, 1]$ is distance measure if:

d(x,y) = d(y,x)
d(x,y) $\leqslant$ d(x,z) + d(z,y)
d(x,x) = 0

For your intuition, it is enough when you limit yourself to what you are familiar with, that is measuring distance in ordinary (Euclidian) space.

Now, in measuring the distance between two images, or segments of images, we may go back to the level of pixels, and establish a distance metric on pixel properties, by comparing all properties pixel-wise and establishing a distance.

pixel properties

- objects with pixel properties $p_1, \ldots, p_n$
- pixels: $(x, y, v1, \ldots, v_n)$
- object contains w x h (n+2)-tuples

Leaving the details for your further research, it is not hard to see that even if the absolute value of a distance has no meaning, relative distances do. So, when an image contains a face with dark sunglasses, it will be closer to (an image of) a face with dark sunglasses than a face without sunglasses, other things being equal. It is also not hard to see that a pixel-wise approach is, computationally, quite complex. An object is considered as

complexity

a set of points in k-dimensional space for k = n + 2

In other words, to establish similarity between two images (that is, calculate the distance) requires n+2 times the number of pixels comparisons.

feature extraction Obviously,we can do better than that by restricting ourselves to a pre-defined set of properties or features.

feature extraction

- maps object into s-dimensional space

For example, one of the features could indicate whether or not it was a face with dark sunglasses. So, instead of calculating the distance by establishing color differences of between regions of the images where sunglasses may be found, we may limit ourselves to considering a binary value, yes or no, to see whether the face has sunglasses.

Once we have determined a suitable set of features that allow us to establish similarity between images, we no longer need to store the images themselves, and can build an index based on feature vectors only, that is the combined value on the selected properties.

Feature vectors and extensive comparison are not exclusive, and may be combined to get more precise results. Whatever way we choose, when we present an image we may search in our image database and present all those objects that fall within a suitable *similarity range*, that is the images (or segments of images) that are close enough according to the distance metric we have chosen.

35

transformation approach Instead of measuring the distance between two images (objects) directly, we can take one image and start modifying that until it exactly equals the target image. In other words, as phrased in MMDBMS, the principle underlying the transformation approach is:

transformation approach

> *Given two objects* o1 *and* o2*, the level of* dissimilarity *is proportional to the (minimum) cost of transforming object* o1 *into object* o2 *or vice versa*

Now, this principle might be applied to any representation of an object or image, including feature vectors. Yet, on the level of images, we may think of the following operations:

$to_1, \ldots, to_r$ – translation, rotation, scaling

Moreover, we can attach a cost to each of these operations and calculate the cost of a transformation sequence TSby summing the costs of the individual operations. Based on the cost function we can define a distance metric, which we call for obvious reasons the *edit distance*, to establish similarity between objects.

cost

- $cost(TS) = \Sigma_{i=1}^{r} cost(to_i)$

distance

- $d(o, o') = min\{cost(TS) \mid TSinTSeq(o, o')\}$

An obvious advantage of the *edit distance* over the pixel-wise distance metric is thatwe may have a rich choice of transformation operators that we can attach (user-defined) cost to at will.

For example, we could define low costs for normalization operations, such as scaling and rotation, and attach more weight tooperations that modify color values or add shapes. For face recognition, for example, we could attribute low cost to adding sunglasses but high cost to changing the sex.

To support the *transformation approach* at the image level, our image database needs to include suitable operations. See MMDBMS.

operations

```
rotate(image-id,dir,angle)
segment(image-id, predicate)
edit(image-id, edit-op)
```

We might even think of storing images, not as a collection of pixels, but as a sequence of operations on any one of a given set of base images. This is not such a strange idea as it may seem. For example, to store information about faces we may take a base collection of prototype faces and define an individual face by selecting a suitable prototype and a limited number of operations or additional properties.

36

example(s) – *match of the day*

The images in this section present a *match of the day*, which is is part of the project *split rpresentation* by the Dutch media artist Geert Mul. As explain in the email sending the images, about once a week, *Television images are recorded at random from satellite television and compared with each other. Some 1000.000.000 (one billion) equations are done every day.*

. The *split representation* project uses the image analyses and image composition software *NOTATION*[26], which was developed by Geert Mul (concept) and Carlo Preize (programming &mp; software design).

research directions – *multimedia repositories*

What would be the proper format to store multimedia information? In other words, what is the shape multimedia repositories should take? Some of the issues involved are discussed in chapter **??**, which deals with information system architectures. With respect to image repositories, we may rephrase the question into *what support must an image repository provide, minimally, to allow for efficient access and search?*. In MMDBMS, we find the following answer:

image repository

- *storage* – unsegmented images
- *description* – limited set of features

[26] homepage.mac.com/geertmul2

- *index* – feature-based index
- *retrieval* – distance between feature vectors

And, indeed, this seems to be what most image databases provide. Note that the actual encoding is not of importance. The same type of information can be encoded using either XML, relational tables or object databases. What is of importance is the functionality that is offered to the user, in terms of storage and retrieval as well as presentation facilities.

What is the relation between presentation facilities and the functionality of multimedia repositories? Consider the following mission statement, which is taken from my research and projects page.

mission

> *Our goal is to study aspects of the deployment and architecture of virtual environments as an interface to (intelligent) multimedia information systems <black>...*

Obviously, the underlying multimedia repository must provide adequate retrieval facilities and must also be able to deliver the desired objects in a format suitable for the representation and possibly incorporation in such an environment. Actually, at this stage, I have only some vague ideas about how to make this vision come through. Look, however, at chapter **??** and appendix **??** for some initial ideas.

37

5.3 documents

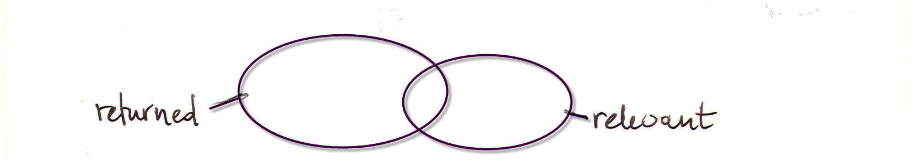
Even in the presence of audiovisual media, text will remain an important vehicel for human communication. In this section, we will look at the issues that arise in querying a text or document database. First we will characterize more precisely what we mean by effective search, and then we will study techniques to realize effective search for document databases.

Basically, answering a query to a document database comes down to string matching. However, some problems may occur such as synonymy and polysemy.

problems

- synonymy – topic T does not occur literally in document D
- polysemy – some words may have many meanings

As an example, *church* and *house of prayer* have more or less the same meaning. So documents about churches and cathedrals should be returned when you ask for information about 'houses of prayer'. As an exampleof polysemy, think of the word *drum*, which has quite a different meaning when taken from a musical perspective than from a transport logistics perspective.



precision and recall

Suppose that, when you pose a query, everything that is in the database is returned. You would probably not be satisfied, although every relevant document will be included, that is for sure. On the other hand, when nothing is returned, at least you cannot complain about non-relevant documents that are returned, or can you?

In MMDBMS, the notions of *precision* and *recall* are proposed to measure the effectiveness of search over a document database. In general, precision and recall can be defined as follows.

effective search

- precision – how many answers are correct
- recall – how many of the right documents are returned

For your intuition, just imagine that you have a database of documents. With full knowledge of the database you can delineate a set of documentsthatare of relevance to a particular query. Also, you can delineate a set that will be returned by some given search algorithm. Then, *precision* is the intersection of the two sets in relation to whatthe search algorithm returns, and *recall* that same intersection in relation to what is relevant. In pseudo-formulas, we can express this as follows:

precision and recall

precision = (returned and relevant) / returned
recall = (returned and relevant) / relevant

Now, as indicated in the beginning, it is not to difficult to get either perfect recall (by returning all documents) or perfect precision (by returning almost nothing). But these must be considered anomalies (that is, sick cases), and so the problem is to find an algorithm that performs optimally with respect to both precision and recall.

For the total database we can extend these measures by taking the averages of precision and recall for all topics that the database may be queried about.

Can these measures only be applied to document databases? Of course not, these are general measures that can be applied to search over any media type!

frequency tables

A *frequency table* is an example of a way to improve search. Frequency tables, as discussed in MMDBMS, are useful for documents only. Let's look at an example first.

example

term/document	d0	d1	d2
snacks	1	0	0
drinks	1	0	3
rock-roll	0	1	1

Basically, what a frequency table does is, as the name implies, give a frequency count for particular words or phrases for a number of documents. In effect, a complete document database may be summarized in a frequency table. In other words, the frequency table may be considered as an index to facilitate the search for similar documents.

To find a similar document, we can simply make a word frequency count for the query, and compare that with the colums in the table. As with images, we can apply a simpledistance metric to find the nearest (matching) documents. (In effect, we may take the square root for the sum of the squared differences between the entries in the frequence count as our distance measure.)

The complexity of this algorithm may be characterized as follows:

complextity

compare term frequencies per document – O(M*N)

where M is the number of terms and N is the number of documents. Since both M and N can become very large we need to make an effort to reduce the size of the frequency table.

reduction

- stop list – irrelevant words
- word stems – reduce different words to relevant part

We can, for example, introduce a *stop list* to prevent irrelevant words to enter the table, and we may restrict ourselves to including *word stems* only, to bring back multiple entries to one canonical form. With some additional effort we could even deal with synonymy and polysemy by introducing, respectively equivalence classes, and alternatives (although we then need a suitable way for ambiguation). By the way, did you notice that frequency tables may be regarded as feature vectors for documents?

38

research directions– *user-oriented measures*

Even though the reductions proposed may result in limiting the size of the frequency tables, we may still be faced with frequency tables of considerable size. One way to reduce the size further, as discussed in MMDBMS, is to apply *latent sematic indexing* which comes down to clustering the document database, and limiting ourselves to the most relevant words only, where relevance is determined by the ratio of occurrence over the total number of words. In effect, the less the word occurs, the more discriminating it might be. Alternatively,the choice of what words are considered relevant may be determined by taking into account the area of application or the interest of a particular group of users.

39

user-oriented measures Observe that, when evaluating a particular information retrieval system, the notions of precision and recall as introduced before are rather system-oriented measures, based on the assumption of a user-independent notion of relevance. However, as stated in IR, different users might have a different interpretation on which document is relevant. In IR, some user-oriented measures are briefly discussed, that to some extent cope with this problem.

user-oriented measures

- *coverage ratio* – fraction of known documents
- *novelty ratio* – fraction of new (relevant) documents
- *relative recall* – fraction of expected documents
- *recall effort* – fraction of examined documents

Consider a reference collection, an example information request and a retrieval strategy to be evaluated. Then the *coverage ratio* may be defined as the fraction of the documents known to be relevant, or more precisely the number of (known) relevant documents retrieved divided by the total number of documents known to be relevant by the user.

The *novelty ratio* may then be defined as the fraction of the documents retrieved which were not known to be relevant by the user, or more precisely the number of relevent documents that were not known by the user divided by the total number of relevant documents retrieved.

