The new technology for multimedia

Autor(en): **Schroeter, Philippe / Burkhard, Thomas / Herrmann, Beat**

Objekttyp: **Article**

Zeitschrift: **Comtec : Informations- und Telekommunikationstechnologie = information and telecommunication technology**

Band (Jahr): **77 (1999)**

Heft 5

PDF erstellt am: **17.09.2024**

Persistenter Link: https://doi.org/10.5169/seals-877021

MPEG-4:

The new Technology for Multimedia

Following the trends of the information society, the Moving Picture Experts Group (MPEG) has recently proposed MPEG-4, the new standard for coding and integration of audio-visual objects. This standard explores all possibilities of digital environment, including the coding of natural audio and video as well as computer generated objects, such as animated graphics, 3-dimensional virtual reality, and sounds. The end user has now the freedom to interact with the scene by deleting, adding, or repositioning objects. Properties of objects can also be modified by a simple mouse click. MPEG-4 has been carefully designed to scale to different transmission platforms such as wireless or IP. It means that the content needs to be coded only once and can be automatically played out at different rates with acceptable quality for the communication environment at hand.

EP2 Section

EP9702 is an Exploration Programme active on the residential market segment. Its main goal is to identify the new technical orientations and business opportunities for this market segment. To achieve this goal, the program focuses on 3 areas: the multimedia technology (MPEG,DVB), the new home network area and the access network for supporting these services. EP9702 uses various working methods, from scenario development to market analysis as well as laboratory demonstrators. EP9702 is also active in most important standardisation bodies and many international projects: ACTS TERA, ACTS ITUNET, ETSI BRAN, EURESCOM BOBAN, EURESCOM HINE, FSAN.

The original scope of MPEG-4 was the development of very-low bit rate coding algorithms, targeting applications such as video-conference. Anticipating the rapid convergence of telecommunication, computer and broadcast industries, the MPEG group widened this scope to meet the challenges of future Multimedia applications and related environments. In particular, MPEG-4 addresses the need for

- universal accessibility and robustness in error prone environments: awareness of the network peculiarities, from mobile to fixed networks;
- high interactive functionality: true user interaction with the content;
- coding of natural and synthetic data: new codecs for audio and video coding;
- compression efficiency: good quality of the reconstructed data at given bitrates;
- decoder downloadability: PC-based software for decoding audio and video data.

In the actual trend of convergence, MPEG-4 gives telcos new opportunities to shift business from network provider to service provider. It is also the first standard that includes nearly all kinds of digital media, from synthetic audio to natural video. Moreover, it really meets the actual customers expectation. Following Swisscom CIT-CT long expertise in the multimedia domain (for example [CT877-2], [CT850-3], [CT-1212-3], [CT-1273-1], and [CT-1274-1]), experts from exploration program 2 give, in the paragraphs below, an insight of the new MPEG-4 functionality and related impacts on the telecommunication world.

MPEG success Story

MPEG was established in January 1988 with the mandate to develop standards for coded representation of moving pictures, audio and their combination. It operates in the framework of the Joint ISO/IEC Technical Committee (JTC 1) on Information Technology and is formally WG11 of SC29. A large part of the MPEG membership is made of individuals operating in research and academia. MPEG first activities resulted in MPEG-1, issued in 1992. This standard was designed for coding progressive video (typi-

PHILIPPE SCHROETER, THOMAS BURKHARD, BEAT HERRMANN, DANIEL LEDERMANN, DENIS SCHLAUSS, BERN

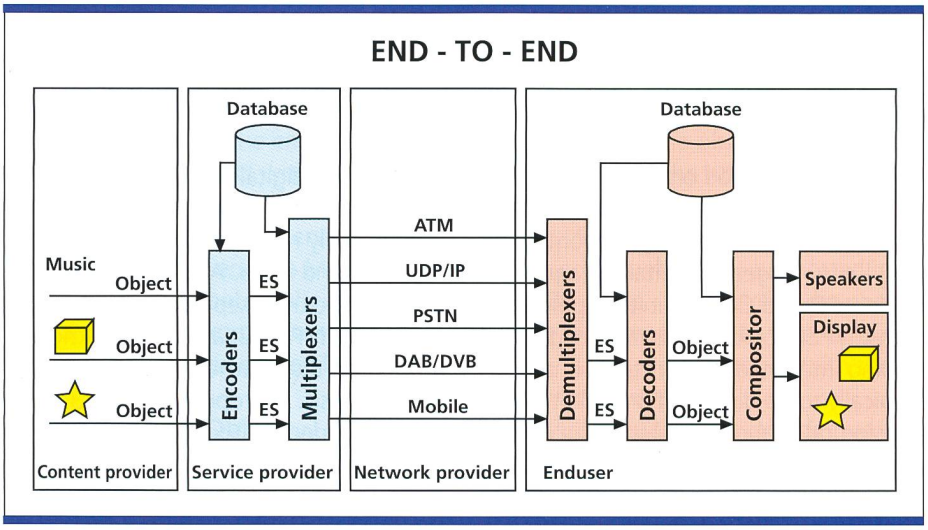

Fig. 1. End-to-end connection, conceptual diagram.

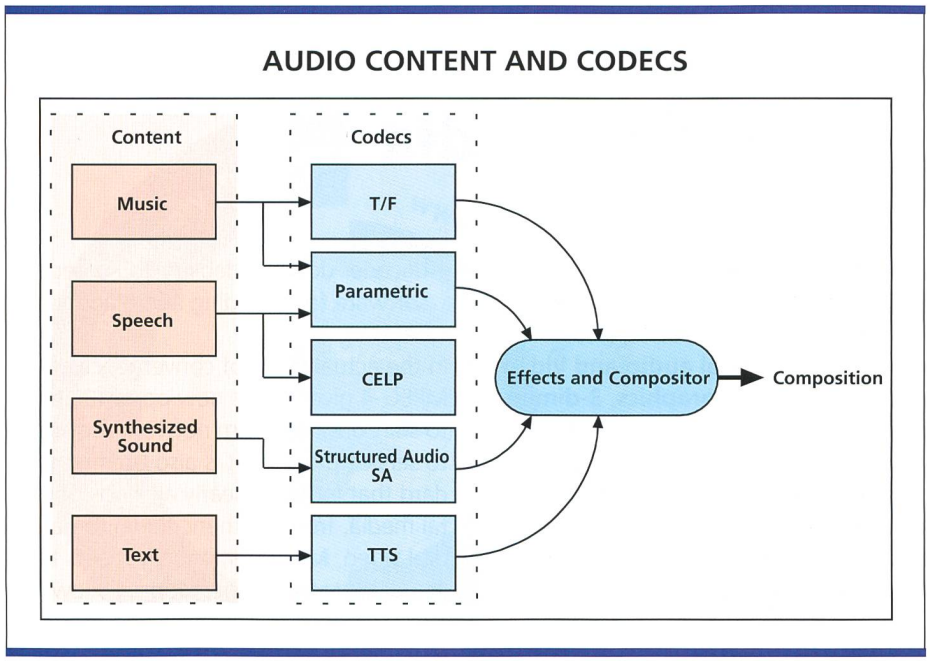

Fig. 2. Four different types of MPEG-4 audio content can be encoded with five different types of codecs. Effects can be added to each of these encoded objects.

