MPEG-Standards für Multimedia-Dienste (Video-Standards für Multimedia)

Von Dr. Jürgen Lohr, Berlin

Inhaltsverzeichnis

MPEG-Standards für Multimedia-Dienste

Dr. Jürgen Lohr, Jahrgang 1962,
Projektmanager für Informations- und
Mediendienste im F&E-Bereich der
Deutschen Telekom, T-Nova

1 Einleitung

Eine grundlegende Weiterentwicklungen und eine Neuorientierung von Produkten bei multimedialen Diensten wird durch die veränderten Provider-Strategien und der Globalisierung der TIMES-Märkte verursacht. Das erfordert eine neue Fokussierung und eine neue Gestaltung der Basis-Technologien bei multimedialen Diensten.

Die Provider in diesem Innovationsfeld stellen u.a. Produkte-und-Dienste-Strategie, Content-Strategie, Netzstrategie und Innovationsstrategie auf. Hier wurden folgende strategisch relevante Funktionalitäten identifiziert:

- Modulares, universal nutzbares Netz
- Technikkonvergenz von Telekommunikationsmanagement-Netzwerk, Intelligente Netze und Informationstechnologie
- Orts- und technikunabhängiger Zugang
- Komplettlösungen und einfache Kundenschnittstelle
- Customer Care
- Globalisierung der TIMES-Märkte in Telekommunikation, Informationsverarbeitung, Medien, Unterhaltungselektronik und Sicherheit
- Aktorik und Sensorik (Anwenderschnittstelle)

Ausgehend von den Basisfunktionalitäten werden strategische Produktgruppen gebildet und die Beziehungen zu den Content-Providern definiert. Da die Inhalte für die Produkte und Dienstleistungen, aufbauend auf Netzleistungen, ausschlaggebend für den Erfolg eines neuen innovativen Produktes sind, werden strategische Partnerschaften angestrebt.

Die unterschiedlichen Funktionalitäten können den Bereichen Anwendung/Produkt, Netz , und Benutzer zugeordnet werden. Zwischen den Bereichen und auf das Endgerät sind optimierende Veränderungen zu erkennen (Abbildung 1).

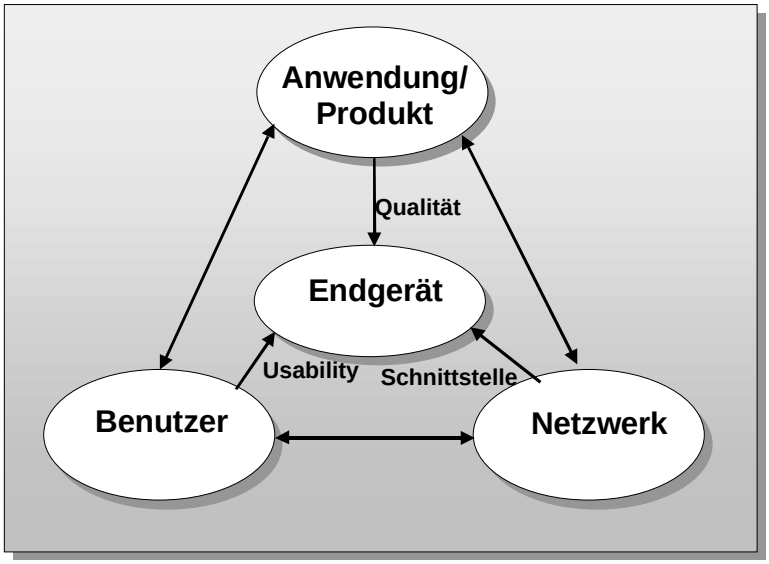

Abbildung 1: Bestrebung der Bereiche

Aufgrund der Globalisierung und der Konvergenz in den TIMES-Märkten ist mit starken Strukturänderungen zu rechnen. Den Strukturänderungen am Markt kann nur mit aktuellen grundlegenden Entwicklungen begegnet werden (Abbildung 2). Die wichtigsten Faktoren der Strukturänderungen liegen in:

- den globalen Zugängen,
- der Navigation und
- der Informationsbeschaffung.

Im Hinblick auf die Provider-Strategien und das Entstehen eines integrierten TIMES-Sektors der Wirtschaft (Telekommunikation, Informationstechnologien, Multimedia, Entertaiment) ist die Entwicklung neuer Formen der Informationsbeschaffung und -präsentation erforderlich, um bestehende Märkte zu sichern und neue Märkte zu erschließen. Einige wichtige Fragestellungen sind:

- Wie kann ein globaler Zugang zu Inhalten, unabhängig von Ort und Endgerät, ermöglicht werden?
- Wie sieht die Navigation der Zukunft aus?
- Wie kann der zunehmenden Informationsflut durch intelligente Inhalte begegnet werden?

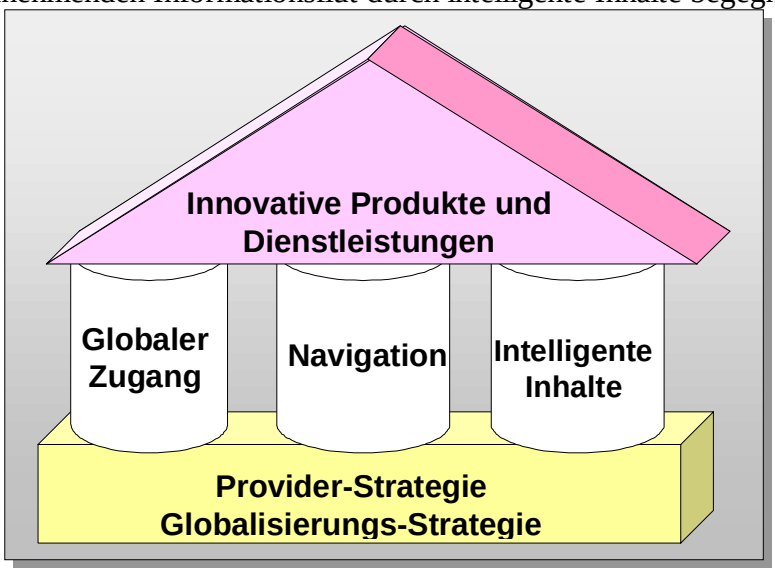

Abbildung 2: Neue Fokussierung

1.1 Globaler Zugang

Die Konvergenz der TIMES-Märkte erzwingt eine Konvergenz der Übertragungswege und der Endgeräte. Es wird zunehmend erforderlich, auf wichtige Informationen weltweit und unabhängig vom Endgerät zugreifen zu können (Abbildung 3). Dies ist bislang in der Regel nur innerhalb einer geschlossenen Nutzergruppe, z. B. Unternehmen möglich. In Zukunft muß der Zugang zu Informationen über verschiedene Endgeräte, basierend auf unterschiedlichen Standards (Webstandards, H320), erfolgen:

- Fernseher,
- PC,
- tragbare mobile Geräte (Laptop, Personal Data Assistent),
- im Fahrzeugen integrierte Geräte (Boardcomputer, Leitsystem)
- Sprachtelefon, Bildtelefone, Screenphone, Fax
- Mobilfunk-Display, Pager

Über beliebige Distributionswege sollte eine Kommunikation möglich sein:

- schmal- und breitbandige terrestrische Netze (PSTN, ISDN, ADSL, ATM, BVN),
- schmal- und breitbandige mobile Netze (GSM, UMTS),
- Satellit und Richtfunk (DVB, MMDS, LMDS).

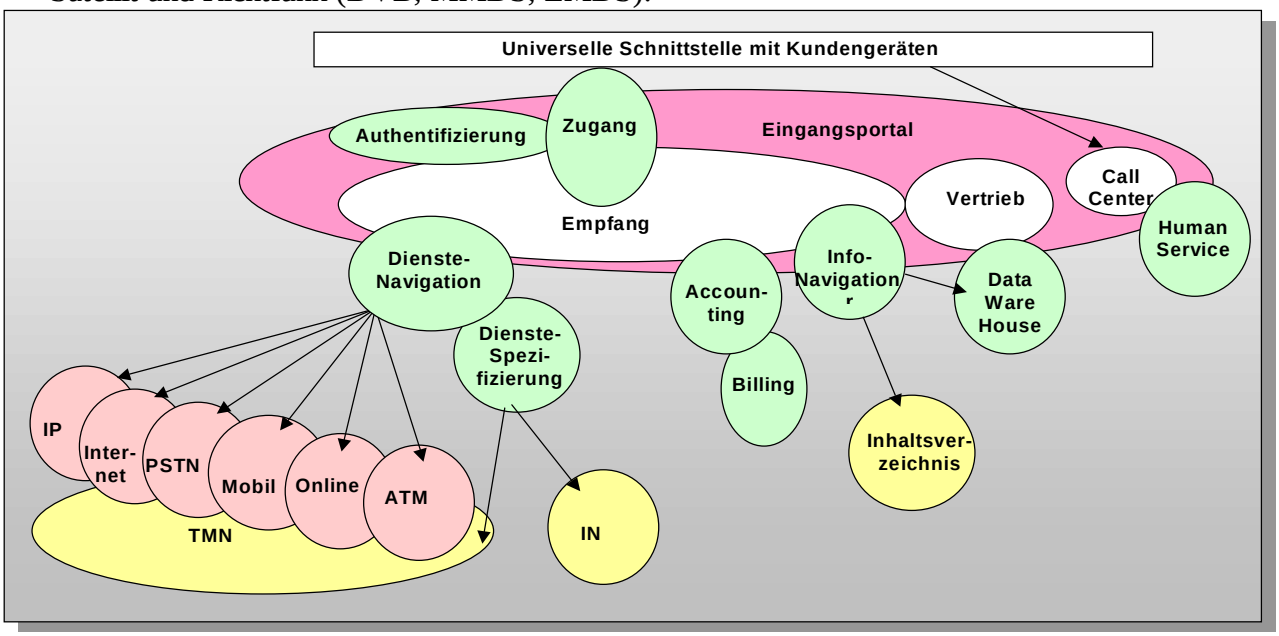

Abbildung 3: Globaler Zugang

Die Aufbereitung der Inhalte stellt einen weiteren Aspekt dar. Je nach Displaygröße der Endgeräte und dem zugrundeliegenden technischen Standard müssen die gleichen Inhalte in unterschiedlicher Form aufbereitet werden. Dies sollte automatisch erfolgen und ist nur möglich, wenn auch für die Gestaltung von Inhalten Grundregeln geschaffen werden, die deren Einsatz als eine Art “modularer Content” sicherstellen.

