43rd International Scientific Colloquium
Technical University of Ilmenau
September 21-24, 1998

M. Mühlhäuser

# Content Development for the Internet as the Fourth Mass Medium

**ABASTRACT:** The Internet is about to become the fourth mass medium - besides Press, Radio, and TV. Personalized news and Video-on-Demand are just initial content types: like the other media, the Internet will bring forth specific genres. In the first part of this article, important characteristics of these future genres are worked out; keywords for these characteristics are: *interactive, personalized,* and *evolutionary.*
It is shown that content creation for the new medium involves software engineering to a large extent. As a consequence, part two of the article relates the above characteristics to major requirements for authoring/development tools and methods. The list of requirements is discussed in detail; it covers a range of aspects such as real-time support, unification of HCI and software engineering, tight integration of browsing support, contents with variable "challenge", etc.
In part three, the claimed characteristics and requirements are backed with experiences from experimental developments of new contents, corresponding software engineering aids, and exploration tools for the new mass medium. Several such example developments were carried out during the last two years in the division headed by the author and were exposed to a large public. Three of them will be discussed in more detail: the *ComicActor* approach to customizable visual agents, the *VideoBrowser* approach to at-a-glance visual description of video, and the *WorldBeat* system for instantaneous, group-wide creation of aesthetic music by novices.
**KEYWORDS:** Internet, multimedia software engineering, mass media, convergence.

## 1. INTRODUCTION

**CONVERGENCE:** The so-called convergence of information technology, telecommunications, and the media businesses (including consumer electronics) is based on the Internet as the single common infrastructure and has a strong impact on each of the three merging large markets [9, 8]. As to the *media businesses,* the Internet is becoming the fourth mass medium, complementing Press, Radio, and TV. While the media-related aspect of convergence may be the slowest one to come about, it may soon become the most revolutionary and far-reaching one.
**MEDIAMORPHOSIS:** in early times of the cinema, recorded theater performances served as cinema contents; in the early days of television, the key TV contents were in turn just cinema movies. The Internet, too, takes off as a mass medium by offering contents known from the other media (this explains why so many believe that Video-on-Demand - VoD - will be a big business for the Internet). The phenomenon is called mediamorphosis in this paper: new media start with old contents and gradually evolve towards new genres which are more appropriate and specific. As mediamorphosis will happen in the Internet, too, VoD is probably *not* going to be an important Internet genre in the long run. So far, so good. But what will the new content types, or genres (used synonymously), look like? And what are the consequences for content development? It would be too ambitious to predict exact details about the new Internet genres; they will evolve in an iterative process driven and limited by the creativity of content creators (from the media business), the taste and budget of the public (with new generations having different taste), and the performances of science and technology (limiting the technical feasibility). This process was emphasized in the initial definition of the term mediamorphosis by Roger Fidler: "The

transformation of a communication medium, usually brought about by the complex interplay of perceived needs, competitive and political pressures, and social and technological innovation." [3].

**OVERVIEW:** rather than predicting the future, we will shed some light on major characteristics of new genres. They are based on experiences gained over the last two years with the development of ambitious exhibits for the Ars Electronica Center AEC, a museum of the future and media laboratory that was set up in the heart of the city of Linz, Upper Austria. The author's division was chartered to set up an entire floor of the AEC, dedicated to the influence of computing, networks, and media on our daily private and work life and on the way we learn. In chapter two, we will provide an overview of the characteristics of new genres. We will then look at several examples which might be interpreted as prototypical new Internet genres: we will reflect these genres in the light of the characteristics described in chapter three, and discuss issues of content development – which will to a large extent turn out to be issues of software engineering.

## 2. THREE CHARACTERISTICS OF NEW INTERNET GENRES

**INTERACTIVE:** the Internet will be the first mass medium to enable truly interactive contents [2]. There have been long debates about whether or not "couch potatoes" i.e. typical TV spectators of today would ever be interested in interactive contents. Many have presumed that interactivity would be regarded as too pretentious to attract the masses. Yet while contents for the masses will probably never be pretentious, they may well become interactive, at least to a certain degree. Even couch-potatoes tend to use their remote control for zapping through channels, although the contents of a single channel may make much more sense than the odd mixture of channel shreds which results from random zapping. And the upcoming "Nintendo™ generation" feeds a huge industry as the kids try to satisfy their (interactive) play instinct.