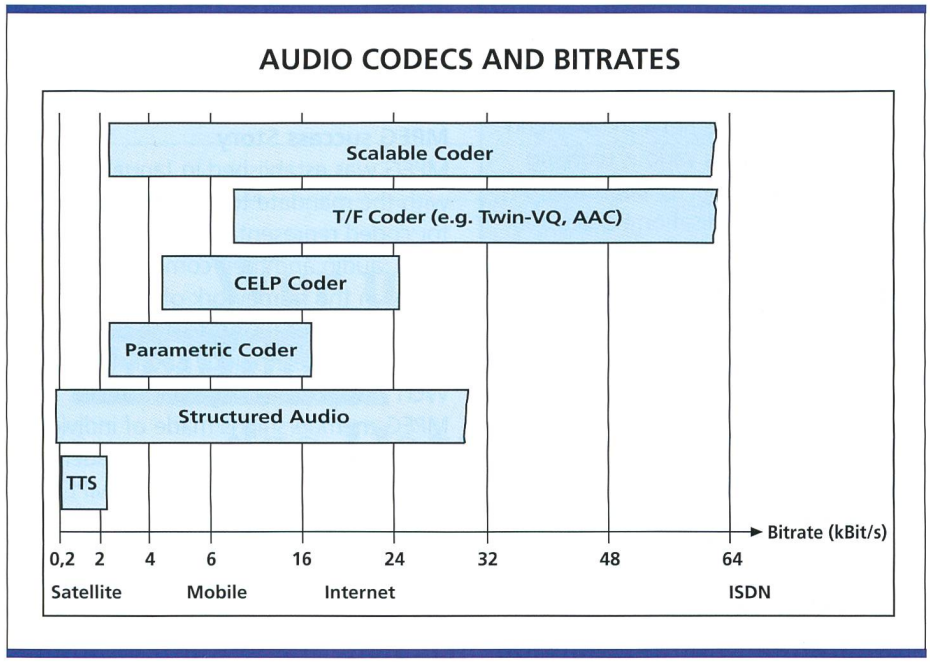

Fig. 3. The MPEG-4 audio codecs and the corresponding bitrate per channel.

cally to be displayed on a PC screen) at a transmission rate of about 1.5 Mbit/s and audio at bitrates ranging from 32 to 448 kbit/s, targeting digital storage (Video-CD and CD-I) as initial applications. MPEG-2 video was designed for coding interlaced sequences of images (typically to be displayed on TV) at transmission rates above 4 Mbit/s. The audio part of MPEG-2 includes multi-channel coding (typically allowing multi-language support, e.g. german, english or surrand sound) and lower sampling rates. MPEG-2 is used for digital video and audio broadcast (DVB, DAB) and DVD, and will progressively replace the traditional analogue TV and radio. Audio compression tools can be used separately. A typical example is MPEG-2 layer III, best known as mp3 on the Internet, for the compression of music. A proposed MPEG-3 standard, intended for High Definition TV (HDTV), was merged with the MPEG-2 standard when it became apparent that the MPEG-2 standard met the HDTV requirements. MPEG-4 is the latest International Standard produced by MPEG and is described in the following sections. The next step, MPEG-7 will standardize a way to describe various types of multimedia information. This description will be associated with the content itself, to allow fast and efficient searching for material that a user may be interested in. These types of information include: still pictures, graphics, audio, video, and information about how these elements are combined in a multimedia presentation.

Audiovisual Objects

MPEG-4 is not just an additional compression algorithm with improved performance characteristics. Although its initial goal was very low bitrate coding, the scope of MPEG-4 evolved to meet the new needs emanating from the convergence of the PC, broadcasting and telecommunication worlds. The answer to this convergence is the object-based representation of audiovisual scenes. Audio and video components of MPEG-4 are known as objects, e.g. a speech, a person, a 3-dimensional graphic. These can exist independently, or multiple ones can be grouped together to form higher-level audiovisual bonds, e.g. a talking person. One strength of this object-oriented approach is that the audio and video can be easily manipulated. Objects in a scene are described mathematically and are given a position in a two- or three- dimensional space (either for audio or video). Scene modifications can be done by users interaction.

Another advantage of the object-oriented approach is that it enables scalable content, also called layered coding. This adds a new dimension to the existing scalability as defined in MPEG-1 and MPEG-2, i.e. the content scales to the available bandwidth at the cost of quality. With layered coding, a decoder can render lower quality content by means of basic layers, and higher quality if the decoding of additional layer is possible (as a function of the network and PC resources). The following kinds of scalability are possible:

- Bit rate scalability allows a bitstream to be parsed into a bitstream of lower bit rate that can still be decoded into a meaningful signal. The bit stream parsing can occur either during transmission or in the decoder.
- Encoder complexity scalability allows encoders of different complexity to generate valid and meaningful bitstreams.

– Decoder complexity scalability allows a given bitstream to be decoded by decoders of different levels of complexity.

Other advantages of the object-based representation can be illustrated by the following examples:

– MPEG-4 coded objects can be coded only once and stored in multimedia databases. Thus, one object can be used interchangeably on many different scenes.
– Objects can be stored locally on the user PC. Thus, on a slow link, only the most "meaningful" objects can be transmitted while the "background" can be generated locally on the PC, or retrieved from the local database.
– All types of objects, synthetic and natural, audio or visual can co-exist in a scene.

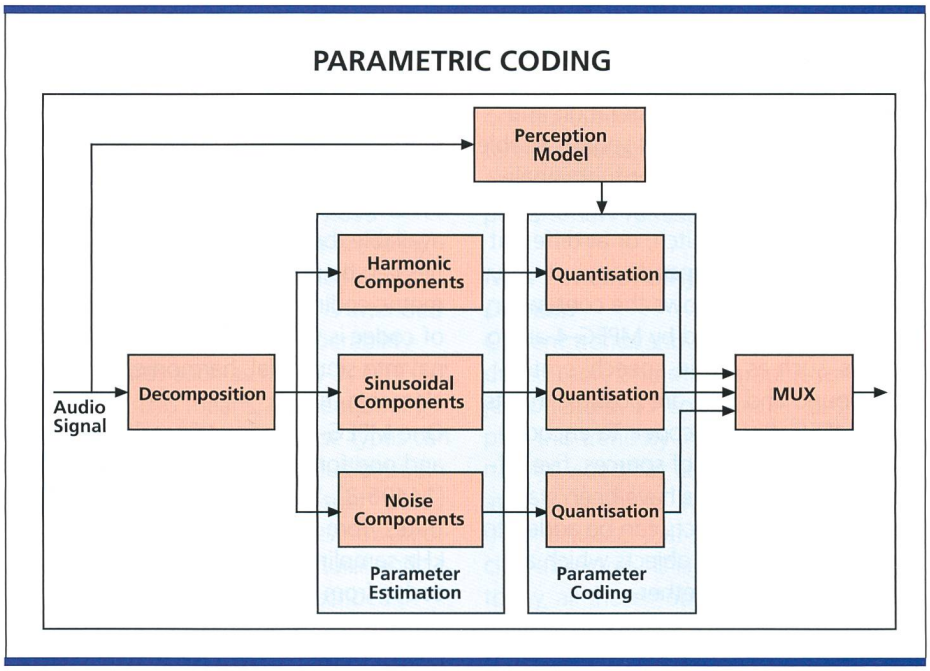

Fig. 4. MPEG-4 Parametric codec block diagram.

The Way to the End-User Screen

The audiovisual objects and compositions have to find their way from the content provider site to the end user. Figure 1 represents a conceptual diagram of an end-to-end connection.

Content provider: In the context of MPEG-4 content providers should more likely be called «object providers». They will provide multimedia objects (Objects in fig. 1) such as music, 2D visual objects, graphics, 3D synthetic objects, speech, etc. How to build these objects is not part of the MPEG-4 standard. However, in order to align to the MPEG-4 needs, media production studios will need to think object-oriented. The TV weather forecast is a typical example. A camera records only the speaker acting in front of a blue screen, whereas the background (weather charts) is emanating from another source (see "Coding of Visual Objects" section for more details). Today, these two signals are combined in the studio prior to broadcasting. With MPEG-4, the background and foreground (talking person), will be delivered separately as two audiovisual objects.

Service provider: Encodes each object using the appropriate MPEG-4 encoder, composes the audio-visual scene and multiplexes the resulting elementary streams (ES). Multiplexing at this stage is used to group ES with similar QoS requirements, to reduce the number of network connections, or the end to end delay. Each multiplexed stream will require a network connection. Note that objects can also be retrieved from multimedia databases.

Network provider: Translates the multiplexed signal into the protocol of each network (ATM, UDP/IP, etc.). Note that the exact translations from the QoS parameters set for each media to the network QoS are beyond the scope of MPEG-4 and are left to be defined by network providers. MPEG-4 has been carefully designed to scale to different transmission platforms such as wireless or IP. It means that the content needs to be coded only once and automatically played out at different rates with acceptable quality for the communication environment at hand.

End-user: On this side, the MPEG-4 signal coming from the network enters a demultiplexer. The resulting elementary streams are decoded and fed into a compositor which reconstructs the audio-visual scene to be played on appropriate devices (PC display, TV screens, speakers). Note that objects can be locally stored in databases. The end-user has the freedom to interact with the content either locally or remotely by sending a message upstream to the service provider. For end-users, the MPEG-4 functionalities can potentially be accessed on single compact terminals such as PCs, set-top boxes or mobile phones.