Das bestehende Bedienungskonzept (Useability) für Endgeräte besitzt unterschiedliche Eigenschaften (Tabelle 4).

Art	TV	PC	PDA
Eingabe	Fernbedienung mit Curser, Fernzeiger	Maus,	Stift
	Funktionstasten	Tastatur	-
	-	Spracheingabe	-
Ausgabe	Großer Bildschirm (81 cm) Großer Sehabstand (ab 3m) Ca 15 Schriftzeilen	Mittlerer Bildschirm (40 cm) Mittlerer Sehabstand (ab 60cm) mind. 30 Schriftzeilen	Kleiner Bildschirm (10 cm) kleiner Sehabstand (bis 30 cm) Ca 15 Schriftzeilen
	-	Sprachausgabe	-
	Tonausgabe	Tonausgabe	-

Tabelle 4: Bedienungskonzept für Endgeräte

1.2 Navigation der Zukunft

Neben der technischen Realisierung des Zugriffs und der Darstellung von Inhalten ist die Navigation zur Erschließung der Inhalte in thematisch strukturierter Form erforderlich. Eine einfache Navigation innerhalb der rasant anwachsenden Menge von Inhalten ein Schlüsselproblem.

Hier müssen neue Lösungsansätze der Navigation, etwa durch Einsatz von folgenden Aspekten, entwickelt werden:

- Sprach Ein-/Ausgabe
- Dreidimensionale virtuelle, grafische Darstellung (VRML)
- Interaktive Videos/Animationen mit Objekterkennung
- Integration von Videoobjekten in 3D-Räumen
- Layer Techniken (Ansätze z.B. MIT)

Hierbei kommt insbesondere der Einsatz von MPEG-4 in Betracht. Im Gegensatz zu MPEG-1 und MPEG-2 will MPEG-4 keine speziellen Anwendungen adressieren, sondern möglichst viele Anwendungsgebiete abdecken. Eine detaillierte Beschreibung des MPEG-Standards erfolgt im nächsten Abschnitt.

1.3 Intelligente Inhalte

Navigatoren erschließen Inhalte in thematisch strukturierter Form und eröffnen so einen Pfad durch die Fülle der Inhalte. Dieser Pfad ist allen Nutzern zugänglich, berücksichtigt jedoch nicht deren persönliche Interessen und Informationsbedürfnisse. Hierzu sind

- persönliche Profile,
- intelligente Agenten sowie
- intelligenter Content (MPEG-7)

erforderlich. Zur Zeit sind erste Internet-Dienste in Richtung persönliche Profile gegangen. Als nächster Schritt ist der Einsatz intelligenter Agenten anzustreben, die automatisiert gezielte Suchvorgänge durchführen und aus Zustimmung bzw. Ablehnung der Suchergebnisse seitens des Nutzers lernen, bis den inhaltlichen Wünschen des Nutzers vollständig entsprochen werden kann. Die Technik der Intelligenten Agenten ist bereits im Prototypen-Status in den Forschungs- und Entwicklungsbereichen der Unternehmen und Universitäten zu finden.

Die zunehmende Bereitstellung digitaler, audio-visueller Contents macht es erforderlich, den potentiellen Anwendern ein Instrument für effizienter und effektiver Suche nach audiovisuellen Inhalten zu geben. Um Suchvorgänge entscheidend vereinfachen zu können, ist intelligenter Content erforder-

lich. Dies gilt besonders für audio-visuelle Inhalte, in denen mit gängigen Methoden keine Volltextsuche oder objektorientierte Suche möglich ist. Mit MPEG-7 soll eine standardisierte Beschreibung verschiedener Arten multimedialer Informationen erreicht werden. Dieser Standard ist zur Zeit in der Definierungs-Phase und wird im unteren Abschnitt erläutert.

2 MPEG-Standard

Die Kompressionstechnologie ermöglicht erst eine Wirtschaftlichkeit der Multimedia-Dienste. Sie basiert auf einer Datenreduktion. Somit benötigen die verschiedenen Medien, wie Bilder, Video und Ton, nicht mehr große Übertragungs- und Speicherleistungen. Die wichtigsten Kompressionstechnologien sind im Audiobereich ADPCM, im Videobereich MPEG und im Bildbereich JPEG.

Die MPEG-Organisation wurde im Januar 1988 ins Leben gerufen, als Teil des "Joint ISO/IEC Technical Committee (JTC 1) on Information Technology", und wird formell auch als Arbeitsgruppe WG11 von SC29 bezeichnet. Jährlich werden drei Treffen abgehalten, die von 300 Experten aus 20 Ländern besucht werden. Mittlerweile sind vier wichtige Standards mit MPEG-1 und MPEG-2 verabschiedet worden sowie mit MPEG-4 und MPEG-7 in Vorbereitung. MPEG3 war ursprünglich für die Anwendungen im Bereich HDTV gedacht. MPEG3 wurden in MPEG2 integriert, so daß MPEG3 heute keine Bedeutung mehr besitzt. MPEG-5 und MPEG-6 wurden nicht definiert und sind demnach nicht von Relevanz.

In den folgenden Abschnitten werden die Themen behandelt:

- Einführung von MPEG-1, -2 und -Audio
- MPEG-4 - Autovisuelle Szenen
- MPEG-4 - Videokompression
- MPEG-4 - Audiokompression
- MPEG-7 - Beschreibungsschnittstelle für Inhalte
- MPEG-7 - Informationssuche von Anwendungen

MPEG-1 und -2 wurden für die Krompession audio-visueller Gesamtszenen entwickelt und erlauben Interaktivität nur auf Programmebene. Sie sind anfällig bei fehlerhafter Übertragung. MPEG-4 wurde für die Kompression audio-visueller Objekte entwickelt und erlaubt eine Szenenkomposition aus Video, Audio und Grafik. Es wird eine Interaktivität auf Objektebene ermöglicht. Bei MPEG-4 wurden Werkzeuge für wirksamen Fehlerschutz, verbesserte Skalierung und Kodiereffizienz definiert. MPEG-7 wurde eine multimediale inhaltsbeschriebende Schnittstelle definiert. Mit MPEG-7 soll eine standardisierte Beschreibung bzw. Attributierung verschiedener Arten multimedialer Informationen erreicht werden. Diese Attribute werden mit dem Content selbst verknüpft und dem Anwender eine schnelle, effiziente Suche nach den gewünschten Inhalten ermöglichen.

2.1 MPEG-1

Die Moving Picture Experts Group hat mit MPEG-1 im Jahre 1993 einen Standard definiert, der zur Komprimierung von kombinierten audio-visuellen Daten dient (Abbildung 5). Als Ein- und Ausgangssignal wird ein CCIR-601-Format (SIF) vorausgesetzt. Dieses Format hat eine aktive Bildfläche von 354*288 Bildpunkten und eine Bildwiederholfrequenz von 25 Hz. Es bietet den wahlfreien Zugriff auf das Bildmaterial, das durch Schlüsselbilder (Inter-Frame-Coding) im Datenstrom ermöglicht wird. Da Video und Audio zeitabhängige Medien sind, ist es möglich, Kompression dadurch zu erreichen, indem anstatt der einzelnen Bilder nur die Unterschiede zwischen aufeinander folgenden

Bildern abgespeichert werden. Außerdem kann eine Verringerung des Informationsgehalts durch die schnelle Bildfolge wettgemacht werden. Es besteht also keine Notwendigkeit, alle digitalen Informationen zu kodieren. Der MPEG-Standard definiert drei Arten von Bildern: Intra Pictures (I-Bilder), Predicted Pictures (P-Bilder) und Bidirectional Pictures (B-Bilder).

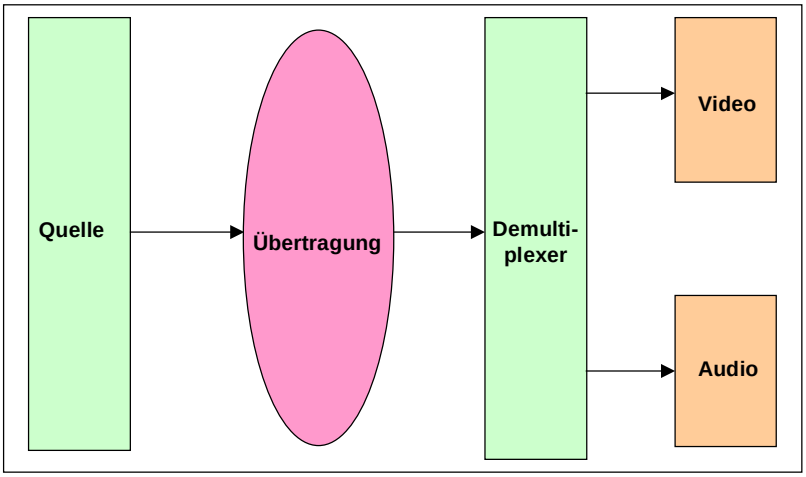

Abbildung 5: MPEG-1

Die I-Bilder werden wie ein Bild im JPEG-Format abgespeichert (Blockzerlegung, DCT, Quantisierung, Lauflängen-Kodierung, Huffman-Kodierung). Die Datenreduktionen werden durch die Kodierung der Lauflängen oder durch Umkodierung nach Häufigkeit erreicht. Bei allmählicher Verbesserung des Bilds wird die Diskrete Cosinus Transformation eingesetzt. Dabei werden für ein Bild mehrere Werte übersandt, bis eine gute Qualität entstanden ist. Es werden im Schnitt 2 Bit pro Pixel verbraucht.