The *relative recall* is obtained by dividing the number of relevant documents found by the number of relevant documents the user expected to be found.

Finally, *recall effort*may be characterized as the ratio of the number of relevant documents expected and the total number of documents that has to be examined to retrieve these documents.

Notice that these measures all have a clearly 'subjective' element, in that, although they may be generalized to a particular group of users, they will very likely not generalize to all groups of users. In effect, this may lead to different retrieval strategies for different categories of users, taking into account levelof expertise and familiarity with the information repository.

40

6.2 video

Automatic content description is no doubt much harder for video than for any other media type. Given the current state of the art, it is not realistic to expect content description by feature extraction for video to be feasible. Therefore,to realize content-based search for video, we have rely on some knowledge representation schema that may adequately describe the (dynamic) properties of video fragments.

In fact, the description of video content may reflect the story-board, that after all is intended to capture both time-independent and dynamically changing properties of the objects (and persons) that play a role in the video.

In developing a suitable annotation for a particular video fragment, two questions need to be answered:

video annotation

- what are the interesting aspects?
- how do we represent this information?

Which aspects are of interest is something you have to decide for yourself. Let's see whether we can define a suitable knowledge representation scheme.

One possible knowledge representation scheme for annotating video content is proposed in MMDBMS. The scheme proposed has been inspired by knowledge representation techniques in Artificial Intelligence. It captures both static and dynamic properties.

video content

video v, frame f
f has associated objects and activities

objects and activities have properties

First of all, we must be able to talk about a particular video fragment v, and frame f that occurs in it. Each frame may contain objects that play a role in some activity. Both objects and activities may have properties, that is attributes that have some value.

property

property: name = value

As we will see in the examples, properties may also be characterized using predicates.

Some properties depend on the actual frame the object is in. Other properties (for example sex and age) are not likely to change and may considered to be frame-independent.

object schema

(fd,fi) – frame-dependent and frame-independent properties

Finally, in order to identify objects we need an object identifier for each object. Summing up, for each object in a video fragment we can define an *object instance*, that characterizes both frame-independent and frame-dependent properties of the object.

object instance: (oid,os,ip)

- *object-id* – oid
- *object-schema* – os = (fd,fi)
- *set of statements* – ip: name = v and name = v IN f

Now, with a collection of object instances we can characterize the contents of an entire video fragment, by identifying the frame-dependent and frame-independent properties of the objects.

Look at the following example, borrowed from MMDBMS for the *Amsterdam Drugport* scenario.

frame	objects	*frame-dependent properties*
1	Jane	has(briefcase), at(path)
-	house	door(closed)
-	briefcase	
2	Jane	has(briefcase), at(door)
-	Dennis	at(door)
-	house	door(open)
-	briefcase	

In the first frame Jane is near the house, at the path that leads to the door. The door is closed. In the next frame, the door is open. Jane is at the door, holding a briefcase. Dennis is also at the door. What will happen next?

Observe thatwe are using predicates to represent the state of affairs. We do this, simply because the predicate form *has(briefcase)* looks more natural than the other form, which would be *has = briefcase*. There is no essential difference between the two forms.

Now, to complete our description we can simply list the frame-independent properties, as illustrated below.

object	*frame-independent properties*	value
Jane	age	35
	height	170cm
house	address	...
	color	brown
briefcase	color	black
	size	40 x 31

How to go from the tabular format to sets of statements that comprise the object schemas is left as an (easy) exercise for the student.

Let's go back to our *Amsterdam Drugport* scenario and see what this information might do for us, in finding possible suspects. Based on the information given in the example, we can determine that there is a person with a briefcase, and another person to which that briefcase may possibly be handed. Whether this is the case or not should be disclosed in frame 3. Now, what we are actually looking for is the possible exchange of a briefcase, which may indicate a drug transaction. So why not, following MMDBMS, introduce another somewhat more abstract level of description that deals with *activities*.

activity

- activity name – id
- statements – *role* = *v*

An activity has a name, and consists further simply of a set of statements describing the *roles* that take part in the activity.

example

{ giver : Person, receiver : Person, item : Object }
giver = Jane, receiver = Dennis, object = briefcase

For example, an *exchange* activity may be characterized by identifying the *giver*, *receiver* and *object* roles. So, instead of looking for persons and objects in a video fragment, you'd better look for activities that may have taken place, by finding a matching set of objects for the particular roles of an activity. Consult MMDBMS if you are interested in a further formalization of these notions.

video libraries

Assuming a knowledge representation scheme as the one treated above, how can we support search over a collection of videos or video fragments in a video library.

What we are interested in may roughly be summarized as

video libraries

which videos are in the library
what constitutes the content of each video
what is the location of a particular video

Take note that all the information about the videos or video fragments must be provided as meta-information by a (human) librarian. Just imagine for a moment how laborious and painstaking this must be, and whata relief video feature extraction would be for an operation like *Amsterdam Drugport*.

To query the collection of video fragments, we need a query language with access to our knowledge representation. It must support a variety of retrieval operations, including the retrieval of segments, objects and activities, and also property-based retrievals as indicated below.

query language for video libraries

- *segment retrievals* – *exchange of briefcase*
- *object retrievals* – all people in v:[s,e]
- *activity retrieval* – all activities in v:[s,e]
- *property-based* – find all videos with object oid

MMDBMS lists a collection of video functions that may be used to extend SQL into what we may call VideoSQL. Abstractly, VideoSQL may be characterized by the following schema:

VideoSQL

```
SELECT – v:[s,e]
FROM – video:<source><V>
WHERE – term IN funcall
```

where *v:[s,e]* denotes the fragment of video *v*, starting at frame *s* and ending at frame *e*, and *term IN funcall* one of the video functions giving access to the information about that particular video. As an example, look at the following VideoSQL snippet:

example

```
SELECT  vid:[s,e]
FROM video:VidLib
WHERE (vid,s,e) IN VideoWithObject(Dennis) AND
        object IN ObjectsInVideo(vid,s,e) AND
        object != Dennis AND
        typeof(object) = Person
```

Notice that apart from calling video functions also constraints can be added with respect to the identity and type of the objects involved.

42

example(s) – *video retrieval evaluation*

The goal of the TREC[27] conference series is to encourage research in information retrieval by providing a large test collection, uniform scoring procedures, and a forum for organizations interested in comparing their results. Since 2003 their is an independent *video* track devoted to research in automatic segmentation, indexing, and content-based retrieval of digital video. In the TRECVID[28] 2004 workshop, thirty-three teams from Europe, the Americas, Asia, and Australia participated. Check it out!

43

[27] trec.nist.gov

[28] www-nlpir.nist.gov/projects/trecvid

research directions– *presentation and context*

Let's consider an example. Suppose you have a database with (video) fragments of news and documentary items. How would you give access to that database? And, how would you present its contents? Naturally, to answer the first question, you need to provide search facilities. Now, with regard to the second question, for a small database, of say 100 items, you could present a list of videos thatb matches the query. But with a database of over 10.000 items this will become problematic, not to speak about databases with over a million of video fragments. For large databases, obviously, you need some way of visualizing the results, so that the user can quickly browse through the candidate set(s) of items.

Video provide an interesting account on how *interactive maps* may be used to improve search and discovery in a (digital) video library. As they explain in the abstract:

> *To improve library access, the Informedia Digital Video Library uses automatic processing to derive descriptors for video. A new extension to the video processing extracts geographic references from these descriptors.*
>
> *The operational library interface shows the geographic entities addressed in a story, highlighting the regions discussed in the video through a map display synchronized with the video display.*

So, the idea is to use geographical information (that is somehow available in the video fragments themselves) as an additional descriptor, and to use that information to enhance the presentation of a particular video. For presenting the results of a query, candidate items may be displayed as icons in a particular region on a map, so that the user can make a choice.

Obviously, having such geographical information:

> *The map can also serve as a query mechanism, allowing users to search the terabyte library for stories taking place in a selected area of interest.*

The approach to extracting descriptors for video fragments is interesting in itself. The two primary sources of information are, respectively, the spoken text and graphic text overlays (which are common in news items to emphasize particular aspects of the news, such as the area where an accident occurs). Both speech recognition and image processing are needed to extract information terms, and in addition natural language processing, to do the actual 'geocoding', that is translating this information to geographical locations related to the story in the video.

Leaving technical details aside, it will be evident that this approach works since news items may relevantly be grouped and accessed from a geographical perspective. For this type of information we may search, in other words, with three kinds of questions:

- *what* – content-related
- *when* – position on time-continuum
- *where* – geographic location

and we may, evidently, use the geographic location both as a search criterium and to enhance the presentation of query results.

mapping information spaces Now, can we generalize this approach to other type of items as well. More specifically, can we use maps or some spatial layout to display the results of a query in a meaningful way and so give better access to large databases of multimedia objects. According to Atlas, we are very likely able to do so:

> *More recently, it has been recognized that the process of spatialization – where a spatial map-like structure is applied to data where no inherent or obvious one does exist – can provide an interpretable structure to other types of data.*

Actually, we are taking up the theme of *visualization*, again. In Atlas visualizations are presented that (together) may be regarded as an *atlas of cyberspace*.

atlas of cyberspace

> *We present a wide range of spatializations that have employed a variety of graphical techniques and visual metaphors so as to provide striking and powerful images that extend from two dimension 'maps' to three-dimensional immersive landscapes.*

As you may gather from chapter 7 and the *afterthoughts*, I take a personal interest in the (research) theme of *virtual reality interfaces for multimedia information systems*. But I am well aware of the difficulties involved. It is an area that is just beginning to be explored!

8.1 virtual context

Imagine that you walk in a museum. You see a painting that you like. It depicts the Dam square in 17th century Amsterdam. Now, take a step forwards and suddenly you are in the middle of the scene you previously watched from some distance. These things happen in movies.