In brief, the danger of interactive content is that it might be perceived as too pretentious, but the advantage lays in giving the user the feeling of being "in control of" or at least "involved in" the course of action. In addition, interaction is a prerequisite for personalized contents, see below.

As to truly interactive contents, we dare to go a step further and claim that successful contents may even foster creativity. Fostering creativity, again, does not seem to be a desirable feature of new contents at a first glance: which genres of which media may ever have fostered creativity? The problem here is the inability of known mass media to provide the interactivity, let alone the simplicity required to make Mr. Average become creative in a way that leads to satisfying results. The common connotations of creativity are artistic skills and intensive learning – painting, performing music, molding, and many other creative tasks take a long time to learn and require skills that many believe not to have. We have shown in our projects that the Internet is the first mass medium which is powerful enough to support creativity-enabling contents for the masses. Up to now, all mass media had some sort of documents as contents. With the Internet, contents may be software. Such software may – just as an example – enable the "consumer" to create paintings or make music of acceptable aesthetic value with almost no learning effort, cf. the WorldBeat example in chapter 3.

The interactivity and creativity of new Internet genres, along with the lack of computing expertise of both potential content developers and potential consumers, impose heavy requirements on three different disciplines of computer science as follows.

1. **Software Engineering:** first of all, interactive or even creative contents may not be mere documents, but *must* be software. This fact boosts the importance of software engineering for content development, a trend which could already be observed with respect to Hollywood productions of recent years (Jurassic Park being maybe the most spectacular one). In contrast to Hollywood movies, however, Internet contents will usually have to be produced by much smaller teams or even by individuals, with much lower budget and for a much smaller customer base. As a consequence, there is a high demand for software engineering concepts which foster cost-effective production: component-based programming and adequate design methods are just examples. While these statements appear to be self-evident or trivial, one has to note that content creation for mass media has been an authoring task for many decades, not a software engineering task; and the effort for putting a media professional in the position to "write programs" (in the largest sense) might be compared to the task of making an average person capable of creating print-ready documents – a task that was considered to be unachievable for a long time: then, the advancements made with the legendary TeX system and the WYSIWIG aspects added by early text processors started the era of desktop publishing.
2. **Human Computer Interaction:** Due to the importance of interactivity and to the challenges involved in designing interfaces for the masses, software engineering must be complemented by adequate aids from the human-computer interaction (HCI) discipline. An important challenge here is the interleaving of the two domains in terms of their concepts, processes, methods and tools: for years, software engineering and HCI lived largely independent from one another; it was not until 1994 that the two communities organized a first joint workshop [11]. To cite an example for the gap between the two disciplines, HCI design traditionally deals with the composite of technical and human components while software design usually "ends" at the *interface* towards the human part. One of the consequences of this "stopping at the interface" is that design methods for software dynamics are not well advanced: the consecutive flow of action usually goes back and forth between humans and technical systems (software) and involves mutual dependencies. Another example is a lack of separation between interaction-related and core functionality in software. Changes in the user interface, or even reflections of the projected mode of interaction (e.g., the terminal devices and modalities used) often require radical redesign of the interaction part; due to the lack of separation, such redesigns presently imply considerable changes in the core functionality.
3. **Real-Time Computing:** real-time behavior is particularly required for contents which foster creativity. The negative effects of rather "batch" than real-time applications could be observed a few years ago: tools existed for creating simple computer artwork based, e.g., on Mandelbrot sets; they would ask the user to provide some cryptic input values (the effects of which could usually not even be perceived) and would then perform calculation and drawing. Unsatisfying results would require the user to try a new turn with new parameters. This "batch mode" of creativity takes away from the user most of the feeling of being in control of the creative process. In contrast to this example, users should be able to immediately sense the effects of their creative act. The challenges here are manifold. For graphics-related tasks, the power of CPUs and graphics cards may still have to be increased until true real-time creative contents can be brought to the masses over the Internet. For the creation of time-dependent media, special hardware will still have to be developed. Most important, appropriate input devices are needed in order to reflect high ergonomic challenges (an aspect which is of course closely related to point 2 – HCI – above) under real-time constraints (cf. "WorldBeat" in chapter 3).