Die P-Bilder nehmen Bezug auf ein vorhergehendes I-Bild oder P-Bild. Man spricht von Forward Prediction. Die Kompressionsrate bei P-Bildern ist ungefähr dreimal so groß wie bei I-Bildern. Die Kompression erfolgt nach folgendem Schema: Hat ein Makroblock in einem P-Bild Ähnlichkeit mit einem Makroblock in einem vorhergehenden I-Bild, so wird der Makroblock des P-Bildes dadurch komprimiert, daß lediglich die Differenz zum ähnlichen Makroblock im I-Bild kodiert wird. Diese Differenz wird als Prediction Error bezeichnet. Die zueinander ähnlichen Makroblöcke in beiden Bildern müssen aber nicht an der genau gleichen Stelle sein. Deshalb wird neben der inhaltlichen Differenz auch die räumliche Differenz der beiden Makroblöcke kodiert. Diese Differenz wird als Motion-Vektor bezeichnet.

Die B-Bilder benutzten sowohl ein vorhergehendes als auch ein nachfolgendes Bild als Referenz. Diese Technik wird als Bidirectional Prediction bezeichnet. Prinzipiell sind vier Arten der Kodierung für Makroblöcke in B-Bildern möglich: JPEG-Standard, Forward Prediction, Backward Prediction und Bidirectional Prediction. Bei der Backward Prediction dient das nachfolgende I-Bild oder P-Bild als Referenz. Bei der Bidirectional Prediction werden sowohl ein Makroblock aus dem vorhergehenden Bild als auch ein Makroblock aus dem nachfolgenden Bild als Referenz für die Kodierung eines Makroblockes im tatsächlichen Bild benutzt.

MPEG-1 arbeitet mit einer Datenrate im Bereich von ca. 1,0 bis 3,0 MBit/s und ist mit Videorecorderqualität vergleichbar.

2.2 MPEG-2

Mit MPEG-2 wurde 1995 eine Erweiterung für neue Videoanwendungen (u.a. Fernsehverteildienst, Multimedia-Kommunikationsdienst) definiert (Abbildung 6). Dieser Standard ist ein Generic-Standard, der sowohl hinsichtlich der Kodiermethoden erweiterbar ist, als auch mit einer Anzahl von Kodiermethoden vorgegeben wurde. Eine wesentliche Änderung ergibt sich aus der Forderung, nicht progressiv abgetastete Sequenzen ebenfalls verarbeiten zu können. Dabei sollte für unterschiedliche Abtastraten (z. B.: schneller Bildvorlauf) bei einer hohen Qualität nicht einfach jedes zweite Halbbild weggelassen werden. Ein weitere Neuerung ist die Möglichkeit, Bitströme zu erzeugen, die in mehreren Qualitäten bzw. Auflösungen dekodierbar sind.

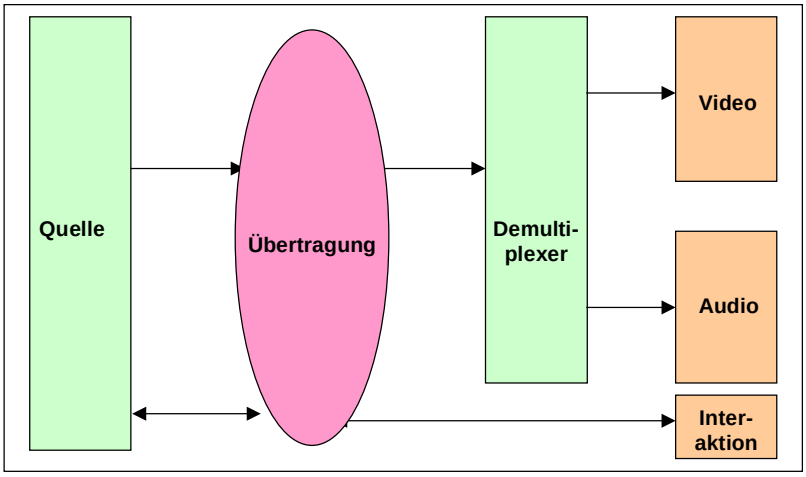

Abbildung 6: MPEG-2

Über mehrere Profile (Low, Main, High1440 und High) wurden unterschiedliche Qualitätsstufen von Videorecorder- über PAL- bis zur HDTV-Qualität festgelegt (Tabelle 7). Im Profil „Low“ werden Kodierung und Dekodierung vereinfacht, indem die Benutzung von B-Bildern ausgeschlossen wird. Die weiteren Profile unterscheiden sich in der An- und Abwesenheit von B-Bildern, der Auflösung der Chrominazkomponenten und der Übertragungsrate.

Level	Max. Auflösung in Pixeln	Bilder/s	Max. Bitrate in MBit/s	Anwendung
Low	352 x 288	25	< 4	Consumer-Video
Main	720 x 576	25	< 15	PAL, Studio-Qualität
High 1440	1440 x 1152	25	< 60	HDTV
High	1920 x 1152	25	< 90	Kinofilm

Tabelle 7: MPEG-2-Profile

MPEG-2 arbeitet zwischen den Datenraten von ca. 1,5 bis ca. 100 MBit/s. Das Format MPEG-2 beinhaltet das Format MPEG-1.

Über die Quellenkodierung hinaus definiert der MPEG-2-Standard auch die Zusammenfassung von mehreren Audio-, Video- und Datensignalen zu einem gemeinsamen Multiplex. Hierbei werden zunächst die Audio-, Video- und Zusatzdaten einzeln in relativ große Einheiten (Packets) unterteilt und mit den erforderlichen Steuerinformationen (Packetizer) versehen. Anschließend erfolgt die Zusammenführung zu einem Datenstrom: entweder zu einem Programm- oder Transport-Multiplex. Beim Programm-Multiplex haben alle Teildatenströme eine gemeinsame Zeitbasis, und die Packets besitzen

eine variable Länge. Beim Transport-Multiplex sind mehrere verschiedene Zeitbasen möglich, und die Packets weisen eine feste Länge von 188 Byte auf. Dadurch können Fehler, die bei Verlust oder Beschädigung eines Packets auftreten, leichter kontrolliert werden.

2.3 MPEG-Audio

Die wichtigste Datenform für Audio ist die Pulse-Code-Modulation. Eine besondere Variation der Technik ist die Adaptive Delta PCM (ADPCM). Bei der Delta Modulation wird eine Datenreduktion durch die Übertragung der Differenz zum vorhergehenden Wert (Signal) erzielt. Bei ADPCM wird eine Datenverringerung durch Vorausberechnung der Differenz des vorhergehenden Wertes (Signals) erreicht.

Die MPEG-Audiokodierung und -komprimierung geschieht durch Reduzierung von nicht hörbaren Signalanteilen. Der Vorläufer der Audiokodierung bei MPEG ist MUSICAM. Bei MPEG-1 und -2 existieren die drei verschiedenen Layer I, II und III. Alle Layer bei MPEG-1 unterstützen bis zu zwei Kanälen sowie die Abtastraten von 48, 44.1 und 32 kHz. Bei MPEG-2 werden auch Surround-, Vielkanal- und Vielsprachen-Systeme berücksichtigt. Diese Kodierung ist abwärtskompatibel, so daß bei einer Stereo-Signal-Abtastung volle Verständlichkeit besteht. Weiterhin arbeitet man an nicht-abwärtskompatiblen Verfahren, die eine bessere Komprimierung und bessere Qualität erlauben (Advanced Audio Canal). Bei MPEG-2 wird durch Halbierung der Abtastfrequenz (16, 22.05, 24 kHz) eine effektivere Komprimierung erzielt.

MPEG-1 und -2 bieten mit höherem Layer niedrigere Bitraten und bessere Qualitäten. Layer I bietet eine Bitrate zwischen 128 und 384 Kbit/s. Im subjektiven Hörtest entspricht eine Kodierung bei einer Rate von 192 kbit/s pro monophonem Kanal CD-Qualiät. Layer II bietet eine Bitrate zwischen 64 und 384 kbit/s. Im subjektiven Hörtest entspricht eine Kodierung bei einer Rate von 128 kbit/s pro monophonem Kanal CD-Qualiät. Layer III bietet eine Bitrate zwischen 8 und 256 kbit/s. Im subjektiven Hörtest entspricht eine Kodierung bei einer Rate von 64 kbit/s pro monophonem Kanal CD-Qualiät. Der Begriff MP3 bezeichnet die MPEG Layer III Kodierung.