Coding of Audio Objects

A big Step forward to the Area of Multimedia

So far MPEG audio in MPEG-1 or 2 has been known for compression of natural music in the area of storage and broadcasting. MPEG-4 audio includes new improved coding algorithms, for all qualities ranging from very low bitrates (below 2 kbit/s) up to higher bitrates (more than 64 kbit/s per channel). In addition to music coding, MPEG-4 includes now speech codecs. However, in order to meet the needs of the fast growing interactive multimedia world, MPEG-4 audio goes much further than just compression.

New Features to allow a whole new Future

Compared to the former MPEG standards the main novelty of MPEG-4 audio is the handling of audio objects and scalability. A composition of several MPEG-4 audio objects can give a scene like a music band consisting of a singer, a piano player, a bass and drums. Each sound generated by these musicians will correspond to one audio object, which can be mono or stereo. The whole composition is described by the scene description (see section "Scene Description: Managing the Objects"), which can be located in a 2D or 3D space. The use of audio objects allows, for example, that a listener plays, say the piano, simultaneously with the band without this latter.

The MPEG-4 audio tools include two methods for synthesizing sounds. The first is based on structured descriptions and the second one on text-to-speech conversion. The source for the genera-

tion of synthesized sound can be text data or so-called instrument descriptions. Additional coding parameters can provide effects, such as reverberation and spatialization. Synthesized sounds enable very low bitrates and other functionalities, such as play-back at different speeds at the same pitch, or at different pitches at the same speed.
Figure 2 (left side) shows the content which can be handled by MPEG-4 audio. Objects, such as music, speech, synthesized sound and text, are possible inputs of an MPEG-4 audio coder. To encode these different types of sources, five different types of codecs have been standardized (fig. 2). Effects can be added to each of the encoded objects which are finally composed together.
The coders are optimised to work at different bitrates and are adapted to different transmission networks (fig. 3).
MPEG-4 standardised a large set of tools rather than a single generic codec. This allows a designer to include only the parts needed for building his application. In addition, this allows to improve only parts of the standard in the future. The encoder side, which includes most of the complexity, is not part of the standard. Only the bitstream and the decoder, which is much simpler than the encoder, are standardised.

Coding of natural Audio Objects

The natural part of MPEG-4 audio is split in two parts: music and speech, with respectively T/F (time/frequency) and parametric codec for music, and parametric and CELP (code excited linear predictive) codecs for speech.
The T/F music codecs are well known and are similar to the MPEG-1/2 Layer I, II, III audio codecs [13818-3]. The state-of-the-art T/F codec is the AAC (Advanced Audio Coding) codec, which delivers stereo FM quality at bitrates around 64 kbit/s and transparent[1] quality stereo at 128 kbit/s and higher bitrates. MPEG-4 has adapted the MPEG-2 AAC codec and added some new functionality. The AAC is available in three profiles called main profile, low complexity profile and scalable sampling rate (SSR) profile. The main profile (resp. the low profile) is intended for use when processing, and especially memory, are not (resp. are) limited. The SSR profile is used when a scalable decoder is needed. The second T/F codec is the Twin-VQ (transform-domain weighted interleaved vector quantization) which performs well at very low bitrates (6 and 8 kbit/s).
When lower transmission bitrates are available, best music and speech qualities can be achieved by means of parametric coding. The idea behind this type of codec is to decompose an audio signal into sinusoidal, harmonics and noise elements (fig. 4).
One MPEG-4 parametric codec for music and one for speech have been specified [14496-3]. The music codec runs at bitrates from 4 to 16 kbit/s with 8 and 16 kHz sampling rates and the speech codec from 2 to 4 kbit/s for narrowband speech. The speech coder runs at the lowest bitrate known for speech coders today.
For speech at higher bitrates (4 to 24 kbit/s) a CELP codec for narrowband and wideband is used. More details can be found in [14496-3].
With this set of codecs MPEG-4 can handle a large variety of natural content at the best quality for a specific bitrate compared to any other existing codec. The use of scalable options (e.g. the base layer could be CELP or Twin-VQ encoded and the upper layer encoded with the AAC) and the handling of audio object offers a lot of new possibilities.

Coding of synthetic Audio Objects

The synthetic part of MPEG-4 audio is separated in the parts of structured audio (SA) and text-to-speech (TTS).
The main idea behind SA is to transmit the description of a sound rather than the compressed sound itself. This way of transmission only needs an extremely low bitrate. The following tools are part of SA:

- Structured audio orchestra language SAOL: This is a describing method for sound generation (synthesis). It can be used for the sound generation with different methods (wavetable, frequency modulation, additive, physical-modeling, and granular synthesis), as well as hybrids of these methods. This allows a user to build his own instrument.
- Structured audio score language SASL: This language allows controlling (e.g. timing) the instruments built with SAOL. It is similar to the known MIDI, which is less powerful but can be used instead of SASL. This allows the author to combine instruments to write songs.

SA audio allows to describe instruments and to control them. This could be for example a synthetic orchestra (designed in SAOL) playing a melody (written in SASL). In addition, SA could even be used to decode natural speech (e.g. CELP format). In order to do this, the CELP decoder, used for natural speech, could be described in SAOL, and transported together with the CELP bitstream (SASL) to the SA decoder. Then, the SA decoder provides exactly the same functionality as a standard CELP decoder. This illustrates that SA can be used in a very flexible and powerful way.

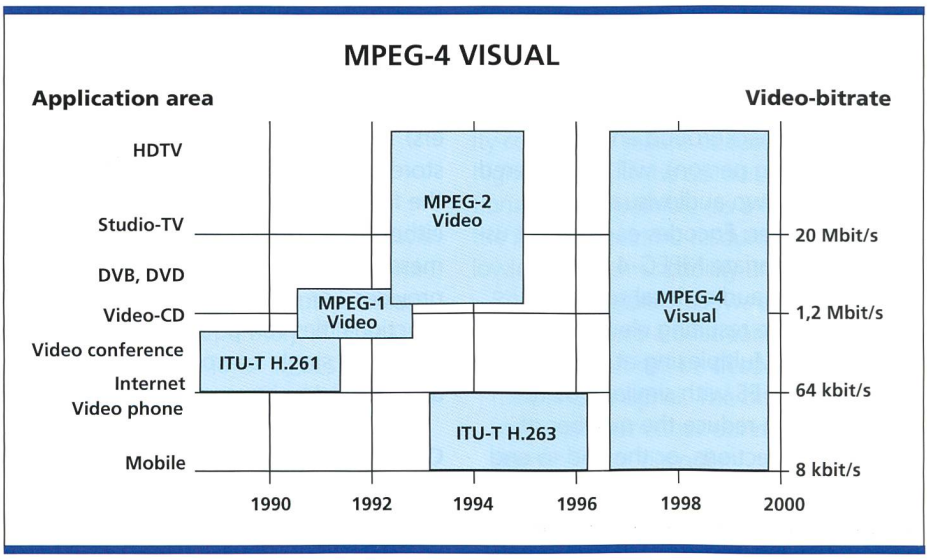

Fig. 5. Visual part of the MPEG-4 standard in relation to earlier standards.

[1] Transparent quality means that there is no audible difference between the original and the encoded sequence.

The text to speech TTS decoder allows to translate a text into speech using phonetic symbols and synthesise them to spoken words. Actually within MPEG-4 only a TTS interface is standardised, which allows to implement TTS within a MPEG-4 system. The speech signal can be synchronised with facial animation and natural video content.

Coding of Visual Objects

Starting with some reflections on the evolution of video compression standards, this chapter introduces the new functionality in the visual domain. Coding of natural and synthetic objects is then described a bit more in details. Finally, robustness aspects in error prone environment is addressed.

Why a new Standard for Video Compression?

Conventional video compression standards bring dedicated solutions with regard to application areas, data rates and functionality. For example, MPEG-2 is targeting high quality digital video broadcast whereas H.263 is more likely used for video phone and video conferencing applications. MPEG-4 is intended to allow some level of interoperability with standards such as MPEG-1, MPEG-2 and H.263. This means, conventionally encoded video objects of these types can be inserted into an MPEG-4 audiovisual scene. Furthermore, in its "Visual" part, the MPEG-4 standard provides new dedicated coding methods for natural and synthetic visual objects as described in the following paragraphs. An overview of the relationship between the different standards is shown in figure 5.

On one hand, those new coding methods built in MPEG-4 conciliate all the different aspects of the previous standards, and on the other hand they provide important new functionality.