**PERSONALIZED:** Internet news tools like Point-Cast Network™ (PCN) and tools based on the upcoming publish/subscribe metaphor (such as Microsoft Active Channels™) are first proofs that the Internet is much better suited for delivering personalized contents than any other mass medium [4]. The reasons are obvious: the Internet is a true two-way network, it supports point-to-point and multicast communication in addition to broadcast, and it supports arbitrary software as contents and as part of the equipment. Today, personalization is based on channel subscription or on personal profiles; both provide only rough descriptions of user preferences. As the mediamorphosis continues, the evolution will lead from canned uniform contents not only to personalized but otherwise traditional contents (which, by the way, will go well beyond news) but also to new personalized genres. Three important requirements can help to advance personalization.

1. **Adequate challenge:** arguing about creative or distracting interactive contents earlier in the paper, it was pointed out that software-based contents can provide the ease-of-use needed to make average users become creative. Yet each individual has of course unique skills and prerequisites. Overcharging the user will make him or her abandon a content in frustration; undercharging on the other hand means boring the user, again a reason for abandoning a content. Computer games usually meet this requirement by offering a fixed number of levels of challenge; obviously, more flexible adaptation is desirable. For other kinds of interactive contents, like creative ones, even less experience exists about adapting the level of challenge (for a sample solution, cf. the WorldBeat system below).
2. **Data taming:** as stated above, channel subscription and user-provided profiles are just the beginning of personalizing contents. Much more elaborate user modeling and data filtering approaches are investigated in many research projects [5], often based on machine learning concepts. But even the best approaches for reducing the information flood will by themselves not be sufficient for taming the exponential growth of the amount of information offered on the market – with the Internet being the fastest-growing such market. To make things worse, multimedia data are up to now difficult to index and filter since their user-level semantics are largely hidden (see below). It can therefore be predicted that even after narrowing down the information as much as possible using filters or queries, the user will be faced with a large amount of multimedia data that cannot be further reduced. Visual or auditory clues are needed here for a quick at-a-glance impression about contents (cf. VideoBrowsers below).
3. **Cooperative computing:** since the Internet is not just a broadcast network like radio and TV and since it is the common infrastructure for all aspects of convergence (see chapter 1), the "mass medium" aspect and the "computer-integrated communication" aspect of the Internet are likely to converge in the context of genres. Thereby, personalization is not restricted to an individual but to a group. As an example, an electronic bulletin board operated by a special interest group (SIG) today would hardly be classified as a mass media genre; but the merger of a SIG bulletin board and of a SIG letter – maybe jointly edited by members of the SIG – may in the future evolve towards such a genre.

**EVOLUTIONARY:** What was called mediamorphosis in the introduction can be generalized into the evolutionary nature of the upcoming mass medium. While the contents as such will have to evolve gradually from known to new genres, the new software will have to relate back to well-known concepts of our present world. This requirement is further detailed below.

4. **Intuitive device:** keyboard and mouse will surely not be appropriate for the general customer base of a mass medium; more appropriate devices will come up and be evolutionary in the sense that their look and feel will be closely related to common non computer devices (e.g., for tactile input such as drawing, a pen-like tool might be appropriate) or communication act (this applies mainly to non tactile input such as speech or gesture).
5. **Intuitive metaphor:** for user interfaces, too, evolution does not mean "evolving from present computers" but "...from present day-to-day tasks". An interface metaphor which resembles a kitchen is much more evolutionary than one which resembles a computer windowing system.
6. **Familiar contents:** seeking for new genres, can we require the contents to be familiar? Yes and no. For one, the requirements stated in this paper need not be fulfilled altogether in all cases; rather, new genres might be evolutionary in just one of the three senses mentioned here, e.g., by offering new devices but old contents. Secondly, new genres might be new in terms of the "how", not always the "what". An evolutionary strategy may for instance be the use of familiar contents in new interactive "clothes" (cf. below: music styles applied in WorldBeat).

**THE MULTIMEDIA SEMANTICS PROBLEM:** A final requirement is orthogonal to all three characteristics of new Internet genres mentioned: whether content types shall be interactive, personalized, or evolutionary, it is always important for the system to understand the meaning of



the data for the user.