2.4 MPEG 4

Im Jahre 1997 begann man MPEG-4 zu spezifizieren. Ende 1998 wurde der neue internationale Multimedia-Standard MPEG-4 (Version 1) verabschiedet. MPEG-4 will keine spezielle Anwendung adressieren, sondern möglichst viele Anwendungsbereiche abdecken (Abbildung 8). Die charakteristischen Neuerungen waren die Integration des Standards H.263 zur Videokodierung, die inhaltsorientierte Interaktivität und ein universeller Zugriffmechanismus. Die Funktionalitäten, die MPEG-4 zusätzlich unterstützen soll, können in die drei Teilbereiche inhaltsbezogene Interaktivität, Kompression und Skalierbarkeit zusammengefaßt werden.

Bei der „inhaltsbezogenen Interaktivität“ soll 1. Ein schneller Zugriff auf die audiovisuellen Daten von audiovisuellen Objekten, z. B.: bei Indizierung, Hyperlinking, Laden bzw. Runterladen und Löschen, garantiert werden; 2. Ein Syntax- und Kodierungsschema bereitgestellt werden, so daß spezielle Objekte einer Szene verändert werden können, ohne daß die digitale Darstellung derselben bekannt ist; 3. Eine Methode bereitgestellt werden, um künstliche mit natürlichen Szenen zu verknüpfen - dies wäre ein erster Schritt zur Vereinheitlichung jeglicher audiovisuellen Information - ; sowie 4. Der zufällige Zugriff auf Teile einer audiovisuellen Sequenz ermöglicht werden.

Bei der „Kompression“ soll 1. eine bessere Qualität bei vergleichbarer Bitrate gegenüber schon bestehenden Standards erreicht werden; sowie 2. sollen mehrere Blickrichtungen oder Tonspuren einer festen Szene kodiert und später wieder synchronisiert werden können, so daß es möglich wird, auch 3D-Objekte darzustellen. MPEG-4 soll dabei die Überschneidungen der unterschiedlichen Blickrichtungen bei der Kodierung ausnützten.
Bei der „Skalierbarkeit“ ist eine räumliche als auch eine zeitliche Auflösung zu skalieren. Außerdem sollen auch einzelne Objekte einer Szene skaliert werden können. Das heißt, daß bestimmte Objekte stärker hervorgehoben werden können bzw. mit stärkerer Auflösung dargestellt werden können.

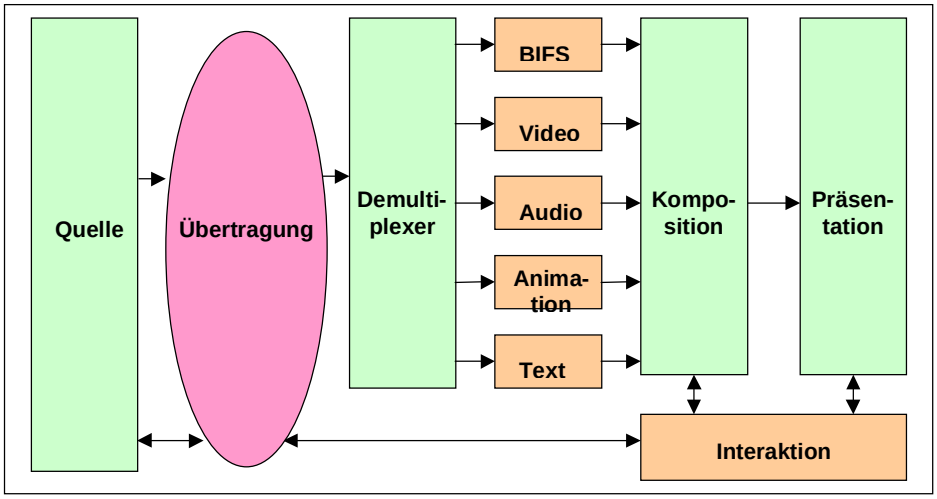

Abbildung 8: MPEG-4

2.4.1 MPEG 4 - Auto-visuelle Szenen

Der MPEG-4 Standard soll die Möglichkeit zur Interaktion mit audio-visuellen Szene-Inhalten vorsehen und die Komposition von audio-visuellen Objekten in eine kohärente Präsentation nun nicht mehr beim Sender, sondern beim Empfänger durchführen. Zur Erfüllung dieser Aufgabe spezifiziert MPEG-4 lediglich ein binäres Format zur Szenenbeschreibung, das "Binary Format for Scene description (BIFS)", das auf einer Erweiterung der "Virtual Reality Modeling Language (VRML)" basiert. Für den Prozeß der Komposition selbst wurden bisher keine normativen Festlegungen getroffen.
Abbildung 9 zeigt ein Modell des MPEG-4-Terminals. Die Elemente der Systemspezifikation sind die Objektbeschreibung, die Szenenbeschreibung, die Synchronisatonsebene und Multiplexebene. Sie bilden einen Rahmen, in den die Dekoder, die einzelne Elementarströme mit Daten für audio-visuelle Objekte verarbeiten, eingebettet sind. Das Transportformat der Datenströme ist ebenso wie die Komposition nicht Bestandteil der Systemspezifikation.
Für eine Videokodierung wird die Szene daher in audio-visuelle Objekte (AVO) eingeteilt, denen ein Video Object Plan (VOP) zugrunde liegt. Ein solches Objekt kann natürliche, synthetische, zwei- oder dreidimensionale, mono-, stereo- oder multi-Sichten beinhalten. Der VOP entspricht den Teilen einer Szene, die der Nutzer manipulieren kann.
Die Szenenbeschreibungssprache BIFS hat, wie auch VRML, mehrere Facetten. Zum einen ist es möglich, hörbare und visuelle Objekte auf hierarchische Weise zu gruppieren und in einem zwei- oder dreidimensionalen Raum, aber auch in der Zeit, zu positionieren. Die Möglichkeiten von VRML

wurden u. a. durch Elemente für zweidimensionale Anwendungen, aber auch durch Elemente zur Integration von zweidimensionalen und dreidimensionalen Szenenteilen erweitert.
Ein weiterer wesentlicher Bestandteil von BIFS ist die Möglichkeit, das Verhalten der gesamten Präsentation zu beschreiben, z. B. in Reaktion auf Benutzereingaben, aber auch einfach über den Verlauf der Zeit. Das Präsentationsverhalten umfaßt z. B. Änderungen von Positionen, Farben, Lautstärken einzelner oder auch ganzer Gruppen von audio-visuellen Objekten. Zusätzlich erlaubt BIFS die Darstellung von graphischen Elementen, angefangen von Linien bis hin zu komplexen Polygonzügen.

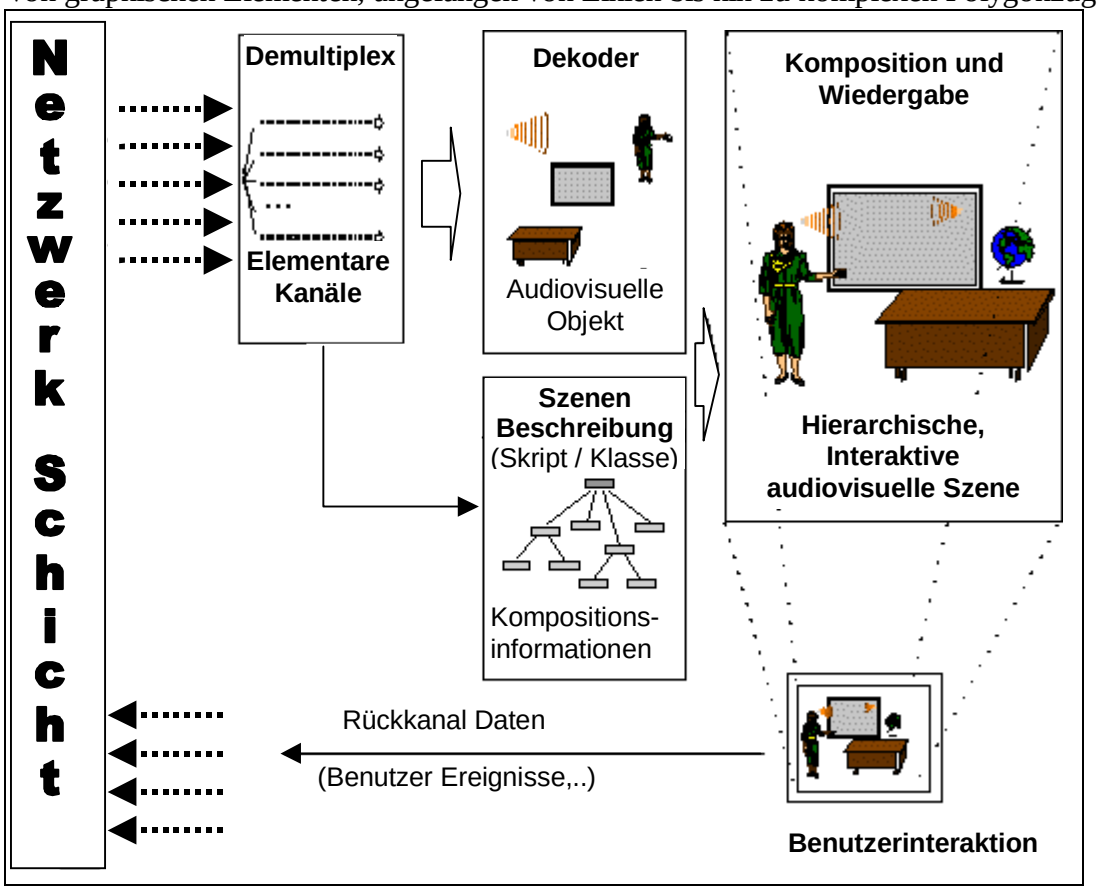

Abbildung 9: MPEG-4 Terminalmodell.