New Functionality in the Visual Domain

Current trends show that video are produced more and more graphic and object oriented. TV commercials best exemplify this trend by incorporating not only natural video but also computer generated 3D graphics, text, charts, and all other artifacts that could make commercials attractive. This means that already today, all these "objects" needed to produce such commercials are generated separately and composed in professional studios prior to be transmitted and displayed on the customers TV. For example, figure 6 shows the mechanism to produce a typical weather forecast TV show. A camera records the speaker in front of a blue background. This allows to separate the different planes relatively easily, and to combine the foreground object (the speaker) with a background. In this example, it corresponds to a weather chart generated by a computer. The resulting scene is then encoded, transmitted either by satellite or cable and displayed on the customers TV. The user interaction is limited to frame oriented functionality, like fast forward, slow motion, or pause. Interactivity is left to the TV show producers to compose the scene.

With MPEG-4, in this example, the foreground and background will be encoded and transmitted separately, and composed only in the customer's equipment (PCs, set-top box, fig. 1). This object-based coding has thus the following implications:

- Extension of the interactivity to the end-user: since the objects are composed at the user side, the customers have potentially similar possibilities as the studio producers to interact with the scene (up to the limit offered by those).
- Higher compression of the content: optimized and dedicated algorithms can be used to encode each type of objects (video, textured images, 2D and 3D synthetic objects). For video, this implies the development of new algorithms for the coding of arbitrary

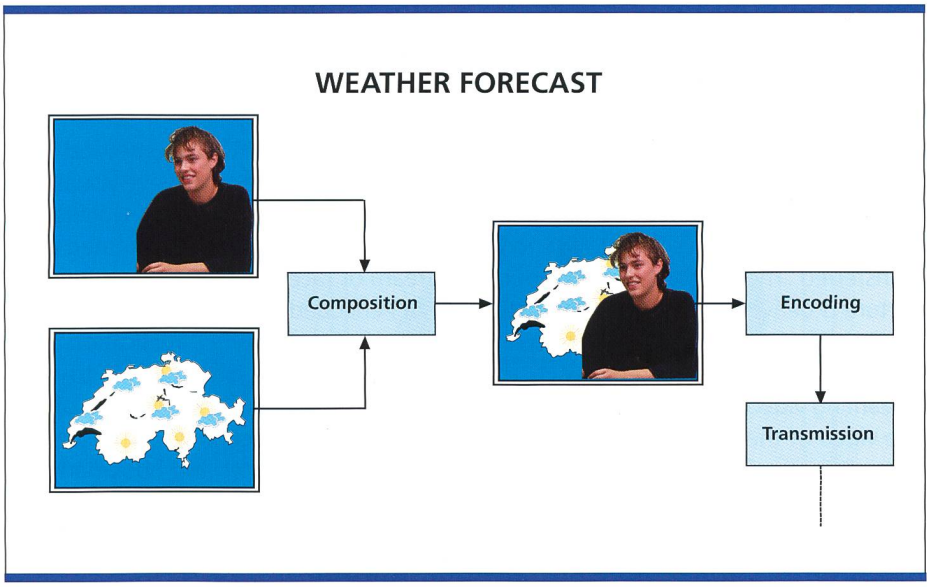

Fig. 6. Production of a weather forecast TV show in existing studio.

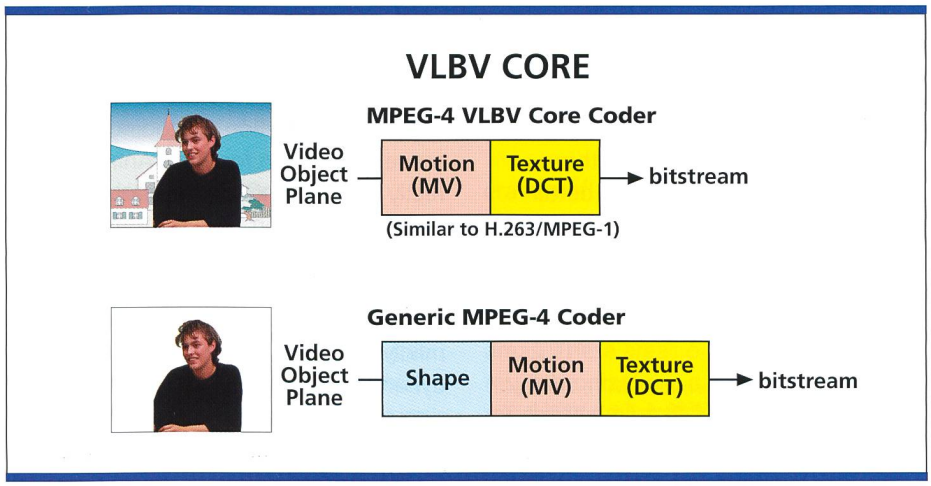

Fig. 7. The VLBV Core and the Generic MPEG-4 Coder.

shaped moving objects (see section "Video coding")
- Scene composition is performed at the user-side: this will modify the production environment and allow the end-user to benefit from the modern production environment.
- Object-based bandwidth allocation (scalability): for example, more bandwidth can be allocated to the foreground when transmission capacities are limited. Also, only the foreground can be transmitted whereas the background could be generated locally on the end-user equipment.

In order to meet the requirements for the encoding of MPEG-4 streams, the scene has to be segmented into different objects according to some semantic meaning. It is important to mention that this segmentation phase is not part of MPEG-4 standard, and is left to the content provider. There exists two solutions for segmentation:
- Separate coding of the objects generated in the studios (as in the example above)
- Computer-based segmentation of scenes, taken up as a whole by a camera. This process may require a certain degree of human supervision since it is very difficult to teach computers what semantic objects means.

Although the computer vision community has made tremendous progress in the recent years, reliable computer-based segmentation remains quite a challenging project. Thus, whenever possible, coding of objects should be performed separately, prior to composition.

Coding of natural Objects

Video coding

A selection of competitive algorithms has been considered at the beginning of the standardisation work. In so called core experiments, these algorithms have been confronted to the wide spectrum of requirements in terms of coding efficiency and functionality. The current algorithm is the result of this evaluation process, but is also open to future improvements.

The basic algorithm for encoding video sequences is based on the hybrid transform algorithm (for example [13818-2]). The basic principle is known from the previous standards, but it has been adapted to the needs of MPEG-4, i.e. the support of arbitrary shapes. Temporal redundancies (e.g. static parts in the video) are reduced by coding only the differences between successive frames in a sequence. Spatial redundancies (e.g. large homogenous areas in an image) are reduced by special mathematical transformations. These two processes form the core of the MPEG-4 video coder and are used to code conventional rectangular images and video (see upper part of fig. 7). Moreover, in order to support arbitrary shapes, the contour information of visual objects needs also to be coded. The details of these processes are described in [N2501].

Figure 8 illustrates a typical usage of the arbitrary shape coding, and of the object-based concept. The video of the tennis player (without background) is coded with the new MPEG-4 video encoding algorithm. The video of the background (play field) can be encoded in the form of so called sprites, which are static images built from image sequences. Thus, for the background, a compression algorithm for static images is rather used (section "Still texture objects"). The background image can be transmitted first and locally stored in the customers PC. Then, in order to play the scene, only camera motion values and the video of the tennis player have to be transferred. The new visible background picture is derived from the locally stored static sprite through geometric calculations in the decoder. This enables streaming video at lower bit rates (only the video of the tennis player is transmitted).

Still texture objects

Still texture coding is used in MPEG-4 for several purposes, like coding of still pictures as such, coding of static backgrounds, and coding of textures for mapping onto the surface of 2D or 3D geometrical models. Still texture objects are encoded with a wavelet[2] transform based algorithm. It has the following advantages compared to standard Discrete Cosine Transform (DCT) algorithms:
- Enhanced coding efficiency (better quality/bits-per-pixel ratio);
- More accurate psycho-visual weighting;
- No blocking artefacts[3];
- Better spatial scalability characteristics with less complexity;
- Easily realizable preview functions.

A similar algorithm is also foreseen for the future JPEG2000 standard. This latter will probably replace the current JPEG standard, which is very popular on the web for coding of still pictures (jpg extension).