One has to recall that multimedia data today are usually represented in a digital form which is closely related to the physical signals perceived by human senses. E.g., digital video and graphics data today represent parameters of the light waves which reach the human eye (luminance and chrominance values, etc.); audio data represent the waveform of the acoustic signal reaching the ear. To make things worth, other aspects of the representation are unfortunate inheritances from the analog media, such as constant sampling rates which do not reflect the event density (cf. spinning wheels in movies which seem to spin in the wrong direction, a result of frame rates which are too low for that particular scene). Briefly spoken, low-level statistical models are used today for representing multimedia.

A myriad of researchers worldwide try to find methods for analyzing these "low-level" representations of (multi)media data in order to extract semantics which are relevant to the human *mind*. This may be called the move from statistical to semantic models. E.g., wordspotting techniques are used to identify keywords in sound waves; using object recognition and video indexing techniques, one tries to "make sense" of video data [10]; gesture and speech recognition, object tracking and many other examples could be mentioned. But while enormous progress has been made in recent years, the results are still not satisfying for many use cases. To summarize, the research activities for building semantic models from statistical-model data are undertaken with enormous efforts but still modest results.

It must be pointed out here that there is a second, hardly recognized pathway from statistical to semantic models of multimedia. This second pathway could be based on the fact that most of the semantic information is usually present when multimedia data are *captured* or *created*. E.g., when a movie is shot, the movie director usually has information at hand (script, storyboard, scenario etc.) which describes the major objects and actors on the scene, the shooting perspective, the camera movements, and more. Part of that information is about to become digitally available anyway (like many other disciplines, movie production becomes computer-supported, a trend which leads to digital on-line storyboards etc.); other parts of that information may be easily captured using appropriate devices (camera movements might be easily captured with sensors, geographical data might be registered using GPS devices, etc.). Considering these facts, it might be a much better idea to capture the semantics of multimedia data at their origin instead of trying to extract them from the (representation of) physical signals way after they have been created - as described in the previous paragraph. Briefly spoken, the second pathway seems to be much less exploited but much more feasible (except for legacy data such as analog movie archives etc.).

In order to store high-level semantics, appropriate models are needed. In the above example, an appropriate movie model would be one that reflects actors and objects, camera movements, time and location of the shooting, course of action and dialogues, etc.

As to multimedia data which are instantaneously created by the users of new genres, the creation process must also be based on an appropriate model. We will discuss this issue in the context of the WorldBeat example in chapter 3.

All in one, present models and representations of multimedia data must move from a level which is close to the physical signals to a level which is close to the representation in human mind (as far as we understand it), but also close to the information which determines the creation process – be it for contents which are "produced" like, e.g., Hollywood movies, be it for contents which are created interactively by the "consumer".

## 3. SAMPLE CONTENTS, TOOLS, DEVELOPMENT AIDS

**WORLDBEAT:** As an example of how different future genres for the Internet might be from what we know today, we mention the so-called WorldBeat exhibit which is used by hundreds of visitors daily since about one and a half years at the AEC Linz (cf. overview in chapter 1). WorldBeat copes with interactive performance of music; fig. 1a shows a picture of a visitor who is about to interactively create music. The system is described in more detail in [1], we will

restrict the discussion to the aspects relevant in the Internet / mass media context. The entire exhibit is controlled via just two wireless batons, tracked by a device which translates two-dimensional baton positions into MIDI data. In one of the WorldBeat modules, the user can conduct a piece by using a baton like a conductor of an orchestra. Many other modules exist which will not be mentioned further, except for one. With this module, the user can play together with a blues band. Using one baton as mouse replacement (pretty much like a light pointer), the user parametrizes the band online until the groove, tempo, bass etc. reflect the user's preferences. S/he then selects a solo instrument and starts to use the batons like the sticks of a vibraphone – just that the vibraphone is virtual i.e. "in the air" in front of the user (and just that the vibraphone sounds like any solo instrument selected). Due to the sophisticated multimedia semantic model of WorldBeat, the system can adjust the keys played so that they fit aesthetically with the accompaniment: the user can not play wrong!! In order to adjust baton strokes in real time to notes which are in harmony with the band but still reflect the speed, velocity and position of the user's strokes (as part of a user-controlled improvised melody), a multimedia semantics model is used which - in this case - is based on a notion of aesthetic music, including aspects of rhythm, harmony, and melody.