Architektonisch ist BIFS insofern flexibler als sein Vorgänger, als die Beschreibungsdaten prinzipiell in Form eines Datenstromes übermittelt werden, der es erlaubt, den Inhalt einer Szene im Laufe der Zeit zu ändern, Elemente hinzuzufügen oder zu entfernen. Für die effiziente gleichzeitige Manipulation vieler Parameter in einer Szene von Seiten der Datenquelle wurde außerdem ein sogenanntes BIFS-Animationsformat definiert.
Die Struktur eines Szenengraphen ist in Abbildung 10 beispielhaft dargestellt. Es enthält ein natürliches audiovisuelles Objekt, eine animierte synthetische Darstellerin, ein aus Text generiertes Sprachereignis (TTS - text to speech) und weitere synthetische graphische Elemente.

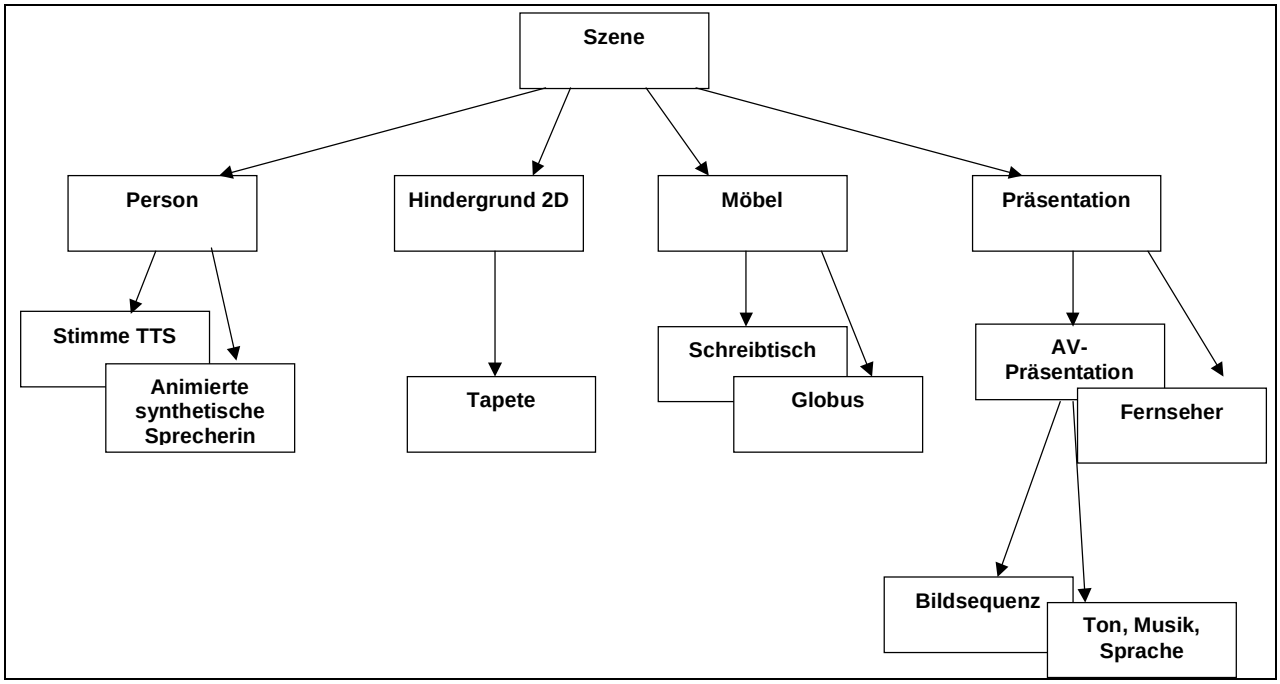

Abbildung 10: Prinzipielle Struktur eines Szenengraphen in einer BIFS-Szenenbeschreibung.

Durch die Aufspaltung von audiovisuellen Präsentationen in einzelne Objekte können dies verschiedenartigen Terminals in heterogenen Netzen effizient erreicht werden. Zum einen kann ggf. für jedes Terminal ausgewählt werden, welche Objekte dargestellt werden, d.h., die Leistungsfähigkeit der Darstellungseinheit im Terminal kann berücksichtigt werden. Hierfür muß ein Terminal zwar die gesamte Szenenbeschreibung einer Präsentation auswerten, kann dann jedoch entscheiden, bestimmte Teile der Szene nicht darzustellen. Unabhängig davon kann aber auch bereits während der Übertragung der Präsentation, die ja nun in Form einzelner Datenströme erfolgt, entschieden werden, daß nur eine Teilmenge der Gesamtdaten übertragen wird, wenn beispielsweise die Datennetzanbindung des Empfängerterminals nicht genügend Bandbreite aufweist.

Während die Szenenbeschreibung in erster Linie Bezüge zwischen audio-visuellen Objekten in Raum und Zeit wiedergibt, erlauben zusätzliche Objektdeskriptoren die separate Beschreibung der beteiligten Datenströme und ihrer Zusammenhänge. Hiermit kann beispielsweise signalisiert werden, daß für ein bestimmtes visuelles Objekt verschiedene Repräsentationen zur Verfügung stehen, deren Datenströme sich in ihrer Datenrate unterscheiden. Auch eine hierarchische Kodierung kann so angezeigt werden. Weiterhin können Datenströme priorisiert werden. Diese Informationen können von Interworking Units (IWU), die verschiedene leistungsfähige Teilnetzwerke miteinander verbinden, als Hinweis verwendet werden, um lediglich eine ausgewählte Teilmenge von Datenströmen weiterzuleiten.

Der wesentliche Unterschied zu ähnlichen Einheiten, die auch heute schon zur Transkodierung zwischen verschiedenen Videokonferenzstandards (z.B. H.320 und H.324) eingesetzt werden, ist, daß die Inhalte der Datenströme nicht dekodiert und wieder rekodiert werden müssen. Eine IWU für ein rein auf MPEG-4 basiertes audio-visuelles Kommunikationssystem in einem heterogenen Netz kann also weniger aufwendig konzipiert werden als bei den genannten existenten Standards.

Abbildung 9 zeigt auch die Schichtenstruktur von MPEG-4. Die Kompressionsschicht ist für die Übersetzung zwischen komprimierter und expandierter Darstellung der audiovisuellen Objekte verantwortlich und enthält auch die besprochenen Funktionalitäten zur Szenenbeschreibung und die Objektdeskriptoren. Darunter befindet sich die Synchronisationsschicht (sync layer). Diese Datenstromspezifikation beschränkt sich darauf, einzelne Datenpakete, sogenannte Zugriffseinheiten (access units), zu identifizieren und mit Zeitmarken zu versehen. Die Transportschicht ist für das Verschachteln von Paketen der verschiedenen Datenströme in einen oder mehrere serielle Multiplexdatenströme verantwortlich. Als Hilfsmittel zur Anpassung an Netzwerke mit großen Paketlängen oder hohen Kosten für die Einrichtung von Transportkanälen wird lediglich ein einfaches Multiplexverfahren vorgeschlagen, genannt FlexMux, das der möglichen hohen Anzahl von Datenströmen mit geringer Datenrate in MPEG-4-Anwendungen Rechnung trägt und bei Bedarf Verwendung finden kann.

Im Rahmen der MPEG-4-Systemspezifikation wird ein generisches Interface zur Transportschicht betrachtet, das DMIF Application Interface (DAI). DMIF, das Delivery Multimedia Integration Framework, ist ein weiterer Teil der MPEG-4 Standardserie, die sich mit der konkreten Spezifikation dieses Layers befaßt. Für jedes mögliche Transportnetzwerk, sei es ein IP Netzwerk, ein DVB Digital-TV-Netz oder ein ATM-Netz, ist eine Adaption zu spezifizieren, um tatsächlich Kommunikation in heterogenen Netzwerken zu ermöglichen.

2.4.2 MPEG-4 - Visuelle Kompression

Der MPEG-4-Standard erlaubt die Kodierung von natürlichen Bild- und Videosignalen zusammen mit synthetischen, im Computer generierten Videodaten. Die Gesamtheit von natürlicher und synthetischer Videoinformation wird in diesem Zusammenhang als "visuelle Information" gekennzeichnet. Der Standardteil zur "visuellen Kodierung" umfaßt neben der Definition eines Videokodierers noch die Beschreibung eines Kopf- und Gesichtsmodells nebst Animationsparametern, sowie die Definition von zweidimensionalen Gittern zur Animation von Texturen.

2.4.2.1 Funktionalitäten des Videoteils

MPEG-4 ist ähnlich wie bereits die Vorgänger MPEG-1/-2 genetisch ausgelegt, d.h. nicht auf eine bestimmte Applikation zugeschnitten. Die Datenrate für die Videodaten liegt typischerweise zwischen 5 kbit/s und 4 Mbit/s, und es werden verschiedene Bildformate von sub-QCIF bis zu TV-Auflösung unterstützt. Die Abtastung kann dabei sowohl progressiv als auch im klassischen Zeilensprungverfahren (interlaced) erfolgen. Um sich an unterschiedliche Kanalkapazitäten anpassen zu können, bietet MPEG-4 die Möglichkeit der Skalierbarkeit, d.h. Videodaten können hierarchisch in einer Weise kodiert werden, die es dem Dekoder ermöglicht, nur einen Teil der Gesamtdaten mit einer entsprechend geringeren Bildqualität auszuwerten. Eine Skalierung kann dabei wahlweise durch Variation der örtlichen oder der zeitlichen Auflösung erfolgen.