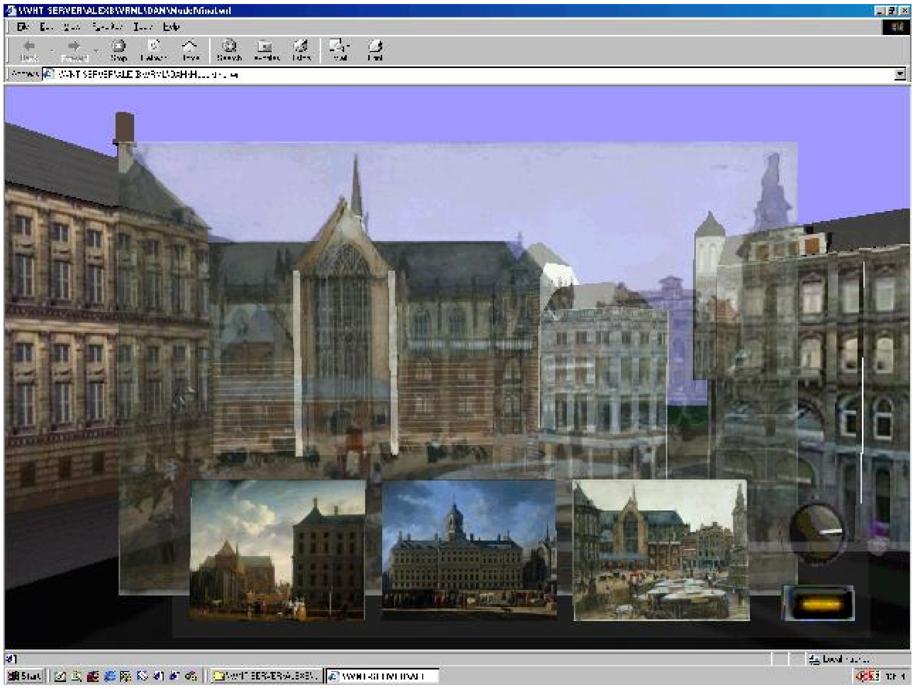
Now imagine that you are walking on the Dam square, some sunday afternoon in May 2001, looking at the Royal Palace, asking yourself is this where Willem-Alexander and Maxima will get married. And you wonder, what did this building and the Dam square look like three centuries ago. To satisfy your curiosity you go to the Royal Museum, which is only a half hour walk from there, and you go to the room where the 17th century city-scape paintings are. The rest is history.

We can improve on the latter scenario I think. So let's explore the options. First of all, we may establish that the Dam square represents a rich information space. Well, the Dam Square is a 'real world' environment, with it has 700 years of (recorded) history. It has a fair amount of historical buildings, and both buildings and street life have changed significantly over time.

So, we can rephrase our problem as

how can we give access to the 'Dam square' information space

But now we forget one thing. The idea underlying the last scenario is that we somehow realize a seamless transition from the real life experience to the information space. Well, of course, we cannot do that. So what did we do?



44

Look at the screenshot from our *virtual context* prototype. You can also start the VRML demo version that is online, by clicking on the screenshot. What you see is (a model of) the Dam square, more or less as it was in 2001. In the lower part, you see a panel with paintings. When you click on one of these painting, your viewpoint is changed so that you observe the real building from the point of view from which the painting was made. Then using the controls to the right of the panel, you can overlay the real building with a more or less transparent rendering of the painting. You can modify the degree of transparency by turning the dial control. You may also make the panel of paintings invisible, so that it does not disrupt your view of the Dam and the chosen overlay.

In other words, we have a VR model of Dam square and a selection of related paintings from the Royal Museum, that are presented in aa panel from which the user can choose a painting. We deploy viewpoint adjustment, to match the selected painting, and we use overlay of paintings over buildings, in varying degrees of transparancy, to give the user an impression of how the differences between the scene depicted in the painting and the actual scene in (the virtual) reality.

We have chosen for the phrase *virtual context* to characterize this prototype, since it does express how virtual reality technology enables us to relate an information space to its original context.

From the perspective of virtual reality, however, we could also have characterized our prototype as an application of *augmented virtual reality*, since what we have is a virtual reality model of a reallife location that is augmented with information that is related to it, (almost) without disrupting the virtual reality experience. In summary, we may characterize our approach as follows.

augmented virtual reality

- give user sense of geographic placement of buildings
- show how multiple objects in a museum relate to eachother
- show what paintings convey about their subject, and how

Considering the fact that many city-scape paintings of Amsterdam have been made, many of which are in the Royal Museum, and that paintings may say many things about their subject, we believe that our approach is viable for this particular instance. The augmented virtual reality approach would also qualify as a possible approach to cultural heritage projects, provided that sufficient pictorial material is available or can be reconstructed.

Although we were quite satisfied with what we accomplished, there are still many things that can be done and also a number of open problems. Guided tours are a wellknown phenomenon. But how to place them in our virtual context is not entirely clear. As another problem, our approach does not seem suited to account for buildings that do no longer exist. Another thing we have to study is how to change the temporal context, that is for example change from a model of the dam in 2001 to a model of the Dam in 1850. We would then also like to have 'viewpoint transitions' over space and time!

Finally, to give better access to the underlying information space we must also provide for textual user queries, and find an adequate response to those queries.

VRML To realize our prototype we used VRML, which limits us to medium quality desktop VR. At this stage, VRML is a good option, since it is a relatively stable format with a reasonable programmatic model. In short, what VRML offers is

VRML

- declarative means for defining geometry and appearance
- prototype abstraction mechanism
- powerful event model
- relatively strong programmatic capabilities

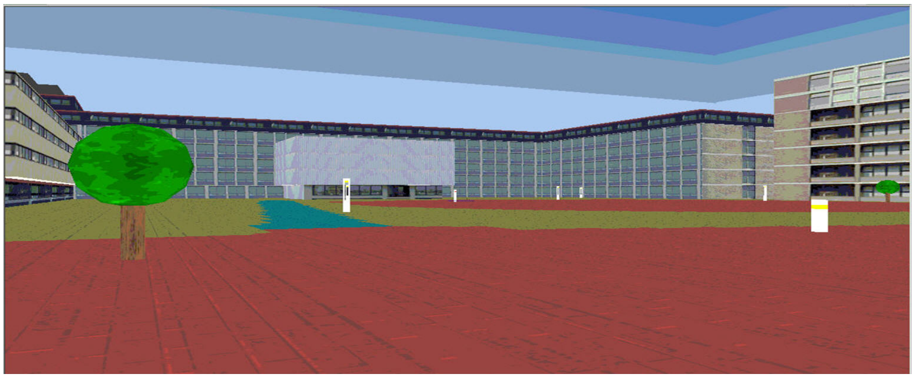
Although VRML allows for writing models (including geometry and appearance) using a plain text editor, many tools support export to VRML. As a consequence, often tools are used to create more complex models.

In addition, VRML allows for defining prototype abstractions, so reuse of models and behavior can be easily realized.

Defining dynamic behavior involves the routing of events that may come from a variety of built-in sensors (for example a TimeSensor for animations) to scripts or so-called interpolators, that allow for the manipulation of geometry and appearance parameters of the model.

In particular, the use of scripts or the *External Authoring Interface* (EAI), that allows for defining behavior in Java, is essential for realizing complex behavior.

Summarizing, VRML is a sufficiently rich declarative language for defining 3D scenes, with a relatively powerful programming model for realizing complex behavior. Some may think that VRML is dead. It isn't. The underlying model is endorsed in both the X3D and RM3D standards, simply since it has proven its worth.



45

research directions– *augmented virtuality*

Given an information space, there is a duality between information and presentation. For an audience or user to be able to digest a presentation, the amount of information must be limited. Effective presentation, moreover, requires the use of proper rethorics (which may be transcoded as *ways of presenting*) that belong to the medium. Using VR, which is (even in its desktop format) a powerful presentation vehicle, one should always beware of the question *what is it good for?* Generally one may ask, what is the added value of using VR? In an abstract fashion the answer should be, to bridge the gap between information content and presentation. Or, in other words, to resolve the duality between information and presentation!

Let's look at an example, a site about archeology, announced as a site offering *Virtual Archeology*. Perhaps it is good to bring to your attention that the *virtual*, in Computer Science, means nothing but another level of indirection to allow for a (more) flexible usage of entities or objects. See OO, section 1.2.

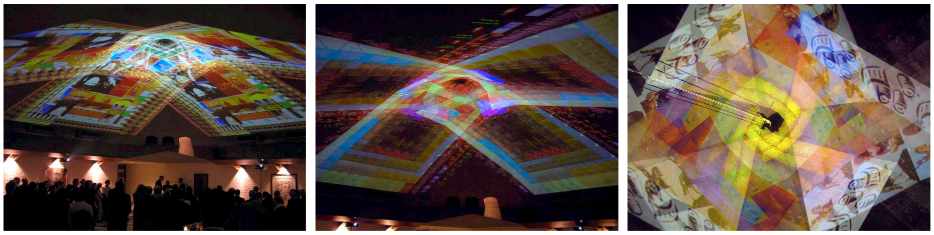
virtual archeology

- variety of archeological sites
- various paths through individual site
- reconstruction of 'lost' elements
- 'discovery' of new material
- glossary – general background knowledge

For a site about archeology, *virtual* means the ability to present the information in a number of ways, for example as paths through a particular site, with the possibility to explore the reconstruction of lost or perished material, and (for students) to discover new perspectives on the material. In addition, for didactic reasons there may also be a glossary to explain concepts from archeology.

Now, how would you construct such a site about virtual archeology? As a collection of HTML pages and links? It seems that we can do better, using VR and rich interaction mechanisms!

So, what is meant by *augmented virtuality*? Nothing that hasn't been expressed by the notion of *augmented virtual reality*, of which an example has been given in this section. The phrase *augmented virtuality* itself is just one of those potentially meaningless fancy phrases. It was introduced simply to draw your attention to the duality between information and presentation, and to invite you to think about possible ways to resolve this duality.



46

9.2 designing the user experience

In a time in which so much information is available as in ours, we may make statements like

postmodern design

> ... postmodern design is of a highly reflective nature ... appropriating design of the past ... in other words, sampling is allowed but no plagianarism

One interpretation might be that it is impossible to be original. But another interpretation might be that it is undesirable to be (totally) original. In other words, as discussed in section 2.3, it is necessary that your design contains references to not only some real situation but also to other work, so that it can be understood and experienced by the user/spectator/client whatever you like to call the people that look at your work. As observed in VR, designing for multimedia does not take into account only the technological or aesthetic issues, but also constraints on what people can perceive and what the experiential context is in which the work is presented, which may be re-phrased more plainly as what expectations the user has.

game design

Let us consider how these observations affect one of the project assignments for our *visual design* course. Also for *game design*, there are several options, dependent on the track the student is most comfortable with.

game design

- style – develop concept, plot and visual assets for *a game of choice*
- content – develop environments, models and animations for *a game of choice*
- effects – develop models, textures and special effects (shaders) for *a game of choice*

To explain, *style* may be considered to involve the whole gamut of concepts, plot and genre, as well as the visual assets or props of the game, those things by which the game differentiates itself from other games. Since the reader of this book will probably be more familiar with games than the author, there is no need to expand on these issues. *Content* is concerned with the actual game environment, including the models and animations. Finally, *effects*, to simplify things, is everything else, those things that are visual but does not belong to the story line or game environment.

Games, perhaps more than any other multimedia application, are appealing, not because they are useful, although they might be, but because the user gets emotionally involved, not to say addicted. Now, following Norman,

> did you ever wonder why cheap wine tastes better in fancy glasses?

Exactly, because cheap glasses do not give us the same emotion. It is, indeed, a matter of style!