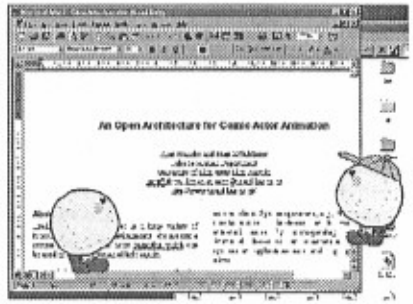

**Figure 1. a) the WorldBeat Exhibit; b) two ComicActors atop a word processor (right)**

Currently, the system is augmented towards a true Internet genre with a new *NetMusic* module: several players – which may be distributed around the world – can play together according to the "can't play wrong" approach just described. Since the latency of audio signals travelling around the world can never be lower than the mere physical signal propagation time (not counting any store-and-forward delays), it is physically impossible to play together worldwide the way one would play together locally (with every player listening and performing at once). Tracks are therefore broadcast in store-and-forward mode so that other players can add tracks by performing in real time but listening to a recording (which might be just a few milliseconds old). Obviously, the system reflects the following requirements from chapter two:

- *Software Engineering:* a novel concept of *musical design patterns* has been developed in WorldBeat, reflecting re-usabilty of software for creative composition and performance.
- *HCI, intuitive metaphors and devices:* light pointer, conductor's baton, and vibraphone stick are all intuitive metaphors - and they are all an identical physical device!
- *Real-Time:* music is created in real time as a piece is conducted or a virtual vibraphone is hit.
- *Adequate challenge:* absolute novice users who never played an instrument can indeed be creative (i.e. perform aesthetic music immediately); for more experienced users, system support can be gradually removed (introducing the risk of playing aesthetically "wrong" notes).



- *Cooperation support:* the NetMusic module can be used by several players almost synchronously. New concepts for playing together (cooperation!) had to be developed, see above.
- *Familiar contents:* common musical styles are applied (other advanced musical systems use uncommon "artistic" styles, often imposed by the limitations of the semantic model applied).
- *Multimedia semantics:* music is the first kind of multimedia data for which a higher-level model was developed: the MIDI standard, based on the concept of musical notes. Our model in turn lays way above MIDI, based on concepts of aesthetic music and supporting real-time.

**COMIC ACTORS:** Fig. 1b shows a use case for novel Internet contents. ComicActors are cartoon-like actors which perform animated actions on the screen. Each ComicActor comes with a repertoire of possible actions like walking, pointing, talking, selecting menus, pressing buttons, dragging windows and objects on the screen, triggering programs and the like. The uniqueness of ComicActors lies in the following combination of features.
The amount of artwork required for creating a ComicActor has to be carried out only once; thereby, the designer is guided in order for the drawing process to yield highly customizable building blocks. The result is packaged into a ComicActorType which can be easily parameterized by a content creator in order to create a particular ComicActor: s/he can determine, e.g., the walking trajectories, the phrases to speak, the GUI objects to manipulate, and more (most of these customizations are input via simple mouse clicks). The clothing, objects to carry along, and cooperation acts between several ComicActors are other examples of choices offered to the content creator. Many decisions and computations are necessary in the background for providing correct actions and smooth transitions. E.g., if the content creator orders a displacement action followed by a point-and-talk action, the system may have to determine the appropriate displacement (walk, fly, hop, fall, ...; left, right, slanted, in zigzag, ...) according to the trajectory, insert a "turn" action in order to smoothly transit from walking to pointing, combine "point" and "talk" building blocks, and perform further computations. Fig. 1b shows two ComicActors of the same ComicActorType who meet atop a word processing system.
ComicActors may be used as help agents, parts of multimedia presentations, visualizations of Internet agents, or avatars (representatives of absent users). For an elaborate description see [6]. ComicActor based visualizations are potential new genres, and ComicActorTypes reduce the creation effort by orders of magnitude. The following requirements from chapter 2 are met:

- *Software Engineering:* the enormous effort for creating an animated cartoon is "shrunk" to a few mouse clicks and parametrization efforts.
- *HCI and familiar metaphor:* animated characters can greatly improve HCI systems; they are familiar enough to be quickly perceived and "abstract" enough to avoid "humanization".
- *Real-Time:* ComicActorss offer unique dynamics: not only humans but arbitrary *software* may specify and order a ComicActor sequence at runtime, to be displayed almost instantaneously.
- *Cooperation Support:* since ComicActors can cooperate, they can be used for visualizing cooperative activities, even between humans and software components.