Ein Kernmerkmal von MPEG-4 ist die Möglichkeit zur Kodierung von Videoobjekten, die nicht notwendigerweise rechteckig sein müssen, sondern eine beliebige Form aufweisen können. Zur Beschreibung dieser Daten wird das Prinzip der "Video Object Plane (VOP)" eingeführt. Ein VOP stellt einen örtlichen Ausschnitt des Videoobjekts zu einem bestimmten Zeitpunkt dar und entspricht damit dem Einzelbild einer klassischen Videosequenz - allerdings mit dem Unterschied, daß das VOP belie-

big geformt sein kann. Um diese Form dem Empfänger mitteilen zu können, besitzt MPEG-4 die Möglichkeit zur Konturkodierung.
Zusätzlich unterstützt MPEG-4 die Übertragung von Videodaten über fehleranfällige Kanäle, wie sie z. B. typisch für Mobilfunknetze sind. Da solche Kanäle in der Regel geringe Bandbreiten besitzen, muß ein geeigneter Kompromiß zwischen Kompression einerseits und Fehlerschutz andererseits gefunden werden.

2.4.2.2 Struktur des MPEG-4-Videokodierers

Abbildung 11 zeigt das grundlegende Konzept, mit dem ausgehend von bekannten Kodiertechniken eine objektbasierte Videokodierung ermöglicht wird. Der MPEG-4-Video Standard setzt sich zusammen aus einem Basiskodierer und einem darauf aufsetzenden generischen Videokodierer.
Der Basiskodierer entspricht im wesentlichen der konventionellen Hybridkodierung, bei der die beiden Komponenten Bewegungsinformation und Textur bzw. Prädiktionsfehler übertragen werden. Das Eingangsmaterial dieses Basiskodierers ist eine Videosequenz rechteckigen Formats. Aufbauend hierauf wird nun die Kodierung von beliebig geformten Videoobjekten dadurch ermöglicht, daß dem Basiskodierer eine Formkodierung vorgeschaltet wird und alle nachfolgenden Verarbeitungsschritte auf den gegebenen Bildausschnitt beschränkt werden.

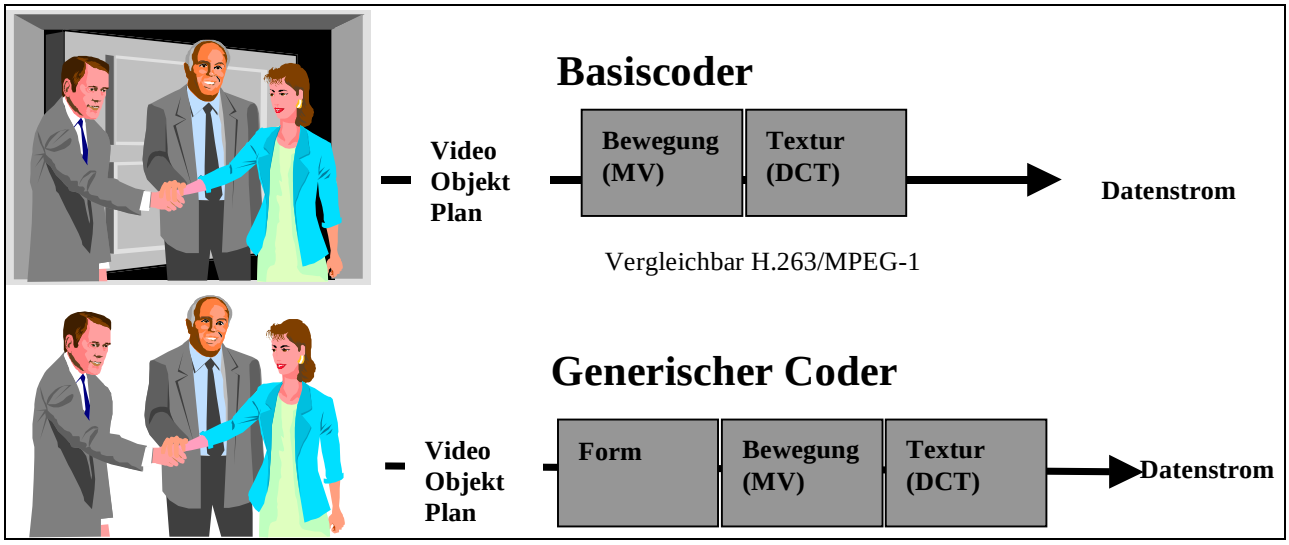

Abbildung 11: MPEG-4-Stufenkonzept des konventionellen Hybridkodierers ergänzt um eine Formkodierung.

Formkodierung: Die Form eines Videoobjekts wird durch eine Binärmaske repräsentiert, die angibt, ob ein Bildpunkt zu einem Videoobjekt gehört oder nicht. Die Kodierung der Objektform entspricht daher einer Kodierung dieser Binärmaske, wobei eine approximative Formbeschreibung durch eine Unterabtastung der Binärmaske vor der eigentlichen Kodierung realisiert wird. Zur Kodierung die Maske wird ein kontextadaptives arithmetisches Verfahren verwendet, bei dem auch zwischen Intra- und zeitlich prädikativer Kodierung umgeschaltet werden kann. Für TV-Anwendungen läßt sich zudem Transparenz kodieren. In diesem Fall trägt die Formmaske keine Binärwerte, sondern eine mit acht Bit aufgelöste Transparenzinformation, deren Kodierung identisch mit der weiter unten beschriebenen Texturkodierung erfolgt.

Bewegungsschätzung und Kompensation: MPEG-4 teilt das Videoobjekt in ein regelmäßiges Blockraster und verwendet dann blockbasierte Bewegungsschätzung und -kompensation, um zeitliche Korrelationen zwischen aufeinanderfolgenden VOPs auszunutzen. Die geschätzten Bewegungsvektoren und der verbleibende Prädiktionsfehler werden ähnlich wie in MPEG-1/-2 oder H.261/263 kodiert, wobei zwischen 8x8-Blöcken und 16x16-Makroblöcken umgeschaltet werden kann, und auch eine Bewegungskompensation mit überlappenden Nachbarblöcken (overlapped block motion compensation) möglich ist. Aufgrund der möglicherweise nicht rechteckigen Form des Bildobjekts ergeben sich allerdings zwei entscheidende Besonderheiten: Zum einen muß das Videoobjekt vor der Bewegungskompensation im Außenbereich mit Bildinformation ergänzt werden, um auch auf dem Rand des Objekts liegende Bildblöcke als sinnvolle Referenz verwenden zu können. Zum anderen wird bei der Bewegungsschätzung für die Randblöcke ein sogenanntes Polygonmatching verwendet, bei dem für die Bestimmung eines geeigneten Bewegungsvektors nur diejenigen Bildpunkte innerhalb des Blocks herangezogen werden, die gleichzeitig auch zum Bildobjekt gehören. Grundsätzlich kann wie in anderen Videokodierstandards auch - bei MPEG-4 zwischen reiner Intra-Frame-Kodierung und vorwärts- oder bidirektional-prädizierten VOPs umgeschaltet werden.

***Texturkodierung*:** Die intraframe-kodierten VOPs wie auch der Prädiktionsfehler nach der Bewegungskompensation werden unter Verwendung der DCT für 8x8-Blöcke kodiert. In einem Makroblock werden dabei vier benachbarte 8x8-Luminanzblöcke und zwei Chrominanzblöcke zusammengefaßt. Ähnlich zur Bewegungskompensation ist auch hier eine besondere Anpassung für die Randblöcke eines Videoobjektes erforderlich. Die in einem Randblock fehlende Bildinformation wird dazu vor der DCT-Kodierung so ergänzt, daß bei einer blockweisen Transformation keine unerwünschten hochfrequenten Spektralanteile auftreten. Alternativ zur Signalergänzung soll in Version 2 des Standards auch eine spezielle formangepaßte DCT (SADCT) verwendet werden, mit der sich insbesondere bei hohen Qualitätsanforderungen weitere Gewinne in der Bildgüte erzielen lassen.

Multiplexer: Die komprimierte Formmaske, die geschätzten Bewegungsvektoren und die DCT-Koeffizienten des Restfehlerbildes mit den zugehörigen Lauflängen werden in einen Datenstrom gemultiplext. Dabei wird im wesentlichen die Makroblock-Syntax aus H.263 verwendet, mit den dort enthaltenen Codewort-Tabellen für DCT-Koeffizienten und Bewegungsvektoren. Diese enge Anlehnung an die H.263-Syntax ermöglicht zudem Rückwärts-Kompatibilität von MPEG-4 mit dem ITU-Standard, d.h., jedes MPEG-4-Terminal kann auch H.263-Datenströme im Basismodus dekodieren. Darüber hinaus gibt es eine spezielle Multiplex-Syntax, die einen besonders fehlerrobusten Datenstrom beispielsweise für Videokommunikation über Mobilfunknetze generiert. Dazu können regelmäßig Synchronisationsworte in den Datenstrom eingefügt werden. Ein zusätzlicher Modus erlaubt die Gruppierung von Bewegungsvektoren und DCT-Koeffizienten mit entsprechenden Synchronisationsmarkern im Datenstrom, so daß die wichtige Bewegungsinformation besser erkannt und gegebenenfalls auch geschützt werden kann. Dazu existiert eine spezielle Tabelle mit rückwärts dekodierbaren Codewörtern, so daß im Fall eines Übertragungsfehlers fehlende Werte auch ausgehend von einem detektierten Marker rückwärts entschlüsselt werden können.