Coding of synthetic Objects

This chapter deals with synthetic visual objects encoding, as described in the visual part of MPEG-4 [N2502]. In addition, there exist general computer graphics elements which are integrated in the BIFS tool described in the system part of MPEG-4 (section "Putting Objects together"). Short descriptions of the main synthetic objects and related usage are listed below:

Face object and facial animation

A 3D (or 2D) face object is a representation of the human face for portraying the visual manifestations of speech and facial expressions. This geometrical model is adequate to achieve visual speech intelligibility and lips synchronization. A face object is animated by a stream of face animation parameters (FAP) encoded for low-bandwidth transmission. An interesting application of the animated face object is text-to-speech (see also section "Coding of synthetic Audio Objects"). A converter translates phonemes and bookmark to FAP. These are used to animate the synthetic face, with changes of expression synchronized with the synthetically spoken text.

Body object and body animation

Body Animation (to be standardised in MPEG-4 Version 2) is being designed by the MPEG-4 community to work in a thoroughly integrated way with face/head animation. Decoding and scene description for Body Animation directly mirror technology already proven in Face Animation.

Mesh object

Meshes are well suited to represent mildly deformable objects. Moreover, 2D dynamic meshes achieve high compression since only the motion of a limited number of points (the nodes) has to be coded. Further information can be found in [N2459].

[2] Like the Discrete Cosine Transform (DCT), which ist used for motion video in MPEG-4, the wavelet transform brings the picture information into a more appropriate from in order to apply further methods for compressing data.

[3] Image dirstortions in form of squares around the objects contours.

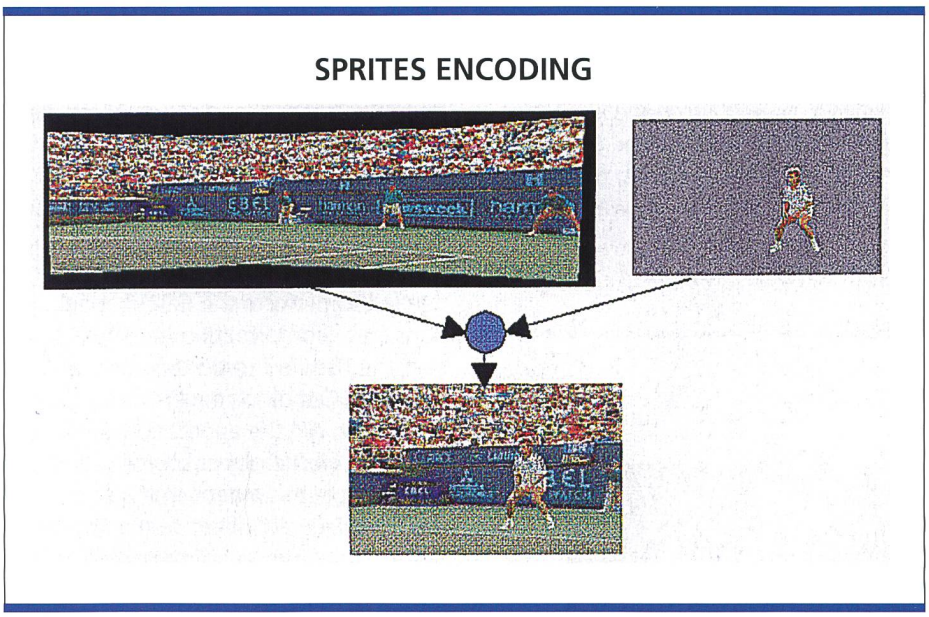

Fig. 8. Encoding of background pictures with sprites.

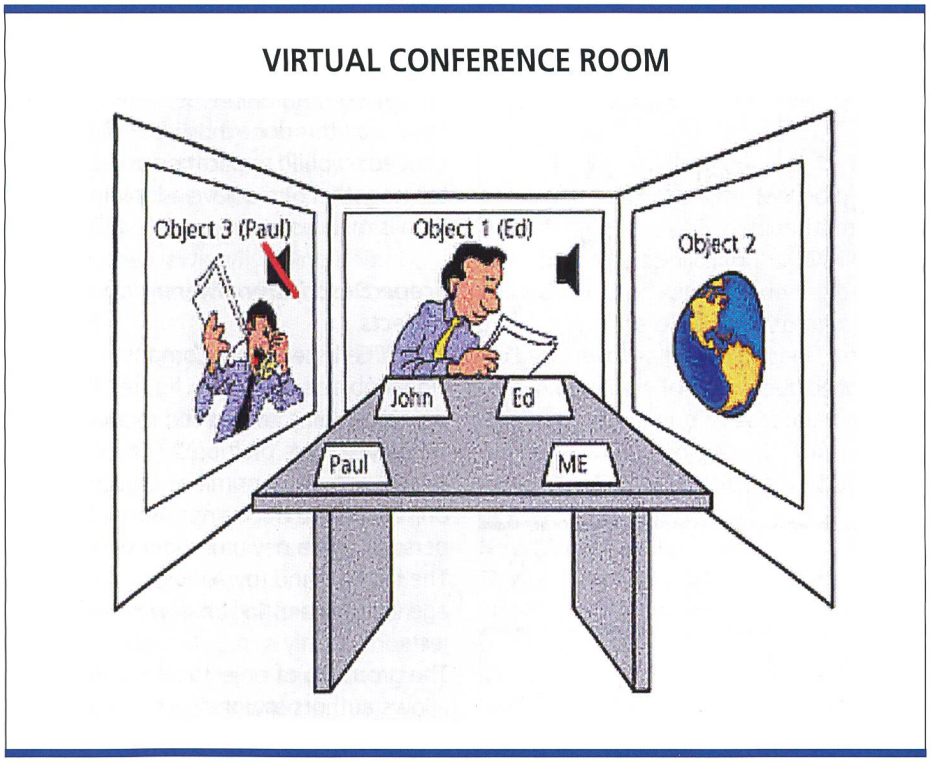

Fig. 9. An MPEG-4 3D Application: Virtual Conference Room.

Abbreviations

AAC	Advanced Audio Coding (MPEG Audio Codec)
ATM	Asynchronous Transmission Mode
BIFS	Binary Format for Scenes
CELP	Code Excited Linear Predictive (Speech Codec)
DAB	Digital Audio Broadcast
DMIF	Delivery Multimedia Integration Framework
DVB	Digital Video Broadcast
DVD	Digital Versatile Disk
ES	Elementary Stream
HDTV	High Definition Television
IP	Internet Protocol
JPEG	Join Picture Expert Group
MPEG	Motion Picture Expert Group
PC	Personal Computer
QoS	Quality of Service
SA	Structured Audio
SASL	Structured Audio Score Language
SAOL	Structured Audio Orchestra Language
T/F Coder	Time-Frequency Codec
TTS	Text-to-Speech
Twin-VQ	Transform-Domain Weighted Interleaved Vector Quantization
VRML	Virtual Reality Modeling Language

Robustness in Error Prone Environment

MPEG-4 provides error robustness and resilience to allow accessing image or video information over a wide range of storage and transmission media. In particular, due to the rapid growth of mobile communications, it is extremely important that access is available to audio and video information via wireless networks. This implies a need for useful operation of audio and video compression algorithms in error-prone environments at low bit-rates (i.e., less than 64 kbit/s). The error resilience tools developed for MPEG-4 can be divided into three major areas. These areas or categories include resynchronization, data recovery, and error concealment [N2459]:

- Resynchronization tools, as the name implies, attempt to enable resynchronization between the decoder and the bitstream after a residual error or errors have been detected.
- Data recovery tools attempt to recover data that in general would be lost. These tools are not simply error correcting codes, but instead techniques which encode the data in an error resilient manner.
- Error concealment tries to hide errors, due to loss of a part of the MPEG-4 stream, by using information that has been received correctly.

Putting Objects together

What is done with all these encoded objects? Somehow they must be put together to build up an audiovisual scene. Consider the following example that illustrates what could be a typical MPEG-4 application.

Today's video conferencing systems consist of some video windows of the partic-

ipant and sometimes of a white board, where everybody can write or sketch a message. There is no intuitive way to help using the conference system. With MPEG-4 there exist tools to make a totally different approach. Not the video is the main focus, but rather an easy to use interactive application, that helps the exchange of information. A virtual 3D room is provided by the conference application (fig. 9) in which the important objects are spatially arranged (e.g. videos of the participants, graphics, documents etc.). The user can virtually walk through the room and have a closer look at the objects, arrange them in a new way, change the properties of objects (e.g. mute Paul's video) and so on. More information on one participant can be obtained by a simple mouse click on the name card, e.g. the company Ed is working for, his address, etc.