**VIDEO BROWSER:** VideoBrowsers provide at-a-glance overviews of video contents. Our system (cf. [7]) selects key frames (representative pictures) from a video and arranges them in a hierarchical way (the most representative ones in the root of a tree, zoom-in groups of frames as lower nodes). Any node i.e. set of frames in the tree may be displayed in a single window called OBVI. An OBVI is a VRML 3D model of a card filer of key frames, subject to 3D manipulation. (stretch & squeeze, isolate objects, make front frames transparent, ...). Due to this approach, the frames - although stuck one behind the other - may be easily inspected at a glance.
Specially adopted video indexing techniques are used which do not only detect scene cuts but measure degrees of dynamics. The following requirements from chapter 2 are met:

- *HCI and familiar metaphor:* the card filer metaphor is intuitive, at the same time saving real-

estate and providing an idea about the course of action.

- *Data taming:* the VideoBrowser described provides an at-a-glance view on hours of video and supports zoom-in on scenes with a few mouse clicks.

## 4. Conclusion

Three major characteristics of future genres for the mass medium "Internet" have been identified, and ten requirements for the corresponding content creation support (which turned out to be a software engineering problem) were derived. These requirements were discussed in the light of three sample developments carried out at the author's group. The future will have to show how new Internet genres will actually look like. The sample genres, tools and development aids discussed could only slightly "lift the curtain". It was enough, however, for showing that the ten requirements stated were actually important, that content creation for the Internet is indeed a matter of software engineering, not only a matter of authoring as formerly, and that the software engineering challenges are considerable and by far not resolved yet.

## 5. References


[1] J. Borchers, and M. Mühlhäuser: Musical Design Patterns: An Example of a Human-Centered Model of Interactive Multimedia, *Proc. IEEE Intl. Conf. on Multimedia Computing and Systems, Ottawa, June 1997*, IEEE CS Press, Los Alamitos, CA, 1997, pp. 63-70.

[2] R. Comerford: Interactive media: an Internet reality, *IEEE Spectrum*, April 1996, pp. 29-32.

[3] R. Fidler: *Mediamorphosis, Understanding New Media*, Pine Forge Press, CA, 1997.

[4] R. Hackathorn: Publish or Perish, *Byte 22(9)*, September 1997, p. 65.

[5] M. Hearst and H. Hirsh (Eds): *Proc. AAAI Spring Symposium on Machine Learning in Information Access*, AAAI, Menlo Parc, CA, 1996.

[6] K. Manske and M. Mühlhäuser: An Open Architecture for Comic Actor Animation, *Proc. 5th ACM Multimedia Conf., Seattle, WA, Nov. 9-13, 1997*, ACM Press, 1997, pp. 251-259.

[7] K. Manske, M. Mühlhäuser, S. Vogl, M. Goldberg: OBVI: Hierarchical 3D Video Browsing, *Proc. 6th ACM Multimedia Conf., Bristol, UK, Sep. 12-16, 1998*, ACM Press, 1998.

[8] D.G. Messerschmitt: The convergence of communications and computing: What are the implications today?, *IEEE Proceedings*, August 1996.

[9] New World: Virtual Venues - Bringing Content to the Media Market, *The Journal of Business Strategy 16 (5)*, Sept. 1995.

[10] Panchanatan, S. and B. Liu (Eds.): Special Issue on Indexing, Storage, Browsing, and Retrieval of Images and Video, *Journal of Visual Communication and Image Representation 7 (4)*, Dec. 1996.

[11] Taylor, R., and J. Coutaz (Eds.): *Software Engineering and Human-Computer Interaction, Proc. ICSE '94 Workshop on SE-HCI: Joint Research Issues, Sorrento, Italy, May 16-17, 1994*, Springer LNCS 896, 1994.