Obviously, games are played for fun. As applications, games may be classified as seductive, which is, see section 2.3, stronger than persuasive. Norman distinguishes between four categories of pleasure.

seduction

- physio-pleasure – of the body
- socio-pleasure – by interaction with others
- psycho-pleasure – due to use of the product
- ideo-pleasure – reflecting on the experience

In other words, games are seductive, or fun to play, because they arouse any combination of pleasure from the categories above. Which combination depends on the kind or genre of game. Quoted from Norman, but originally from Wolf we can list, not exhaustively, the following genres of video game:

genre(s)[29]

> Abstract, Adaptation, Adventure, Artificial Life, Board Games, Capturing, Card Games, Catching, Chase, Collecting, Combat, Demo, Diagnostic, Dodging, Driving, Educational, Escape, Fighting, Flying, Gambling, Interactive Movie, Management Simulation, Maze, Obstacle Course, Pencil-and-Paper Games, Pinball, Platform, Programming Games, Puzzle, Quiz, Racing, Role-Playing, Rhythm and Dance, Shoot Em Up, Simulation, Sports, Strategy, Table-Top Games, Target, Text Adventure, Training Simulation, and Utility.

When you develop a game it is good to reflect on what genre your game belongs to, because that will directly affect the user's expectation when playing the game, due to the conventions and rules that exist within a particular genre. For video games, which can be characterized as *a mixture of interactive fiction with entertainment*, interaction is evidently another distinguishing factor in determining the success of the game.

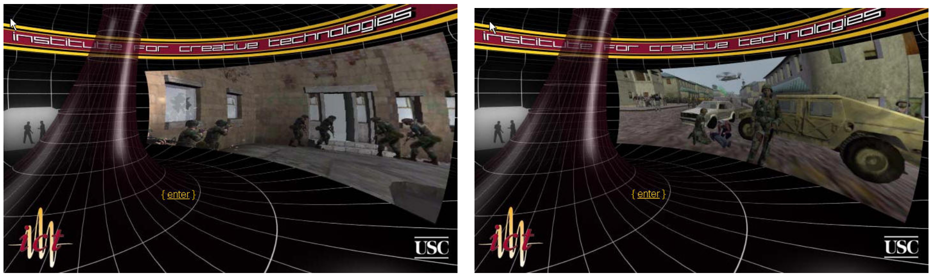
Corresponding to the kind of pleasure a user may experience, Norman distuinguishes between three levels of design:

levels of design

- visceral – what appeals to our intuition (*affordance*)
- behavioral – is all about use (*performance*)
- reflective – its all about message, culture and meaning

Of these, the latter two should be rather obvious, although we will elaborate on the notion of *usability* later on. But what does *affordance* mean, and how is it related to our intuition.

[29] www.robinlionheart.com/gamedev/genres.xhtml


institute for creative technologies
{ enter }
USC
institute for creative technologies
{ enter }
USC

47

affordance – ecology of behavior

The notion of affordance[30] has a long history. According to Don Norman, the word "affordance" was invented by the perceptual psychologist J. J. Gibson, to refer to the properties of the world that 'invite' actions, as for example a char invites one to sit. Originally, however, the notion of affordance dates back to the beginning of the 20th century, when it was used in phenomenologist philosophy to describe how the world around us presents itself as meaningful. Affordance, in other words, is a concept that explains why it seems natural for us to behave in a particular way, either because it is innate, as the reflex to close one's eyes by sudden exposure to light, or because we have learned that behavior, as for example surfing the web by clicking on links. In game or product design, thinking about 'affordance' may help us to find the most natural way to perform certain actions. Natural is in this context perhaps not the best phrase. What we must take into account is what is perceived as an affordance, and how actions fit in with what we may call an exology of behavior (with the system).

How does this diversion in abstract philosophy help us design better games? To answer this question, I'd like to recount my visit at the Virtual Humans Workshop, held in october 2004 at the Institute of Creative Technologies[31], in Los Angeles. Of interest in particular is the ICT Games Project:

ICT Games Project

> The goal of the ICT games project is to develop immersive, interactive, real time training simulations to help the Army create a new generation of decision-making and leadership-development tools.

As further explained on the website:*with the cooperation of the U.S. Army Research, Development and Engineering Command, Simulation Technology Center (RDECOM STC), Training and Doctrine Command (TRADOC), and commercial game development companies, ICT is creating two training simulations that are intended to have the same holding power and repeat value as mainstream entertainment software.*

The two training applications developed by ICT are:

- Mission Rehearsal Exercise – to solve a potential conflict after a car accident
- Language Training Simulation – to learn how to contact local leaders in arabic

[30] www.jnd.org/dn.mss/affordances-and-design.html
[31] www.ict.usc.edu

The *mission rehearsal exercise* is situated in former Yugoslavia. The trainee is confronted with the situation after a car accident in which a boy got injured. The mother of the boy is furious, and a potentially hostile crowd is waiting. An army convoy is on its way to a nearby airport, and needs to pass the crossing where the accident took place. The trainee must decide what to do and give appropriate orders, knowing that the wrong decision may lead to serious trouble.

The *language training simulation* is situated in the Middle-East, and is meant to teach the trainee not only some basic arabic but also proper ways of conduct, in conformance with local customs to gain confidence.

Both applications are highly realistic, with impressive graphics.[32] The both support speech input. The challenge in both simulation games was to come up with a natural way to indicate to the trainee what options for actions were available. Natural means, in this context, that it should fit within the simulation or game environment. Obviously, a menu or a row of pushbuttons does not fit naturally within such an environment, and would break what we have previously, in section 2.3 called 'immersion'.

I was invited at ICT for the Virtual Humans Workshop because of my involvement with embodied conversational agents (ECAs), as discussed in section 8.3. The topic of the workshop was, among others, to investigate whether the notion of affordance could help in analyzing and evaluating the interaction of a user/trainee with the simulation game. These were the questions we tackled:

Virtual Humans Workshop[33]

- Is it more appropriate to construct a frame of analysis that encompasses both user and ECA in a single interaction graph?
- Is it fitting to think in terms of a fixed graph that the user comes to recognize, or is the graph itself a dynamic structure?
- Is it even appropriate to focus on "affordances to act," or is it more fitting to consider cues that influence the mental interpretations that lead to action (e.g., affordances of control, affordances of valence of potential outcomes, etc.)? How does this relate to intrinsic motivation?

This workshop was a follow-up on a seminar in Dagstuhl on Evaluating Embodied Conversational Agents[34], where we discussed the tipics of interaction and affordance in a special interest group. In the *research directions*, I will further discuss an evaluation study that we did on agent-supported navigation in a virtual environment.

Back to our question, how can affordance help us in designing a game? In the *mission rehearsal exercise*, described above, it would be much more easy to have a menu with all available options listed. However, such amenu would defeat the purpose of the simulation, since such menus will not likely occur in real life. Immersion is, in other words, necessary to maintain the emotional involvement with the application, and affordance is the key to immersion. But, although it sounds like an answer, it does rather lead to another question, *how can we define the usability of a game*?

[32] A lot of attention has been devoted to creating the models and environments. Both simulations are implemented using the Unreal game engine.

[33] www.ict.usc.edu/∼vhumans/2004/

[34] wwwhome.cs.utwente.nl/∼zsofi//eeca

48

usability and fun

In interaction design there is a clear, yet unresolved, tension between usability and fun. Usability is, despite the many disputes, a well-defined notion:

usability (ISO DIS 9241-11)

> ... the effectiveness, efficiency and satisfaction with which specified users can achieve particular goals in particular environments ...

This is the ISO DIS 9241-11 definition, cited from Faulkner. In sectiona 10.3 we will further investigate usability as a means to evaluate systems from an interaction perspective. Now, I wish to focus on why artefacts or games might be appealing even if these same aspects may compromise usability in the traditional interpretation.

In describing a fancy juice squeezer, designed by Philip Starck Norman observes, following KS, that is:

emotional involvement

- entices by diverting attention – unlike the common
- delivers surprising novelty – not identifiable to its function
- goes beyond obvious needs and expectations – it becomes something else
- creates an instinctive response – curiosity and confusion

The phrase *satisfaction* in the definition of usability above seems somewhat meagre to explain the emotional involvement with games, and even inappropriate as one realizes that, in the *mission rehearsal exercise*, frustration might actually be beneficial for the learning experience.

49

example(s) – *visual sensations*

The dutch *visual sensations*[35] festival is an annual contest for VJs. In 2005, in cooperation with the festival, a parallel seminar seminar was held discussing the topic of the history of VJ-ing, a aplenary discussion of the relation between club-VJs and the established art circuit. In addition there were two guest speakers, Geert Mul and Micha Klein[36], both visual artists who also have a ten-years experience as VJ.

50

Above is another work of Geert Mul, in cooperation with DJ Speedy J. It was shown a dance event in cooperation with Rotterdam Maritime Museum. On the right, the cranes are swinging on the rhythm of the music.

The portfolio of Geert Mul[37] starts with a quote from Film:

form and content

> Very often people assume that "form" as a concept is the opposite of something called "content". This assumption implies that a poem or a musical piece or a film is like a jug. An external shape, the jug, contains something that could just as easily be held in a cup or pail. Under this assumption, form becomes less important than whatever it is presumed to contain.
>
> We do not accept this assumption. If form is the total system, which the viewer attributes to the film, there is no inside or outside. Every component functions as part of the overall pattern that is perceived. Thus we shall treat as formal elements many things that some people consider content. From our standpoint, subject matter and abstract ideas all enter into the total system of the artwork (....)

I totally agree with this. And perhaps this is why I have a preference for artworks that are slightly out of the main stream of tradional art.

research directions– *engaging with fictional characters*

What do you need to evaluate your game or multimedia application? There are many ways to gain insight in how your system is being used, see section 10.3. But if you want to establish functional properties of a multimedia application, for example the effectiveness of using an agent in navigating a virtual environment, in a scientifically more rigorous way, you need to have:

experimental validation

[35] www.visualsensations.nl

[36] www.michaklein.com

[37] e.mac.com/geertmul2

- a theory – in our case: PEFiC
- a test scenario – for example, memory tasks in a digital dossier
- the technology – to realize applications

In this section, I will briefly describe our efforts in experimentally validating the use of ECAs in virtual environments. As technology, we use our *intelligent multimedia technolgy*, described in sectiona 8.3 and appendix E. So what must be explained is the theory we adopt and the test scenarios we use.

PEFiC is a theory developed by Johan Hoorn and Elly Konijn, to explain *Perceiving and Experiencing Fictional Characters*, see PEFIC. The PEFiC theory may serve as the basis for the experimental evaluation of user responses to embodied agents. In summary, PEFiC distinguishes between three phases, encoding, comparison and response, in analyzing the user's behaviour towards an agent. Encoding involves positioning the agent (or fictitious character) on the dimensions of ethics (good vs bad), aesthetics (beauty vs ugliness) and epistemics (realistic vs unrealistic). Comparison entails establishing personal relevance and valence towards the agent. Response, finally, determines the tendency to approach or avoid the character, in other words involvement versus distance.