So far only "local interactivity" was needed to perform the described actions, i.e. no information is sent back over the network to do the interaction. With MPEG-4 much more is possible. Via a common API the application can have influence on the object source itself. In our example this means that a participant, say Ed, can influence the application of his communication partners. They hear a bell ringing when Ed wants to have their attention, or Ed's video is growing and his audio is louder than the ones of the others when he wants to say something. Or, more interestingly, Ed could show the 3-dimensional model of a new car prototype to illustrate his presentation.

How is all this done in MPEG-4? What is needed to build applications that can talk together? Lets have a look into more details of the system part of MPEG-4.

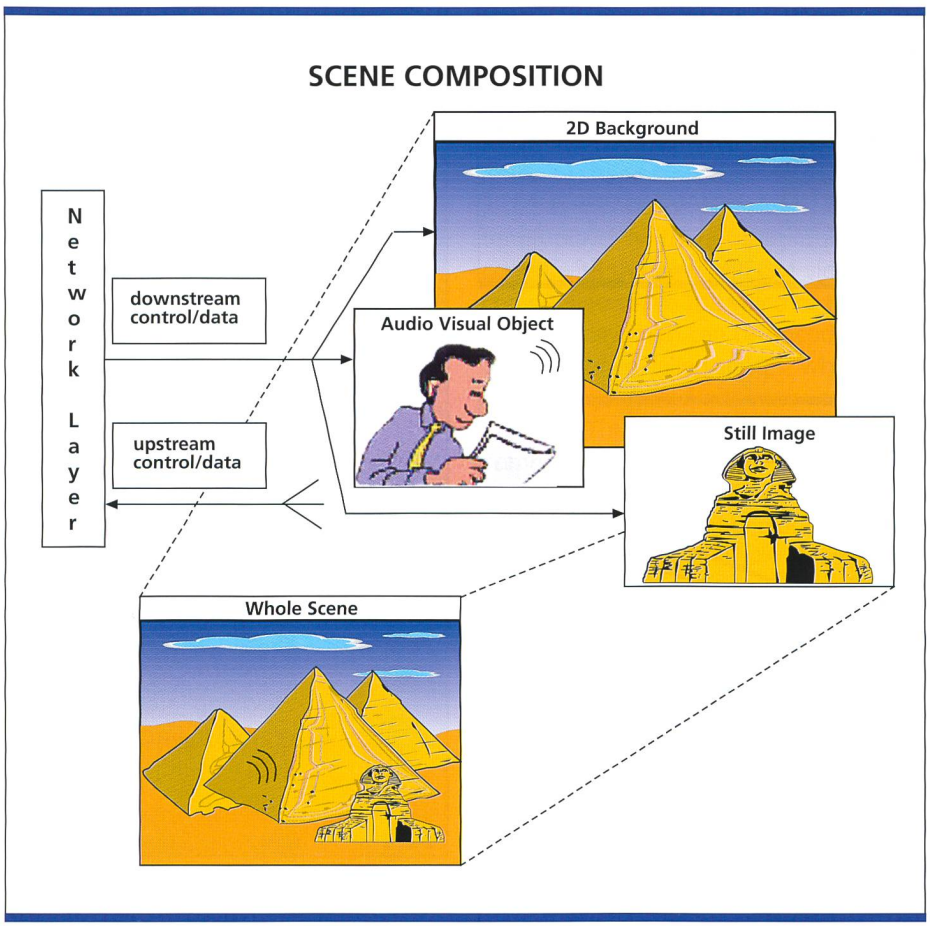

Fig. 10. The composition of an MPEG-4 scene.

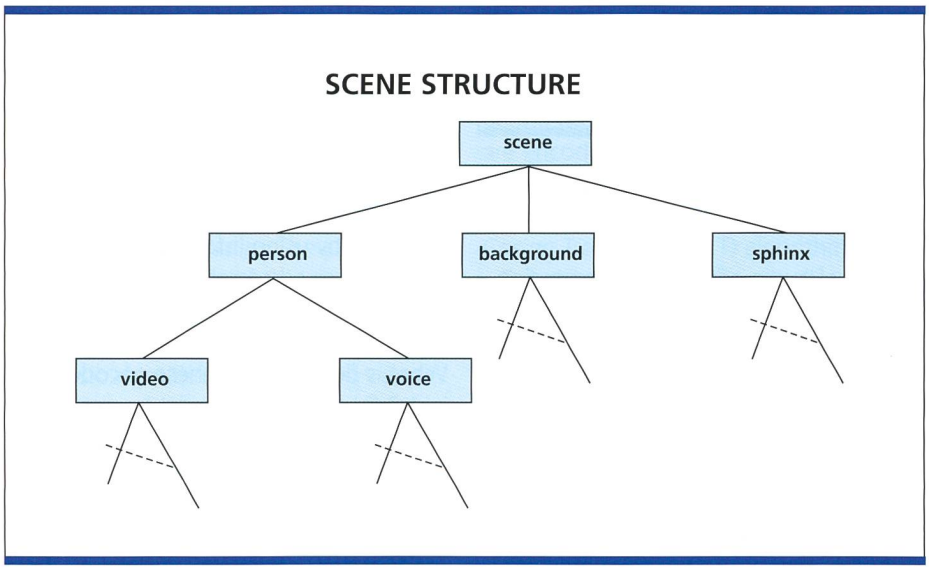

Fig. 11. Logical structure of a scene.

Scene Description: Managing the Objects

In MPEG-4 scenes are composed of individual objects as seen in figure 10. The figure contains compound media objects, which are built up by objects themselves, and other primitive objects[4]. In our figure the video and the voice of the person form a new compounded object. The background (pyramids) and the image (sphinx) are the other primitive objects.

The grouping of objects to new objects allows authors to construct complex scenes, and enables consumers to manipulate meaningful (sets of) objects. More generally, MPEG-4 provides a standardized way to describe a scene, allowing for example to:

- place media objects anywhere in a given coordinate system;
- apply transforms to change the geometrical or acoustical appearance of a media object;
- group primitive media objects in order to form compound media objects;
- apply streamed data to media objects,

[4] Primitive objects are elementary objects, in contrast to compound objects.

in order to modify their attributes (e.g. moving texture belonging to an object; animation parameters animating a moving head);
– change, interactively, the user's viewing and listening points anywhere in the scene.

A special language for describing and dynamically changing scenes has been included in MPEG-4 and is called Binary Format for Scenes (BIFS). A lot of the functionality of BIFS has been adapted from the Virtual Reality Modeling Language (VRML).

A BIFS scene follows a hierarchical structure, which can be represented as a tree. Each node of the graph is a media object, as illustrated in figure 11 (note that this tree refers back to fig. 10). The tree structure is not necessarily static; node attributes (e.g. positioning parameters) can be changed while nodes can be added, replaced, or removed.

In general, the user observes a scene that is composed according to the design of the author. Depending on the degree of freedom allowed by the author, however, the user has the possibility to interact with the scene. Operations a user may be allowed to perform include:
– change the viewing/listening point of the scene, e.g. by navigation through a scene;
– drag objects in the scene to a different position;
– trigger a cascade of events by clicking on a specific object, e.g. to start or to stop a video stream;
– select the desired language (German, English,...) when multiple language tracks are available;
– more complex kinds of behavior can also be triggered, e.g. a virtual phone rings, the user answers and a communication link is established.

Interaction

In order to make an interactive application possible, a whole communication architecture for managing the different datastreams is defined, the so called Delivery Multimedia Integration Framework (DMIF). DMIF provides the MPEG-4 programmer with a common interface to establish and close connections to local and remote data streams in both directions (up- and downstream). In addition, DMIF is responsible for multiplexing the different datastreams. The delivery technology, which encompasses transport network technologies (e.g. Internet, ATM infrastructure), broadcast technologies or local storage, is totally hidden from the DMIF user (the application) by DMIF. Additional functions are to hide the delivery details from the DMIF user, to manage QoS sensitive real time channels, to allow service providers to log resources for usage accounting, and to ensure interoperability between end-systems. More details on DMIF can be found in [N2506a].