2.4.3 MPEG-4 - Audio Kompression und Repräsentation

Unter MPEG-4-Audioverarbeitung versteht man die Integration von Sprach- und Audiokodierung natürlicher Signale, von synthetischer Sprach- und Klangerzeugung sowie der Komposition aller

Audioinhalte (Objekte) zu einem Klangbild. Ein Audioobjekt ist ein hörbares Ereignis, das zum Beispiel mit einem oder mehreren Mikrofonen (Mono oder Mehrkanal) aufgenommen sein kann. Je nach Audioobjekt und der erwünschten Übertragungsqualität kann eine unterschiedliche Kodierung gewählt werden. Eine Bahnhofsszene könnte zum Beispiel aus dem Dröhnen eines einfahrenden Zugs im Mehrkanal-Format, eines Gesprächs im Stereo-Format, einer Bahnhofsdurchsage in Telefonqualität und synthetischer Hintergrundmusik bestehen. Die mit MPEG-4 übertragene Bahnhofsdurchsage kann dann im Abspielgerät (Terminal) nachbearbeitet und aus unterschiedlichen Richtungen wiedergegeben werden.

2.4.3.1 Kodierung von natürlichen Audioobjekten

Die MPEG-Audio-Untergruppe hatte sich zum Ziel gesetzt, mit MPEG-4 deutlich kleinere Datenraten für die Repräsentation von natürlichen Signalen zu verwenden, als es bisher mit den MPEG-1 und MPEG-2 Layer-3 Standard möglich war.

Parallel zu diesen Aktivitäten wurde noch am MPEG-2 Standard Teil 7 - Advanced Audio Kodierung (AAC) - gearbeitet. Dieser kann bis zu 48 Kanäle (ch) mit 64 kbit/s/ch kodieren, ohne daß hörbare Unterschiede zum Originalsignal (ca. 700 kbit/s/ch) wahrnehmbar sind. AAC braucht damit ca. die halbe Datenrate für eine vergleichbare Qualität gegenüber MPEG-1 und -2 Layer II.

AAC wurde dann als hochwertiges Kodierungsverfahren ebenfalls für MPEG-4 übernommen. Für niedrigere Datenraten mit "FM"-, "AM"- oder "Telefon"-Qualität werden zusätzlich neue Algorithmen verwendet. Die durchgeführten Tests zeigten deutlich, daß die Qualität nicht nur von der Datenrate, sondern auch von der Art des Eingangssignals abhängt:

- Komplexe Signale wie Pop- oder Orchestermusik können mit Zeit/Frequenz-Verfahren (T/F Kodierung), zu denen auch AAC gehört, mit Datenraten von ca. 24 bis 64 kbit/s/ch kodiert werden.
- Bei diesen Verfahren wird das Zeitsignal in den Frequenzbereich transformiert und der redundante Anteil durch Prädiktionsverfahren aus dem Spektrum entfernt. Anschließend werden die Spektralkoeffizienten unter psychoakustischen Gesichtspunkten quantisiert und kodiert.
- Sprachsignale verlangen dagegen entweder spezielle CELP-Sprachkodierer (Code Excited Linear Prediction), die mit Datenraten von 6 bis 24 kbit/s arbeiten, oder sprachoptimierte parametrische Verfahren für sehr niedrige Datenraten von 2 oder 4 kbit/s.
- Bei der CELP-Kodierung werden LPC-Koeffizienten (Linear Predictive Coding) ermittelt und nach einer Umformung skalar oder vektoriell quantisiert. Das LPC- Restsignal wird durch einen Index auf einen Codebucheintrag ersetzt. MPEG-4 erlaubt hierbei eine feine Skalierung der Datenrate (FineRateControl) und Abtastraten von 8 kHz und 16 kHz.
- Beim parametrischen HVXC-Verfahren (Harmonic Vector eXcitation Coding) wird das Restsignal zusätzlich noch transformiert und anschließend parametrisiert.
- Für klare, harmonische Klänge eignen sich besondere musikoptimierte parametrische Verfahren (ab 4 kbit/s).
- Beim parametrischen HILN-Kodierer (Harmonic and Individual Line plus Noise) werden harmonische und dominierende Frequenzanteile sowie Rauscheigenschaften extrahiert und übertragen.

Bei beiden parametrischen Verfahren lassen sich Tonhöhe und Abspielgeschwindigkeit im Dekoder unabhängig voneinander verändern.

In Bild 12 ist der Einsatzbereich der einzelnen Verfahren, bezogen auf Datenrate und Audiobandbreite, grafisch dargestellt. Der obere Balken zeigt hier den skalierbaren Audiokodierer, der den Datenratenbereich lückenlos abdeckt. Dazu erzeugt eine erste Kodierung und Dekodierung einen Basis-

bitstrom mit geringer Datenrate. Im zweiten Schritt wird die Differenz aus dem dekodierten und dem Originalsignal gebildet und anschließend daraus ein Erweiterungsbitstrom generiert.
In Version 2 des MPEG-4-Standards soll ein speziell auf die Audiokodierung angepaßtes Fehlerschutzverfahren eingeführt werden. Das Fehlerschutzverfahren soll wichtige Informationen besser sichern (unequal error protection), um den Standard auch für fehleranfällige Übertragungskanäle attraktiv zu machen.

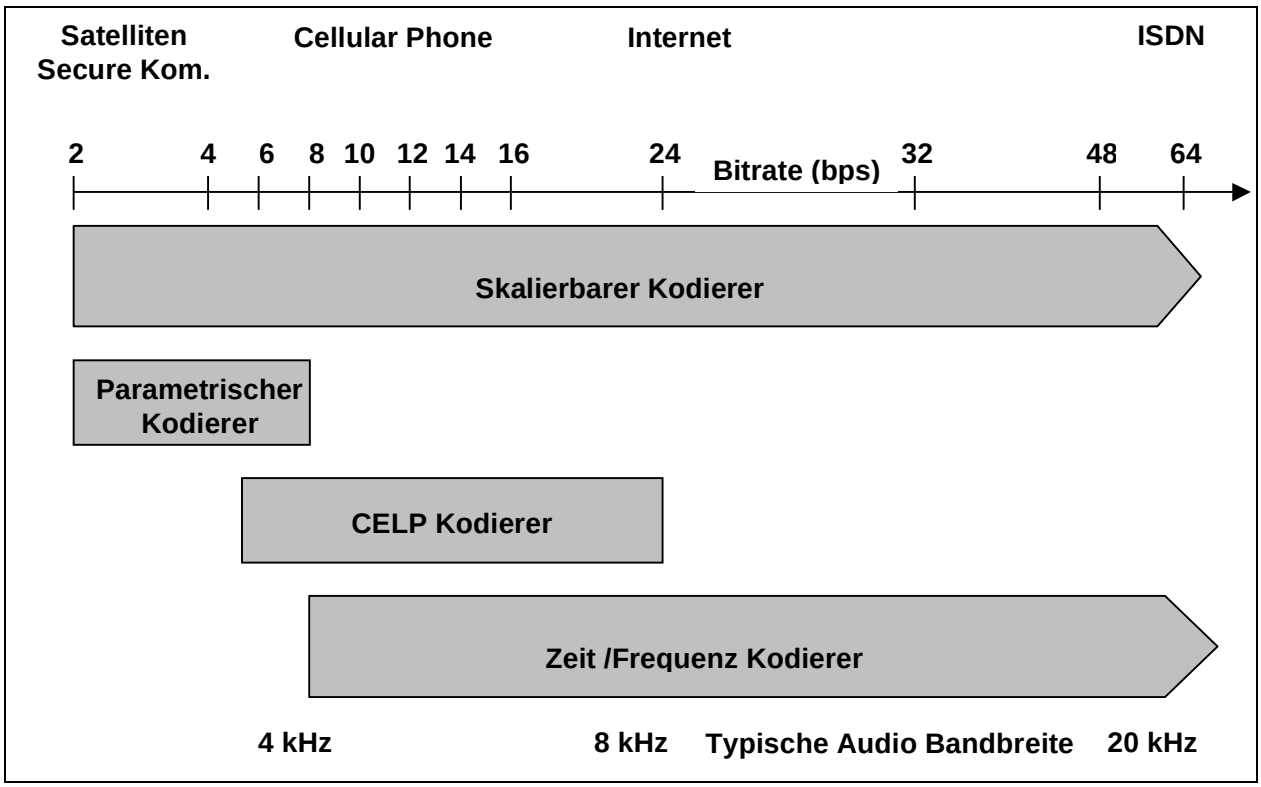

Bild 12 Übersicht über MPEG-4 Audio Kodierungsverfahren.

2.4.3.2 Übertragung von synthetischer Sprache und Klängen

Neben natürlichen Audiosignalen können in MPEG-4 auch Sprache und Klänge synthetisch erzeugt werden. Es war innerhalb von MPEG-4 unmöglich, für jede Sprache dieser Welt ein TTS-Verfahren (Text-To-Speech) zu standardisieren. Deshalb wurde ein Interface definiert, um Texte, prosodische Informationen und Lippenformen zu repräsentieren. Prosodische Informationen beinhalten linguistische Angaben, wie Satzintentionen (Aussage-, Frage-, Befehlssatz), Akzente, Pausen, Silbenrhythmus. Die Lippenformen können u. a. zur Ansteuerung von graphisch animierten Gesichtern benutzt werden. Die Umsetzung des Textes in die entsprechende Sprache bleibt dem Implementierer überlassen.
Als synthetisches Musikformat wurde unter anderem MIDI (Musical Instrument Digital Interface) mit der DLS-Erweiterung (Downloadable Sounds) gewählt. Zusätzlich wird das "Structured Audio-Format" (SA), eine Erweiterung von CSound, eingesetzt. SA besteht aus einer Instrumentenbeschreibungssprache SAOL (SA Orchestra Language) und dem dazugehörigen Steuercode SASL (SA Score Language). Die SAOL-Sprache enthält, neben den üblichen Befehlen einer Programmiersprache, eine umfassende Bibliothek von ca. 100 vorinstallierten Audio-Funktionen. Der besondere

Vorteil ist, daß hiermit ein Musiker den Klang seiner Instrumente ganz genau spezifizieren kann, was mit MIDI allein nicht möglich ist.