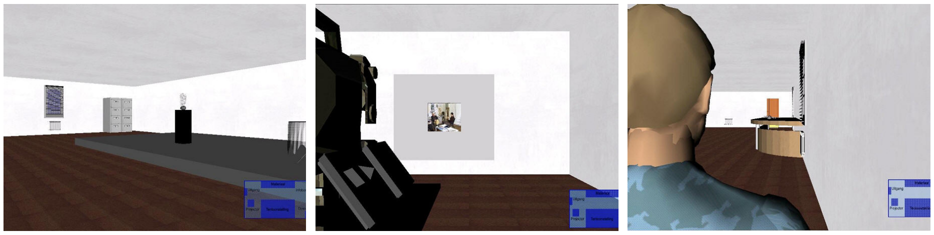
In general, having a virtual environment, there is, for developing test scenarios, a choice between:

validation scenario(s)

- navigation – pure interactivity
- guided tours – using some narrative structure
- agent-mediated – navigation and guided tours

For our application, a virtual environment of an artist's atelier, we have three experimental conditions, navigation without an agent, navigation with a realistic agent and navigation with a cartoon-like (unrealistic) agent. To ensure that these conditions can be compared, the acrtual information encountered when using the application is in all conditions the same.

The independent variable in our experiment, the degree of realism of the agent, corresponds with the epistemic and to some extent the aesthetic dimension of appraisal in the PEFiC theory. As dependent variables we have, among others, user satisfaction, believability, that is estimated usefulness of the agent, and also the extent to which the relevant information is retained.



51

The application is a *digital dossier* for the Dutch artist Marinus Boezem. The spatial metaphor we used for the dossier is the artist's atelier. We created a virtual environment containing a display of the artworks, in 3D, a file cabinet with textual information, a workbench for inspecting the artist's material, and a video projector, with which the user can display a video-recorded interview with the artist.

The actual task to be performed by the user is to learn what constraints do apply to the installation of one of the artworks, *Stone and Feather*:

Stone and Feather

- feather: 70 cm, from ostrich, curved
- stone: 13.5 cm, white marble
- position: alignment with pedestal, no glue
- environment: 50 lux of light max.

The items mentioned in this list must be reproduced by the user in a subsequent memory test, and in another experiment the user must be able to choose the right materials and reconstruct the artwork.

Our assumption in designing this test scenario was that the gestural nature of positioning the artwork will be favorable for the condition with a gesturing agent, whereas believability will be positively affected by the degree of realism of the agent.

10.2 digital dossier(s)

After a first round of the *multimedia casus*, in which the students produced an application giving an overview of the INCCA information archive, the participants, but only incidental information about the artists and their artworks, we decided to focus on case studies of individual artists, and we introduced the notion of *digital dossier*:

digital dossier

> *Create a VR that realizes a* digital dossier *for a work of a particular artist. A* digital dossier *represents the information that is available for a particular work of art, or a collection of works, of a particular artist. The* digital dossier *should be* multimedia-enhanced*, that is include photographs, audio and other multimedia material in a compelling manner.*

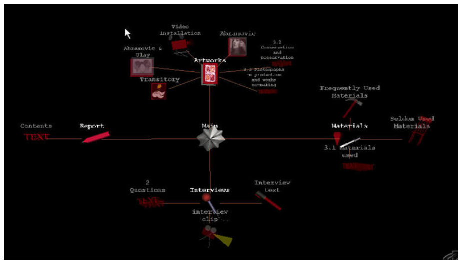
Like a medical dossier, the *digital dossier* was meant to give the information about the artist and the works of art readily at hand, so that it could effectively be used for the task of conservation and the re-installation of the artworks.

Since we were in doubt whether the phrase *dossier* actually existed in the English language, we looked it up in a dictionary:

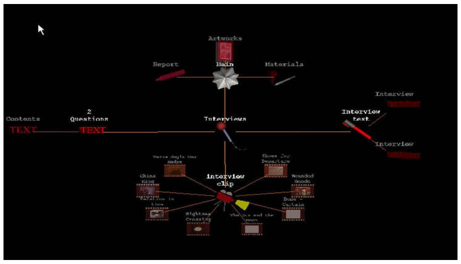
Webster New World Dictionary

- dossier (dos-si-er) [Fr < dos (back); so named because labeled on the back] a collection of documents concerning a particular person or matter
- archive – 1) a place where public records are kept ... 2) the records, material itself ...

We chose for the phrase *digital dossier*, and not for archive or library, to stress that our focus lies on presentational aspects. Although issues of data representation and content management are clearly important, our primary interest was with issues of presentation and navigation.



main node



interviews

52

the *abramovic dossier*

For the 2004 autumn group, we decided to take the work of Marina Abramovic, a serbian-dutch artist who became wellknown in the seventies with performances with her partner Ulay, and has since then produced numerous installations, videos and performances with what I would like to call 'high existential impact'. The directive with which the students where set to work was, quoting Ted Nelson:

> *everything must be highly intertwinkled*

Since virtual museums are by now a common phenomenon, and the virtual atelier for Marinus Boezem may be considered to be just a variant of this, the 2004 autumn group decided to explore alternative ways of presentation and navigation.

As material for the *abramovic dossier* there was an interview with Marina Abramovic from ICN, made in cooperation with the Dutch Foundation for the Visual Arts, and a great collection of videos from Montevideo. In addition, a transcription of the contents of the interview made by Michela Negrini, a student of media art at the University of Amsterdam, who also provided an interpretation as well as a categorization of the works of art. Given the material and the categories along which this material was classified, the students decided to explore the use of concept graphs as an instrument for navigating the information space.

navigation – *concept graphs*

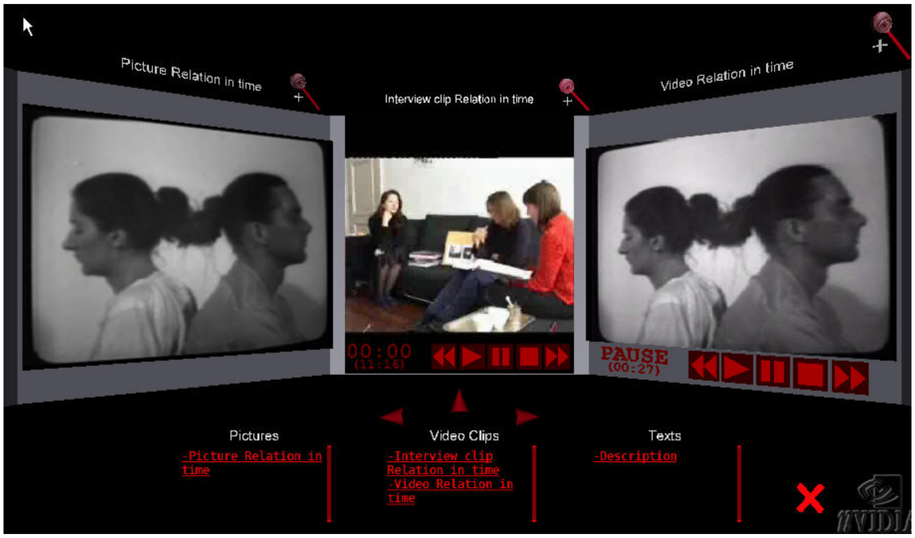
The reader has already encountered concept graphs in chapter 1, when the notions of multimedia, medium, television and communication were explained by indicating their relations to other concepts.

Concept-relation graphs are a familiar tool in linguistics and have also been used for a long time in Artificial Intelligence to describe the semantic relationships in complex domains. As a navigation instrument it is, to my knowledge only used in a kanji learning tool[38] and the Visual Thesaurus[39].[40]

[38] www.rikai.com/perl/KanjiMap.pl?

[39] ualthesaurus.com

[40] The Visual Thesaurus allows also for invoking Google image or document search from any of the elements of the concept graph.



presentation of video clips from Marina Abramovic

After the initial idea was there, one of the students of the group, Olaf van Zon, an AI student, managed to get a first version of a 3D concept graph working in VRML. This prototype implementation demonstrated the potential of the concept graph as a navigation instrument in the *abramovic dossier*.

presentation – *gadgets*

The original idea of presenting information, that is the actual interview, the videos and images of the works of art, as well as the textual information, was to use *rooms*, where the information could be projected on the walls. The *room* metaphor, which obviously stems from the virtual museum approach, did however not seem appropriate since it conflicted with the concept graph used for navigation. After some discussion, information rooms were abandoned in favor of *information gadgets*, that could be expanded from and collapsed into the concept graph.

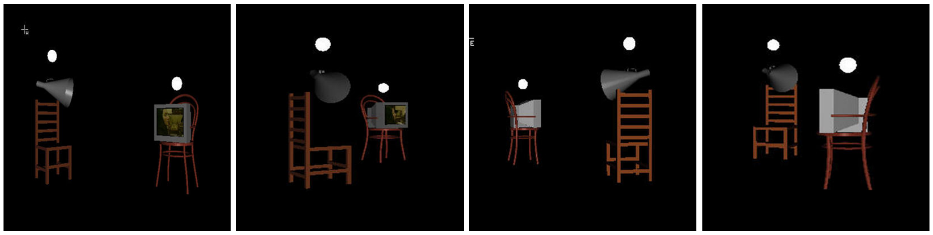
In the original *abramovic dossier*, the presentation gadget consists of three panes that can simultaneously show a video of the work, the interview, that is the fragment in which Abramovic speaks about that particular work, and the textual information related to the work and the interview. However, it appeared that in some cases there was not enough information, because the work was not spoken about in the interview, and in other cases there was too much information, for example multiple recordings or text documents. It was then decided to extend the presentation gadget with lists of alternative material that the user could select from and direct to one of the panes for further inspection.

To enable the user to focus on one of the panes, for example to get a better view of the video material a zoom in/out button was provided. All these enhancements, however, did complicate the interaction, as became clear when the *abramovic dossier* was presented at Montevideo.

In the course of the project, another interesting presentation feature was added, namely the reconstruction of one of the video installations in 3D, incidentally demonstrating the advantages of using 3D.

reconstruction – *recreating the installation*

In discussing the *abramovic dossier* with Bart Rutten from Montevideo, who provided us with all the video material, another project was mentioned which was concerned with 3D-recordings/models of existing installations. Having full confidence in the technical capabilities of my students, I promised to show that such a reconstruction of an installation would naturally fit within our approach.



Reconstruction of Terra della Dea Madre *in VRML.*

54

The installation for which the reconstruction was made is *Terra dea degli madre*, and installation with two chairs and a television, which was exhibited in the Stedelijk Museum of Amsterdam, in 1986. As a starting point, we took a video produced at the time of the exhibition, which shows the installation in an exposition room in the Stedelijk Museum, and which contains, apart from comments from Abramovic, also the video shown on the televison in the installation.

At this point, we can only speculate how useful such a reconstruction can be as a tool for the conservator responsible for the re-installation, to play around with the presentation parameters, the positioning in space, the overall size, light and ambient effects.