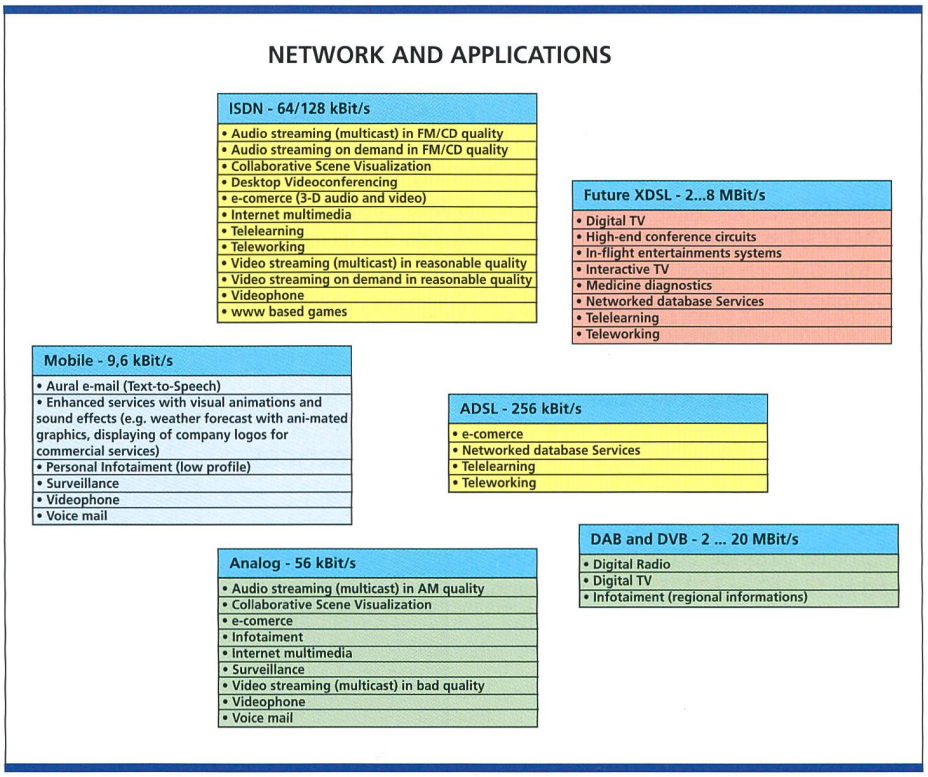

Fig. 12. Network types and the possible MPEG-4 applications.

New World of Business

This last section concludes the article by highlighting some consequences of MPEG-4 on the information society. In the first paragraph, a classification of MPEG-4 applications as function of existing (and future) networks is given.

Possible Applications on current and future Networks

The available bitrate on a network is one key parameter which allows applications to work with a sufficient quality. Figure 12 shows possible applications enabled with MPEG-4 classified by network types. In general an application designed for a specific network will also function properly on network with higher bandwidth.

Some explanations for a choice out of these applications are displayed below.

Internet multimedia

The Internet community will progressively adopt MPEG-4 as the standard for compressed audio and video in web pages. The great benefit for service providers is that they can prepare the content in a single form for all different kinds of service platforms.

For the users, new and sophisticated voice and video mail systems will be made available by MPEG-4.

Video streaming

Video on demand: The customer has the possibility to choose among a large number of movies. For each movie which is watched, an individual stream (audio or audio/visual) is transferred over the net. The user is able to control the stream via "VCR-like" buttons, e.g. fast-forward, play, pause, etc.

Near video on demand: The customer has the possibility to choose from a limited number of programs. They are "joining in" the existing streams.

Digital TV

Digital Television (DTV) will change the nature of television broadcasting since digital data together with digital audio/video can be delivered to consumers. Digital data can enhance the consumers' viewing experience by providing a more interactive environment.

By its very nature, MPEG-4 is perfectly suited to provide this functionality. Many new applications will be made possible through MPEG-4 in combination with digital TV broadcasting.

Surveillance
The MPEG-4 standard offers compression tools, which meet the requirements of surveillance applications. Examples of such requirements are low encoder complexity, low delay, robustness in error-prone environments, low bitrate mode and scalability.

Videoconferencing
MPEG-4 Videoconferencing applications will provide extended functionality compared to H.320 or H.323 (see paragraph "Putting Objects together" for a comprehensive example).

Video phone
The new object oriented features of MPEG-4 will allow to enhance the quality of the scene representation in particular for low bitrate applications.

Collaborative Scene Visualization
Collaborative Scene Visualization supports a class of Computer Supported Cooperative Work (CSCW) applications where groups of people are working simultaneously in distributed locations to accomplish a task by sharing a common visual information space.
A trend of this kind of applications is that they will provide Augmented Reality (AR). The objective of AR is to create an environment in which a user perceives both real and virtual/synthetic (generated with a computer) objects in a seamless way.

Tele-learning
The standard of MPEG-4 is able to fulfill the requirements of telelearning applications:
- Learning over distance in closed user groups;
- Possibility to interact with an instructor;
- Interaction between participants;
- Rich tool set for scene representation;
- Networked database Services.

MPEG-4 provides mechanisms to efficiently access multimedia contents from databases. Such content is described by attributes, like keywords or numerical values. The process of segmenting and describing of multimedia content is part of MPEG-7.

The new Culture
If we look at the evolution of the Internet over the last years, we can observe a steady growth of traffic volume. In this context the transported content is moving from poor text to audio and video. Interactivity and multimedia content is also the general trend on other platforms (mobile, broadcast). As highlighted in this article, MPEG-4 has the potential to generate these new types of applications which enable a new culture of content, service and transport. Techniques like VRML, Java, compression, natural and synthetic content, scalability and objects have been adapted from different areas and improved to build a new powerful standard. MPEG-4 has so the potential to make true multimedia happen.

References

[AMUSE] Project AMUSE/phase 2 (Acts project AC011) Advanced multimedia services to residential users. http://www.snh.ch/projects/amuse/html/basle_site.html (Swisscom CT Exploration Project EP9702_29).

[N2457] ISO/IEC JTC1/SC29/WG11 Document N2457 (Atlantic City, October 1998): MPEG-4 Applications.

[NN2431] ISO/IEC JTC1/SC29/WG11 Document N2457 (Atlantic City, October 1998): MPEG Audio FAQ Version 9.

[AES-1/2-97] Bosi, Brandenburg "Overview of MPEG audio"; Journal of the audio engineering society; Vol 45, No.1/2 1997.

[IEEE 2-99] Koenen "MPEG - multimedia for our time"; IEEE Spectrum 2-99.

[N 2424] ISO/IEC JTC1/SC29/WG11 Document N2423 (Atlantic City, October 1998): Report on the MPEG-4 speech codec verification tests.

[N 2425] ISO/IEC JTC1/SC29/WG11 Document N2423 (Atlantic City, October 1998): Report on the MPEG-4 audio on Internet verification tests.

[N 2157] ISO/IEC JTC1/SC29/WG11 Document N2423 (Tokyo, March 1998): Report on the MPEG-4 NADIB verification tests.

[N2459] ISO/IEC JTC1/SC29/WG11 Document N2459 Overview of the MPEG-4, Standard, October 1998, Atlantic City.

[N2460] ISO/IEC JTC1/SC29/WG11 Document N2460 MPEG-7: Context and Objectives, October 1998, Atlantic City.

[13818-2] ISO/IEC 13818-2: Information technology – Generic coding of moving pictures and associated audio information (MPEG-2) – Part 2:Video;1995.

[13818-3] ISO/IEC 13818-3: Information technology – Generic coding of moving pictures and associated audio information (MPEG-2) – Part 3:Audio;1995.

[14496-3] ISO/IEC 14496-3: Information technology – Generic coding of moving pictures and associated audio information (MPEG-4) – Part 3:Audio;1998.

[N2459] ISO/IEC JTC1/SC29/WG11 Document N2459 (Atlantic City, October 1998): Overview of the MPEG-4 Standard.

[N2501] SO/IEC JTC1/SC29/WG11N2501 INFORMATION TECHNOLOGY – CODING OF AUDIO-VISUAL OBJECTS Part 1: System ISO/IEC 14496-1 Final Draft of International Standard Version of: 18 December, 1998.