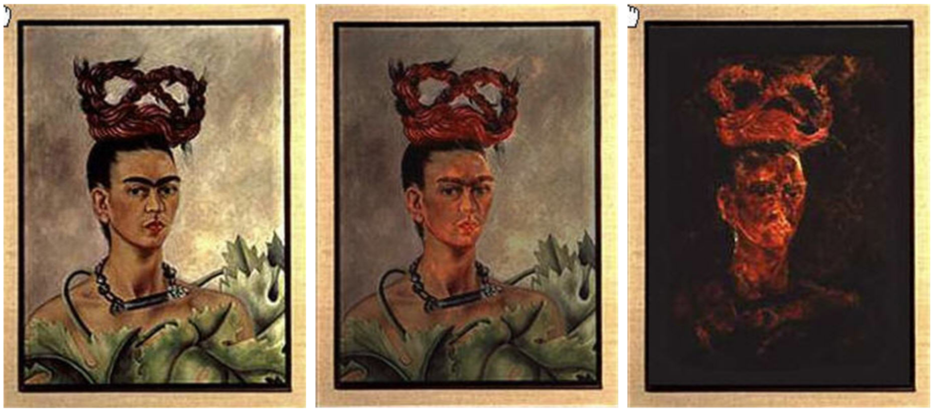
style issues – *how to improve the dossier*

The *abramovic dossier* does also provide a facility for search, as well as online help. However, as already mentioned, when demonstrating the application to the interested parties, that is ICN and Montevideo, a number of issues came along, that I will here summarize as a list of questions:

style issues

- what icons should be used to identify the elements of the concept graph?
- what categories and relationships are most appropriate?
- how should the information be displayed, simultaneously or more focussed?
- how do we allow the user to choose between multiple information items?
- how do we avoid visually disturbing elements?

Obviously, although the *abramovic dossier* was very positively received, these issues must be dealt with to make it a success. Having a first prototype, we need to rethink our application, not only with regard to its style of presentation, but as we will discuss in section 10.3, also in terms of its underlying data representation.



no light half light full light

55

example(s) – *conservator studio*

Ever thought of becoming a conservator? Seattle Artmuseum's Conservator Studio[41] gives you the opportunity to explore this career options:

> Explore four paintings from the Mexican Modernism exhibition through the eyes of a conservator (what's a conservator? you'll find that out too!). You'll have a new perspective on the paintings as well as how they are handled and prepared for display.

The illustrations above show what occurs when manipulating *transmitted light* on the painting *Self-Protrait with Braid*, oil on canvas, from the Mexican painter Frida Kahlo. As explained in the accompanying text: *when a light is shone through this painting one can see that the hair and the flesh areas are painted with thin layers of paint.*

These series of images are part of an interactive *flash* application developed by the Seattle Artmuseum to engage the general audience in the conservation of art, and to arouse an interest in art in general. The application allows the user to experiment with the various techniques used for the analysis and conservation of oil paintings.

research directions– *establishing usability*

In the March 2005 volume of CACM, an assessment is given of the current state of *user-centered design* practice. User-centered design is, quoting UCD, *a multi-disciplinary design approach based on an active involvement of users to improve the understanding of user and task-requirements, iterative design and evaluation.* In the article, which is based on a survey among user-centered design practitioners, user-centered design is claimed to

[41] www.seattleartmuseum.org/exhibit/interactives/mexicanModernism/enter.asp

have been beneficial for, among others, customer satisfaction and enhanced ease of use. Other measures mentioned are mostly relevant for e-commerce applications, which, as the authors observe, *have greatly bolstered the the appeal of usability and user-centered design, as users can take their business elsewhere with just one mouse click*.

In our case, the competition is fortunately less threatening. Nevertheless, usability issues such as legibility of text, ease in navigation and adequate task support are equally relevant. As a first step after completing the *abramovic dossier*, we have developed a test-plan and a sample task, and (the students) executed two test-sessions with participants from ICN and Montevideo, who where asked to work with the system thinking aloud. The test-sessions were recorded on video, and the participants were requested to complete a questionnaire.

In UCD, a list of approaches is given, which were reported to have been used by the respondents of the survey:

user-centered design methods

> field studies, user requirement analysis, iterative design, usability evaluation, task analysis, focus groups, formal/heuristic analysis, user interviews, prototype (without user testing), surveys, informal expert review, card sorting, participatory design

The three most frequently used methods in this list are, respectively, iterative design, usability evaluation and task analysis. These three methods were also considered to be important by the respondents. Frequently used, but not considered to be as important, were informal expert reviews. And less frequently used, but considered important, were field studies. This distinction can, according to UCD, attributed to cost-benefit trade-offs, since clearly field studies are much more costly.

Usability evaluation looks, according to Preece to issues such as:

usability evaluation

- *learnability* – time and effort to reach level of performance
- *throughput* – the amount of work done
- *flexibility* – accomodating changes in the task
- *attitude* – of users to the system

To conclude this section, let's take a closer look at task analysis.

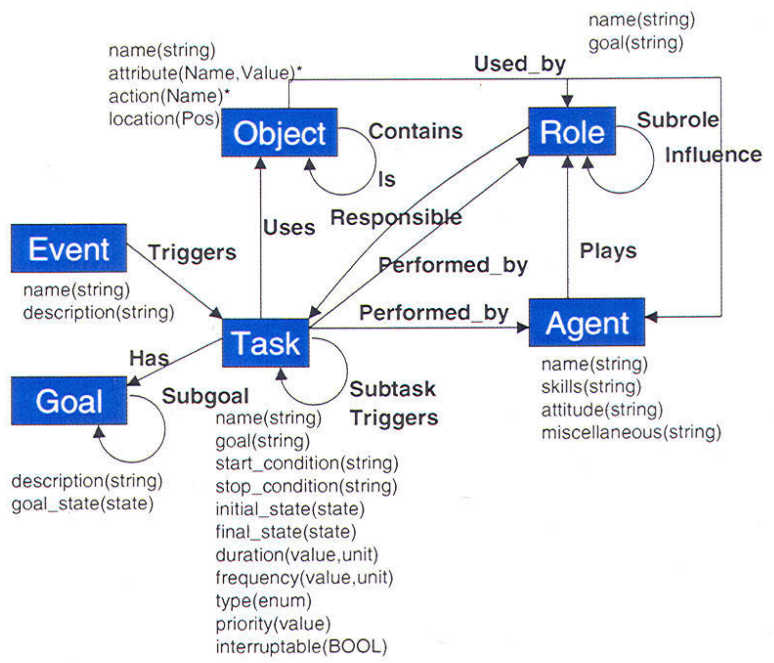
task analysis Task analysis may be characterized as the decomposition of a task into subtasks or steps, to arrive at a sufficiently detailed description of the task and its relation to the environment.

In Euterpe, a description is given of what might be understood as the task world ontology, the concepts and relations that play a role in performing a task analysis. The main concepts figuring in the task world ontology are, following Euterpe:

task world ontology

- *task* – activity performed by an agent to reach a certain goal
- *goal* – a desired state in the task world or system
- *role* – a meaningful collection of tasks
- *object* – refers to a physical or non-physical entity
- *agent* – an entity that is considered active
- *event* – a change in the state of the task world

As indicated in the diagram above, these concepts are related in various ways. Example relations include *uses*, *triggers*, *plays*, *performed_by*, *has*, etcetera.



56

Creating a task model based on this, or a similar, ontology may help us understand what a user needs to accomplish and how this may be supported by an information system. As such, creating a task model should be considered to be an essential ingredient of the software engineering life cycle, OO.

11.1 constructing a game

Since ancient times, games have been an essential part of culture. According to Huizinga, game playing is not merely meant for entertainment, but is (one way or another) fundamental for our survival:

> ... *in the game we confront a function of the living creature which cannot be determined either biologically or logically* ...

As a terminological aside, in the english language there is a distinction between *play* and *game*. The difference seems to be primarily that a game is subject to additional rules, which determine the success in the game, whereas play seems to be less constrained and more phantasy-driven. In addition, games seem to have a competitive element, a challenge, where you may either win or lose.

Anyhow, as observed in Semiotics:

visual culture

> *games are an increasingly important element in our visual culture.*

Nowadays, with rapid advances in technology, the game industry is said to be overtaking both the film industry as well as the software industry, in terms of total budget. This situation has a somewhat paradoxical effect on game development, however. On the one hand, it is more easy to develop a game, considering the great variety of tools and technologies that are available. On the other hand, to make a successful game has become more difficult, since the competition is much more harsh and does not allow for amateurism. Nevertheless, game development is worthwhile in itself, even if it does not lead to economic success. As a computer science or multimedia student, you should at least be familiar with the process of game development, and gain experience with (a selection of) available game technology.

game engine component(s) Game development covers a wide range of activities, and although preferably done in a team, requires a correspondingly wide set of skills, artistic as well as technical. When speaking about *game programming* it is important to make a distinction between:

- game play programming
- game engine programming

Game play programming, usually, consists of scripting the behavior of characters or elements of a game scene, using a scripting language such as, for example, Lua[42]. It may considered to be part of game design, and certainly requires an artistic approach.

In contrast, game engine programming is a much more technical activity and concerns the development of functional components of a *game engine*, which according to Ultimate encompass:

game engine component(s)

- rendering system – 2D/3D graphics
- input system – user interaction
- sound system – ambient and re-active
- physics system – for the blockbusters
- animation system – motion of objects and characters
- artificial intelligence system – for real challenge(s)

Although it is possible to build one's own game engine using OpenGL or DirectX, or the XNA[43] framework built on top of (managed) DirectX, in most cases it is more profitable to use an existing game engine or 3D environment framework, since it provides the developer with a load of already built-in functionality. In section 4.4, a brief comparative overview of game engine(s), that may be used to built virtual worlds, has been given. This overview did not include the flex 2 SDK, nor related flash engines, that may be used for the development of online (flash) games.

[42] www.lua.org

[43] msdn.microsoft.com/directx/XNA

57

elements of game design Game development projects are similar to multimedia projects, or for that matter software development projects, in that they may be subdivided in phases covering the definition of requirements, the actual development, as well as testing and delivery. However, instead of looking at these phases, or trying to explicate the differences between software development and game development projects, it seems to be more worthwhile to look at what characterizes game developement, and, as a consequence, what must be the focus of game design.

Eventhough the question *what is a good game?* is rather elusive, it is evident that a game must be fun to play! When creating a game, we may according to GameDesign capture fun in two areas:

fun

- in the general flow of the game experience and
- in the individual moments during a playing session.

which, together, determine the player's unique experience in playing the game.

As a first working definition, GameDesign states:

> *a game is a series of processes that takes a player to a result.*

which covers both the skill aspect of game playing, as well as the fact that a player generally must make a series of decisions. Refining this initial definition, GameDesign proposes to characterize a game as follows:

interactive electronic game

> A game is a play activity comprised of a series of actions and decisions, constrained by rules and the game world, moving towards an end condition. The rules and the game world are delivered by electronic media and controlled by a digital program.
>
> The rules and game world exist to create interesting situations to challenge and oppose the player. The player's actions, his decisions, excitement, and chances, really, his journey, all comprise the "soul of play".
>
> It is the richness of context, the challenge, excitement, and fun of a player's journey, and not simply the attainment of the end condition that determines the success of the game.

Although this definition could certainly be improved in terms of conciseness and clarity, it does cover the main characteristics of game playing, in particular *context* and *challenge(s).*

varoom thung blam

58

With respect to challenge(s), we may observe that these may take the form of *occasional battle(s)*, as we have called them in the PANORAMA application, discussed in section 5.4. A battle may be characterized by the following conditions:

battle condition(s)

- confrontation on well-established area
- delimited in space/time
- audience/participants who judge victory/loss

It is in this context interesting to note that being mentioned in the *hall of fame*, that is a list of players with the highest scores, seems to be an intrinsic motivation for players to frequently return to the game, either to acquire or maintain a position on that list.

The actual development of a game is a complex process, and, in our experience, for each game unique, since there are many forces influencing a project. Nevertheless, whether you work on your own or in a team, inevitably the following aspects or roles have to be taken in consideration:

design team role(s)

- manager(s) – keep everything together
- producer(s) – maintaining focus
- programmer(s) – solve problem(s)
- tester(s) – control quality
- designer(s) – elaborate idea(s)

In our experience, for example in developing VU Life, that we will discuss in section 11.2, one person may take up several roles, but at some point it becomes clear what the strength(s), and for that matter weaknesses, of each individual participating in the project are.

history of game(s) There are many overviews of the hostory of (video) games, for example at wikipedia[44]. One overview that takes an interesting technological perspective is worth mentioning. Algorithms distinhuishes between eight phases, where each phase is characterized by one or more unique technical development(s):

history

1. phase i: before space war – hardwired
2. phase ii: spacewar on atari – console with game

[44] en.wikipedia.org/wiki/History_of_video_games

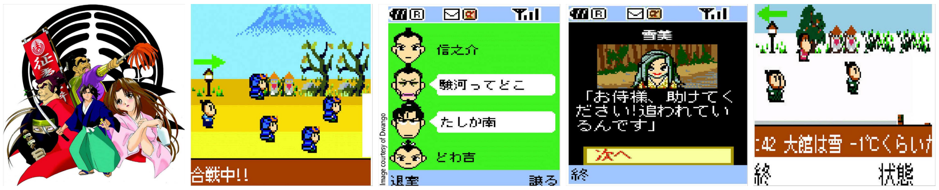
3. phase iii: game console and PC – separate game development
4. phase iv: shakedown and consolidation – player code in data files
5. phase v: advent of the game engine – user level design
6. phase vi: the handheld revolution – the GameBoy
7. phase vii: the cellular phenomenon – larger installed user base
8. phase viii: multiplayer games – from MUD to Everquest

For those interested in more details, Algorithms provides as characteristic examples for each phase the following (additional) data: 1) tennis for two William Higinbotham Brookhaven National Labs New York, 1950' 2) Steve Russell, 1961, MIT, spacewar, two player game on Digital PDP-1 3) Atari VCS, Apple II, IBM PC (Dos) 4) Donkeykong, Pacman -> Nintendo 5) Doom -> Valve Halflife 6) Gameboy with well-established collection of game 7) NTT Docomo I-Mode, Samurai Romanesque 8) MUD (1979), MULE (1983), Ultima/Everquest 1600 hours/year.

59

Technical progress does not stop here. Two developments worth mentioning are the rise of Second Life as a 3D immersive interface to the web, and the recent availability of a new 3D engine for flex 2, PaperVision3D[45], both of which may influence the choice of (online) game platform(s) for the near future. Another development that we may expect for the near future is according to Algorithms: *a truly cinematic gaming experience.*

[45] papervision3d.org



駿河ってどこ

たしか南

「お侍様、助けてください!追われているんです」

:42 大館は雪 -1℃くらいだ

Screens from Samurai Romanesque.

60

example(s) – *samurai romanesque*

Samurai Romanesque, available on Japan's NIT DoCoMo packet-switched i-mode network, is an example of a mobile game with a large following. This massive multi-player game is developed by the japanese game developer Dwango. It runs on the Java 2 platform Micro Edition (J2ME). Players take, as we read in Samurai a virtual journey through 15-th century Japan, engage other players in real-time battles, visit historical towns and villages, practice the art of Zen, engage in romances and even can have children. This massive multiplayer role-playing game can accomodate up to half a million players, and is accounted to be a huge success in Japan. *Samurai Romanesque* is an example of a mobile game incorporating features such as position awareness, player history, chatting, and effective graphics. In Samurai, it is further explained how the technology with which the game is implemented positions itself in the *battle for mobile cyberspace*.

research direction(s)– *serious games*

Serious games are here to stay. Currently there is for example already a great offer of business management games. When googling on *game*, *business*, and *management*, we find, among many other offerings games to train leadership[46] (which provides urgent problem situations in a variety of areas, including military and health care applications), and entrepreneurship[47] (which provides a eight round cycle of sessions instruction how to start a business, get clients, etc., with extensive feedback in the form of reports and comments after each round). A general observation we may make here is, however, that the games we have seen so far primarily focus on functionality and offer at best an efficient interface, which we do not find very appealing from a more artistic perspective.

There are many (more) resources on serious games[48]. To indicate what it is all about, we present a quote from virtual heroes[49]:

> *Serious games and simulations are poised for a second revolution. Today's children, our workforce and scientists are increasingly playing, learning, and inventing in visually intensive "virtual" environments. In our increasingly experiential economy, immersive educational and training solutions are needed to advance the workforce of tomorrow. Game-based learning and technologies meet this challenge.*

[46] www.experiencepoint.com
[47] www.marketplace-simulation.com
[48] www.cs.vu.nl/∼eliens/media/resource-serious.html
[49] www.virtualheroes.com

However, regardless of the fuss being made, apart from the euphorics their is little attention to determine in a more scientific way what the actual value of the game is in the learning process, and what elements or aspects of the game contribute to the learning experience. To provide such a foundation, we will propose a game reference model in section 12.1, that we have applied to formulate criteria for effective service management games in Serious.

There is a wide choice of technology available for the realization of serious games. For example, in the *climate game* project, we did explore various technologies, including interactive video with flash, as well as the use of the HalfLife2 game engine, with which we gained experience in developing a promotional game for our faculty, VULife. With regard to the use of 3D we may remark that since ancient times a walk in space has served as a mnemonic device, and as such spatial memory may aid in retention and understanding, which might also provide a decisive argument for the use of 3D in aa serious game, such as a service management game!

As explained in section 3.4, we found great inspiration for Clima Futura, our climate game, in *Peacemaker*[50], that provided us with an example of how to translate a serious issue into a turn-based game

From an interview with the makers:

peace maker(s)[51]

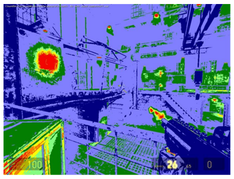
> Q: With the lion's share of strategy games on the market being devoted to ending a conflict through violence, why was it important to you to emphasize the need for a peaceful solution?
>
> A: When we started to work on the project and looked around at other video games, we encountered the notion that war is much more challenging and conflict is essential to engage players. Many people we talked to perceived peacemaking as mere negotiations, where a group of diplomats sit at a table for lengthy discussions and sign agreements. We tried to shed light on what we see as the other side of peacemaking how challenging it is for a leader to gain trust and understanding in the face of constant violence. How difficult it is to execute concessions, while your own population is under stress or feeling despair. In a sense, peacemaking can be more complicated, sophisticated and rewarding than war making, and it is a message that we would like to convey to young adults, the future generation of leaders.

In summary, Peacemaker covers both political and social issues, with appealing visuals, not sacrificing the seriousness of the topic. By presenting real-time events using video and (short) text, awareness is created by allowing a choice between the points of view of the various parties involved. Such awareness may lead to political action and will no doubt influence choices, also when it comes to issues of climate change. Essentially, serious games aim at attitude change, the rest follows automatically ...

black/white

greyscale

sepia


thermal

[50] www.peacemakergame.com

[51] seriousgamessource.com/features/feature_071806_peacemaker.php

61

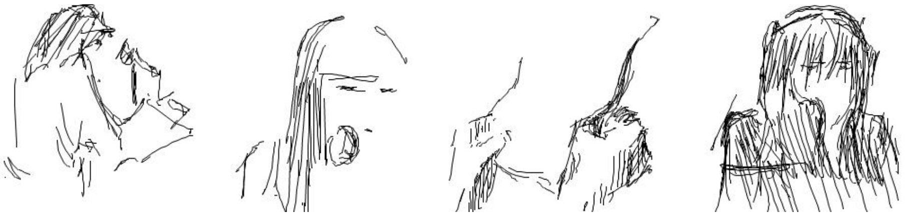
12.1 a game model

Games present challenges, invoke involvement, and are essentially interactive. Although it might seem far-fetched to regard game playing as a paradigm for interaction, it is definitely worthwhile to have a closer look at *game theory* for inspiration. According to HalfReal, from a theoretical perspective games may be said to have the follwoing properties:

game theory[52]

- system – (formal) set of rules
- relation – between player and game (affectionate)
- context – negotiable relation with 'real world'

In particular, *relation(s)* and *context* determine the meaning of the game for the player, both on an individual/existential level and in relation to a societal context.



62

To characterize the defining characteristics of games in a more precise way, HalfReal presents a classic game model that may act as a reference for the description and evaluation of games:

classic game (reference) model

- *rules* – formal system
- *outcome* – variable and quantifiable
- *value* – different valorisation assignments
- *effort* – in order to influence the outcome
- *attachment* – emotionally attached to outcome (hooked)
- *consequences* – optional and negotiable (profit?)

For current day video games, HalfReal observes that there is a tension between rules and the fictional or narrative component of the game:

rules vs fiction

> game fiction is ambiguous, optional and imagined by the player in uncontrollable and unpredictable ways, but the emphasis on fictional worlds may be the strongest innovation of the video game.

[52] www.half-real.net

In some cases it might even not be clear what the *rules of the game* are, as for example in Second Life, where *presence* and *expecience* seem to be prevalent. In general, role playing games seem to be less constrained than skill-based games. Nevertheless, in both cases does the visual environment augment the experience, adding to the narrative context.

63

So, returning to our original question:

theory of interaction

> are *games* relevant for a theory of interaction?

our tentative answer is yes!