MPEG-4 allows new services and applications independently of the transport media. The same application or encoded sequence can adapt to different kinds of transport networks. Furthermore an MPEG-4 scene can contain a variety of different object types, which can be any kind of synthetic and natural audio, video and graphics. All these objects can be independently encoded and reused in new scenes. All this allows much more flexible services.
The object based concept of MPEG4 enables an easy way to include interactivity to an application. This interactivity can be local or remote and is supported by the system part of the MPEG4 standard.
Many of the new applications will produce asymmetrical traffic. Real time and QoS (e.g. guaranteed bandwidth) aspects become more and more important.
Error correction is done within the compression process which makes the transport of the content much more robust.
The compression tools provided are based on today's most efficient and flexible techniques.
MPEG-4 is now an available standard. The first applications have been shown by Microsoft's multimedia player, which operates according to the standard. The next applications will probably be in the domain of mobile services and equipment. In the broadcast domain, MPEG-4 has been evaluated by the European Narrowband Digital Audio Broadcasting Group to replace the current analog AM-broadcasting. [4, 7]

Acknowledgments
The authors would like to thank Mr. Johannes Schneider for the proofreading of this document and for his constructive remarks.
This article is the result of a teamwork with an equal engagement of all members of the authoring team.

References

[N2502] ISO/IEC JTC1/SC29/WG11 N2502a (Atlantic City, October 1998) INFORMATION TECHNOLOGY – GENERIC CODING OF AUDIO-VISUAL OBJECTS Part 2: Visual ISO/IEC 14496-2 Final Draft of International Standard Version of: 13 November, 1998, 15:27.

[N2506a] ISO/IEC JTC1/SC29/WG11 Document N2527 (November 1998) INFORMATION TECHNOLOGY – GENERIC CODING OF AUDIO-VISUAL OBJECTS Part 6: Delivery Multimedia Integration Framework ISO/IEC 14496-2 Final Draft of International Standard Version of: 15 November, 1998.

[N2460] ISO/IEC JTC1/SC29/WG11 Document N2460 (Atlantic City, October 1998), MPEG-7: Context and Objectives (version - 10 Atlantic City, October 1998).

[N2527] ISO/IEC JTC1/SC29/WG11 Document N2527 (October 1998) MPEG Systems (1-2-4-7) FAQ, Version 7.0a.

[CT877-2] MPEG-2 over ATM transfer trials between Bern and Leidschendam.

[CT919-1] Überblick Projekt MPEG/DVB.

[CT850-3] Übertragung von MPEG-2 Transportströmen über ATM-Netzte (Schlussbericht zum Projekt MPEG-2 over ATM).

[ATM Express] OVL Report (Swisscom AG, KPN Research, Telia), May 1998, High quality video conferencing over ATM, ATM Express technical report (CT Exploration Project EP9705_19).

[CT-1212-3] KaDAS (Untersuchung der Eigenschaften von kaskadiertenk, datenkomprimierten Audiosignalen) – Schlussbericht.

[CT-1273-1] Audio Visual Quality, the AVQ project.

[CT-1274-1] MPEG-4 deliverable.

[COMTEC-9/10 1996] DAB: Vom Diensteanbieter zum Nutzer.

Zusammenfassung

MPEG-4 – die neue Multimediatechnologie
Den Trends der Informationsgesellschaft folgend, hat die «Moving Picture Coding Expert Group» (MPEG) einen neuen Standard MPEG-4 für die Codierung und Integration von audiovisuellen Objekten vorgeschlagen. Dieser Standard befasst sich mit allen Möglichkeiten eines digitalen Umfeldes: einerseits mit der Codierung von natürlichem Audio und Video, und andererseits mit computergenerierten Objekten, wie beispielsweise animierte Grafiken, künstliche 3D Welten und Töne. Der Benutzer hat nun die Freiheit mit der Szene zu interagieren, indem er Objekte löscht, einfügt oder neu positioniert. Die Eigenschaften eines Objektes können durch einen einfachen Mausklick modifiziert werden. MPEG-4 wurde mit der Absicht geschaffen, verschiedene Übertragungsplattformen, beispielsweise drahtlose oder IP basierte, zu unterstützen. Das bedeutet, dass der Inhalt nur einmal codiert werden muss, und mit verschiedenen Bitraten bei ansprechender Qualität wieder abgespielt werden kann.

Beat Herrmann *received his degree in Electrical Engineering from the HTL Burgdorf, in the spring of 1983. Study projects were building of a computer graphics hardware and realization of an audio multiplexing transmission system. In 1990 he received a postgraduate degree in Software Engineering from the SWS Bern. Examination project was the implementation of digital audio filters on a transputer network. He joined Swisscom Corporate Technology in the autumn of 1983. Working areas are TV transmission systems, video server technology and DVB. He was responsible for projects for the realization of a framestore based measurement system for digital television signals, for building a codec for studio TV signals, and for the transmission of MPEG-2 transportstreams over ATM networks.*

Denis Schlauss *received his degree in Electrical Engineering form the ETH Zürich 1993. He then joined the former Research & Development Department of Telecom PTT, today's Corporate Technology of Swisscom. He first worked in the Digital Video Broadcasting (DVB) domain, coding of video (MPEG-1, MPEG-2) and simulation of transmission systems. Denis Schlauss is now working on new multimedia technologies like MPEG-4 and MPEG-7 with focus on video and system parts. He is also dealing with problems of making subjective and objective perceptual quality measurements for low bitrate video.*

Thomas Burkhard *received his degree in Electrical Engineering from the HTL Biel in 1990. In the same year he joined the Research & Development Department of Swisscom. In 1993 he received a postgraduate degree in Software Engineering from the SWS Bern. He was working on broadband multimedia technologies with focus on test and demonstration platforms. Since March 1999 Thomas Burkhard is working as Security Manger for Swisscom Mobile.*

Philippe Schroeter *received his diploma in Electrical Engineering from the Swiss Federal Institute of Technology EPFL, Lausanne, Switzerland, in January 1991. In April 1991, he joined the Signal Processing Laboratory at the EPFL as a Ph.D. student. His research interests included image and video compression, image segmentation, computer vision and motion analysis. He developed a three dimensional tool for the automatic segmentation of the brain obtained by Magnetic Resonance Imaging system. In October 1996, he received his Ph.D. in communication systems. In January 1997, he joined the research and development division of Swisscom. His competencies includes fiber based technologies in the access network, multimedia services and call centers solutions. He is active in FSAN (Full service access network) and ITU-T standardization, EURESCOM projects and is currently working in the field of MPEG digital technologies and residential services.*

Daniel Ledermann *was born in 1970 in Laupen. He received a degree in Electrical Engineering from the HTL in Biel in 1996. His graduation project was in the domain of digital audio signal processing. In January 1997 he joined the Signal Processing and User Interface group of the Swisscom Corporate Technology. He worked on perceptual audio, video and speech quality (end-to-end), mainly in the field of objective audio and video quality in the broadcast domain. Daniel Ledermann is currently working on new multimedia technologies like MPEG-4 and 7 and on subjective/objective perceptual quality for low bitrate video.*

J.C. Russ

The Image Processing Handbook
Springer-Verlag GmbH, Heidelberg.
Jointly published with CRC Press in Cooperation with IEEE Press. 3rd ed., 1999, 750 pp., hardcover, Fr. 225.–, DM 150.–, öS 1818.–,
ISBN 3-540-64747-3.

The Image Processing Handbook covers methods for two different purposes: improving the visual appearance of images to a human viewer and preparing images for measurement of the features and structures. The handbook presents an extensive collection of image processing tools, enabling the user of a computer-based system to understand those methods provided in packaged software and to program additions needed for particular applications. With balanced and complete descriptions, the text outlines frequency-space methods with an extensive mathematical presentation and spatial-domain processing requiring only a modest technical background in mathematics or computers.

G. Guekos

Photonic Devices for Telecommunications
Ho to Model and Measure
Springer-Verlag GmbH, Heidelberg.
1999, 404 pp., 211 figs., 22 tab., hardcover, DM 159.–, öS 1161.–,
Fr. 144.–, ISBN 3-540-64318-4.

This book focuses on the basic topics of modeling and measurement for photonic devices and discusses the most modern tools for simulation and experimentation available to engineers and physicists. It presents and compares powerful methods for the numerical modeling of photonic waveguide structures and for waveguide characterization. Further, it explains extensively how to model distributed feedback (DFB) lasers, the key optical source for advanced communications systems, and how to experimentally determine accurately their main characteristics. Finally, it investigates the potential of non-linear properties of semiconductor optical amplifiers (SOAs) for fiber communications through a detailed theoretical and experimental examination of the four wave mixing (FWM) effect. This work provides extensive referencing that covers the latest research results.