In an attempt to formulate criteria for effective service management games, developed in cooperation with Getronics-PinkRoccade, Serious, we gave a characterization in terms of the reference game model, as outlined below:

effective service management game(s)

- *rules* – service management protocols
- *outcome* – learning process
- *value* – intellectual satisfaction
- *effort* – study procedures
- *attachment* – corporate identity
- *consequences* – job qualification

There is no need to emphasize that this is only a first approximation, and for that matter a rough one. What we must keep in mind, however, is that the model is not only applicable on a macro-level, to characterize an entire game, but more importantly may also be applied on a micro-level, to establish the criteria for each (relevant) step in the game play. To emphasize the relevance particular aspects of service management games, we added two more criteria to the model:

- *scenario(s)* - problem solving in service management
- *reward(s)* - service level agreements

After all, the goal of playing a service management game is to be trained in, as stated above, problem solving in service management situations, and reaching acceptable service level agreement(s)!

game (interaction) design pattern(s)

Game play is an experience that requires active involvement of the player and may stand as a model for interaction, that is interaction with high emotional load. Types or patterns of interaction that may occur in game playing are analysed in GamePatterns, which characterizes *game play* as:

game play

> ... structure of interaction with game system and other player(s)

GamePatterns explicitly avoid giving a definition of either game(s) or game play, but instead focus on *interaction patterns*, that is how players interact with the game system (and other players) to effect a change in the state of the game.

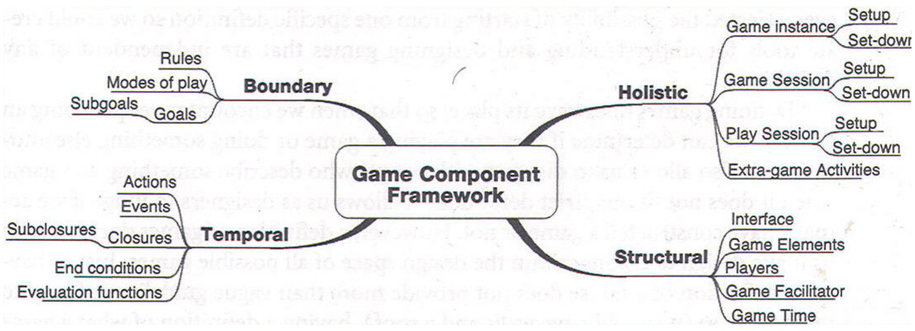
For the description of interaction patterns, GamePatterns introduce a component framework, that distinguishes between the following types of components:

component framework[53]

- holistic – *playing games as an undividable activity*
- boundary – *limit the activities of people playing games*
- temporal – *describe the flow of the game (interaction)*
- structural – *physical and logical elements of the game system*

The various components, each affect game playing in a particular manner, either by the way the game presents itself through an interface (structural component), or by the rules that determine the game (boundary component).

An overview of the aspects or elements that belong to each component is given in the figure, taken from GamePatterns, below.



64

For narrative structure(s), that we will discuss in the next section, obviously the *temporal component* is most important, containing elements such as *closures* as well as *actions* and *events* that may occur during game play.

Referring to GamePatterns for details, an impression of what (types of) interaction patterns may exist is given in the following list:

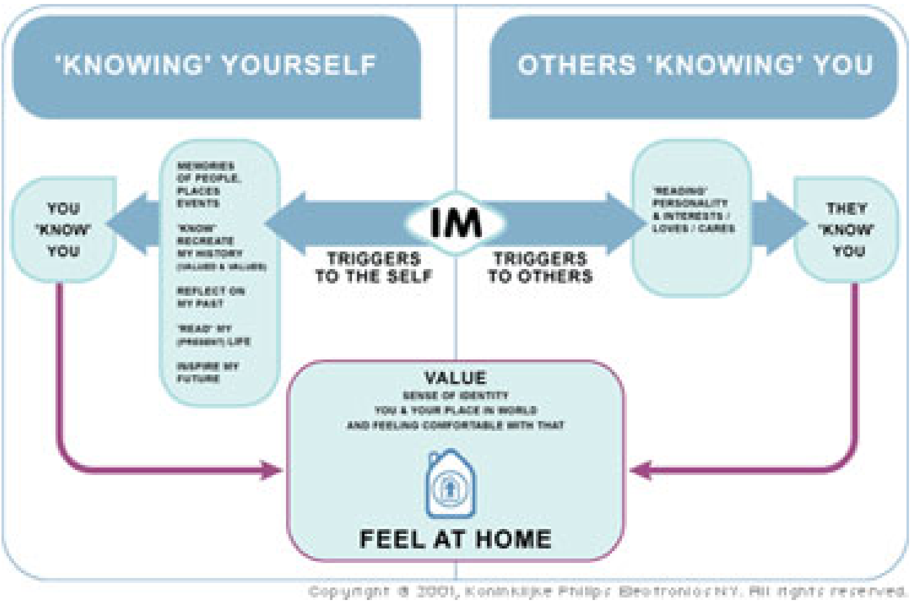
pattern(s)

- *resource management* – resource types, control, progress
- *communication and presentation* – information, indicators

[53] www.gamedesignpatterns.org

- *actions and events* – control, rewards and penalties
- *narrative structures and immersion* – evaluation, control, characters
- *social interaction* – competition, collaboration, activities
- *mastery and balancing* – planning, tradeoffs
- *meta games and learning* – replayability, learning curve(s)

For example, with respect to actions and events that may occur during game play, there are various ways rewards and punishments may be dealt with. Also, as we mentioned in section 10.4 when discussing *interaction markers*, there exists a variety of patterns by which to present information and indicate (opportunities for) interaction.


'KNOWING' YOURSELF
OTHERS 'KNOWING' YOU
YOU 'KNOW' YOU
MEMORIES OF PEOPLE, PLACES EVENTS
'KNOW' RECREATE MY HISTORY (VALUED & VALUES)
REFLECT ON MY PAST
'READ' MY (PRESENT) LIFE
INSPIRE MY FUTURE
IM
TRIGGERS TO THE SELF
TRIGGERS TO OTHERS
'READING' PERSONALITY & INTERESTS / LOVES / CARES
THEY 'KNOW' YOU
VALUE
SENSE OF IDENTITY
YOU & YOUR PLACE IN WORLD
AND FEELING COMFORTABLE WITH THAT
FEEL AT HOME
Copyright © 2001, Koninklijke Philips Electronics N.V. All rights reserved.

65

example(s) – *intimate media*

From the company that used the slogan "let's make things better", and now advertises its products with "sense and simplicity", there is the MIME[54] project, not to be confused with the multipart internet mail standard, which focusses on *Multiple Intimate Media Environments.*

As concepts embodying their ideas they propose, among other:

intimate media object(s)

1. *glow tags* – a subtle way to trigger the person who has placed it or who sees it
2. *living scrap book* – to capture and collect information and media digitally
3. *picture ball* – as an object of decoration and a focus for storytelling

[54] www.design.philips.com/about/design/section-13484

4. *lonely planet listener* – enabling people to listen to a real time connection to another place

On a more abstract level, seven core qualities are identified which *capture the essence of the intimate media experience*:

intimate media experience(s)

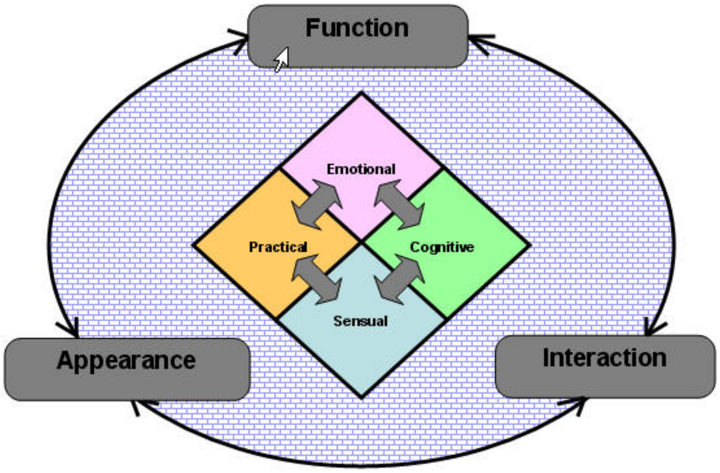
- sensorial – *experience is visual, audible, tactile, olfaric*
- personalized – *objects embody meaning and memories*
- analogue – *people relate to physical objects*
- enhancement – *people already have extensive intimate media collections*
- serendipity – *it supports unstructured and flexible usage*
- longevity – *objects may exist over generations*

As can be read on their website: *intimate media describes the things that people create and collect to store and share their personal memories, interests and loves.* And: *intimate media is central to how people make sense of their world by representing roots, heritage and a sense of belonging, achievement and connection.*

research directions– *experience as meaning*

For the design and appreciation of the new category of digital systems, including games, we may, with (forward) reference to our discussion of the history of thought in section 12.4, well take *pragmatist aesthetics* as a common ground, since it does justice to the existential dimension of aesthetic awareness, and allows for a process of aesthetic literacy, that is becoming sensible to aesthetic awareness and reflection. You may wonder though, how we get to this conclusion.

In Presence it is observed that *the aesthetic potential of the narrative space centered on the consumer product has received surprisingly little attention.* The authors then argue that, motivated by insights from phenomenology, there should be a shift of attention from *use* to *presence*, where presence does not merely mean appearance but a more complex dialectic process of appearance and gradual disappearance dependent on the role the object plays in the life of the user/subject. The notion of *expressional* is then introduced, to convey the expressive meaning of objects, and in particular interactive objects, in our surroundings. For the *design of presence*, *aesthetics* is then considered as a *logic of expressions*, in which *expressions* act as *the presentation of a structure in a given space of design variables.*



66

So far, this makes perfect sense. We may further mention here the framework set up by Dhaval Vyas, Meaning, which characterizes the user's experience a the result of the process of constructing meaning. In diagrammatic form, the process of constructing meaning may be depicted as above.

In more detail, the validity of the *experienc as meaning* framework may be substantiated by the foloowing observations:

experience as meaning

- experience occurs during the interaction between the user(s) and the interactive system(s) in the lived environment
- designers convey meaning (consciously or unconsciously) through the appearance, interaction and function of the system
- user(s) construct a coherent whole that is a combination of sensual, cognitive, emotional and practical forms of experience

In other words, an *interactive system* is determined by *function*, *interaction* and *appearance*. As such the framework may be called *pragmatist*, and has indeed been nfluenced by Pragmatics.

Returning to our argumentation, for objects that are not designed for usability in the functional sense the notion of *use* is too strict and is, using a dialectic argument, subject to the dialectics of *presence*, as argued in Presence. Conversely, using a similar dialectic argument, for new categories of objects, *presence* requires *use*, or getting used to, in other words a process in which the user becomes interested and familiar with the object. We may even speak of *aesthetic affordance*, with the realization that the notion of *affordance*, explaining an ecology of behavior, originally stems from the late-idealist phenomenology expounded in Sein.

But, however appealing the notion of *expressional*, in the light of our discussion in section 12.4, where we distinguish between aesthetic awareness as a given, or a priori, sensibility and aesthetic judgement as being of a more empirical nature, we would prefer to consider *aesthetics* as a *logic of sensibility*, which includes a dimension of self-reflection in the sense of its being aware of its own history. Put differently, to characterize the contextual aspect of aesthetics, as it certainly applies to art, we may speak of *aesthetic literacy*, that is aesthetic awareness that is self-reflective by nature.