2.4.3.3 Komposition der Audioobjekte zu einem Klangbild

Audio-Szenenkomposition bedeutet, daß ein MPEG-4-Terminal mehrere Audiodekodierausgänge, die jeweils ein sinnvolles Audioobjekt darstellen, zu einem Tonstück arrangiert. Diese "Mischpult-Eigenschaft" wird durch einige spezielle Audio-BIFS-Knoten beschrieben. Zu den Standardverfahren gehören Misch-, Umschalt-, Verzögerungs- und 2D/3D-Funktionen. Es sei an dieser Stelle erwähnt, daß bei unterschiedlichen Abtastfrequenzen alle Audioobjekte auf die höchste vorkommende Abtastfrequenz konvertiert werden. Zusätzlich gibt es noch Effekt-Funktionen, die mit "Structured Audio" gesteuert werden. Damit lassen sich alle denkbaren Nachverarbeitungen, wie Filter (Tiefpass, Hochpass, Equalizer), Amplitudenänderungen (Kompression, Balance), Verzögerungen (Nachhall, Echo) und Effekte (Chorus, Flanger), frei programmieren.

Durch zusätzliche Audio-BIFS-Knoten in Version 2 wird es unter Verwendung von physikalischen Modellen möglich sein, die Raumakustik von möblierten Räumen oder Konzerthallen nachzubilden.

2.4.4 MPEG-4-Profile

Der MPEG-4-Standard läßt sich für typische Anwendungsklassen konfigurieren und damit in der Komplexität und im Leistungsumfang an bestimmte Hardwareplattformen anpassen. Dies erfolgt mit Hilfe von spezifizierten Profilen, wobei jedes Profil in Form von nachgeordneten Ebenen (Levels) feste Rahmen für die freien Koderparameter, wie Datenrate, Bildgröße, Audio-Abtastfrequenz, Anzahl der Szenenobjekte, festlegt. Da die Palette der MPEG-4 Anwendungen noch nicht absehbar ist, können zusätzliche Profile auch nach Abschluß der Standardisierung noch erstellt werden.

Beispielhaft erfolgt hier eine Kurzbeschreibung der bereits spezifizierten Profile, die speziell auf die Kodierung natürlicher Videoobjekte abgestimmt wurde:

- Das Simple Visual Profile bietet effiziente Kodierung von rechteckigen Videoobjekten und die Möglichkeit zur fehlerrobusten Bildübertragung. Es ist damit im wesentlichen für zukünftige mobile Netzwerke, wie UMTS, konzipiert.
- Das Simple Scalable Visual Profile bietet darüber hinaus noch die Möglichkeit zur zeitlichen und örtlichen Skalierbarkeit. Damit eignet es sich für Anwendungen, die auf Grund von Beschränkungen in der Datenrate oder in der Dekoderkomplexität mehr als eine Qualitätsstufe benötigen. Dies ist z. B. bei einem softwarebasierten Videoabruf der Fall.
- Das Core Visual Profile ergänzt zusätzlich noch die Funktionalität der Formkodierung und eignet sich damit im besonderen für Anwendungen, die von inhaltsbasierter Interaktivität profitieren, wie Multimediakommunikation im Internet.
- Das Main Visual Profile erlaubt ergänzend die Kodierung von semi-transparenten Videoobjekten im Zeilensprungverfahren und zielt damit insbesondere auf hochqualitative Fernsehdienste und DVD-Anwendungen.

2.5 MPEG-7

MPEG-7 wurde formal mit „Multimediale Inhaltsbeschriebende Schnittstelle - Multimedia Content Description Interface“ definiert. Im Oktober 1998 erfolgte der Call for Proposals, Dezember 1999 wird der erste Entwurf erstellt, der internationale Standard ist für September 2001 geplant. MPEG-7 beschäftigt sich damit, den Inhalt von audio-visuellen Sequenzen, Bildern und Graphiken in einer effizienten und zweckmäßigen Darstellung zu repräsentieren, so daß damit eine Informationssuche

möglich wird. Das Ziel von MPEG-7 ist es, die zur Zeit existierenden eingeschränkten Möglichkeiten proprietärer Lösungen für die Identifizierung von Content zu erweitern, insbesondere durch die Einbeziehung einer größeren Menge von Datentypen. Mit MPEG-7 soll eine standardisierte Beschreibung bzw. Attributierung verschiedener Arten multimedialer Informationen erreicht werden. Diese Attribute werden mit dem Content selbst verknüpft und werden dem Anwender eine schnelle, effiziente Suche nach den gewünschten Inhalten ermöglichen.
MPEG-7 spezifiziert einen Standard-Satz von Beschreibungs-Elementen, die zur Beschreibung von unterschiedlichen Typen von Multimedia-Informationen benutzt werden. MPEG-7 will neben den Beschreibungs-Elementen auch eine Struktur (Beschreibungsschemen - Description Schemes /DS) und deren Beziehungen untereinander festlegen. Diese Beschreibungen, die eine Kombination aus den Beschreibungs-Elementen und den Beschreibungsschemen sind, können mit dem eigentlichen Inhalt direkt verbunden werden. Dieses erlaubt eine schnelle und effektive Suche der Inhalte nach den Vorstellungen eines Anwenders. MPEG-7 möchte außerdem eine Sprache zur detaillierten Definition der Beschreibungsschemen standardisieren (Description Definition Language /DDL - Beschreibungs-Definitionssprache). Das Material, das mit MPEG-7 Daten ergänzt wurde, kann indexiert und gesucht werden Dieses Material könnte aus folgenden Bestandteilen bestehen: Standbilder, Grafiken, 3D-Modelle, Audio, Sprache, Video und Informationen. Die Informationen charakterisieren die Verknüpfungen und Beziehungen der Elemente in einer Multimedia-Präsentation (scenarios composition information - Information der szenischen Komposition). In bestimmten Fällen könnten diese generellen Datentypen auch Gesichtsausdrücke und persönliche Eigenschaften beinhalten.
Der MPEG-7 Standard baut auf den einzelenen Definitionen, wie analoge Signalverarbeitng, Pulse Code Modulation (PCM), MPEG-1, -2 und -4, auf. Eine Funktionalität des Standards erzeugt Verweise auf mehrfach nutzbare Teile. Zum Beispiel: Ein Form-Beschreibungs-Element aus MPEG-4 könnte in einem MPEG-7 Zusammenhang nützlich sein, und die gleiche Information könnte für den Bewegungsvektoren der DCT-Felder in MPEG-1 und -2 angewendet werden.
Die MPEG-7 Beschreibungs-Elemente sind nicht von dem Verfahren abhängig, wie die zu beschreibenden Inhalte kodiert und gespeichert wurden. Es ist möglich, das ein MPEG-7 Beschreibungs-Element an einem analogen Video oder an einem Bild, das zum Druck auf Papier vorgesehen ist, angehängt wird. Gerade deswegen sind MPEG-7 Beschreibungs-Elemente nicht Abhängig von der kodierten Repräsentation der Inhalte, da MPEG-7 in Zusammenhänge mit MPEG-4 steht. Der MPEG-4 Standard ist mit Werkzeugen ausgestattet, die die audio-visuellen Inhalte als Objekte mit bestimmter Zuordnung in der Zeit (Synchronisation) und Raum (auf dem Bildschirm für Video, in dem akustischen Raum bei Audio) enkodieren. Wird der MPEG-4 Enkoder genutzt, wird es möglich sein, Beschreibungs-Elemente zu Objekten innerhalb einer Szene hinzuzufügen. Das können audio oder visuelle Objekte sein. MPEG-7 möchte unterschiedliche Detaillierungsstufen der Beschreibungs-Elemente erlauben, um die Möglichkeit von verschiedenen Ebenen von Unterscheidungen anbieten zu können
Weil die beschriebenen Merkmale bedeutungsvoll im Kontext der Anwendung sind, werden sie unterschiedlich für verschiedene kulturelle Regionen und bei verschieden Anwendung sein. Dies deutet an, daß das gleiche Material mit verschiedenen Typen von Merkmalen je nach Einsatz im Anwendungsfeld beschrieben werden kann. Im Folgenden wird ein Beispiel für das visuelle Material gege-

ben. Eine niedrigere Abstraktionsebene benötigt Beschreibungs-Elemente, wie Form, Größe, Textur, Farbe, Bewegungsrichtung und Position. Für Audio-Material gibt es Beschreibungselemente, wie Tonart, Stimmung, Klang, Tempo, Tempowechsel und Position im akustischen Raum. Auf der mittleren Ebene können dies Objekte und Strukturen, wie Schauspieler und Melodie der Geige, sein. Die oberste Ebene gibt semantische Information an: "Das ist eine Szene mit einem bellenden braunen Hund auf der linken Seite und einem blauen Ball auf der rechten Seite, der gerade im fallen begriffen ist. Im Hintergrund ist ein Geräusch von vorbeifahrenden Autos zu hören". Alle diese Beschreibungs-Elemente werden natürlich in einer effizienten Sprache kodiert - effizient im Sinne der Suche. Weitere dazwischen liegende Abstraktionsebenen können ebenfalls existieren.

Die Abstraktionsebenen stehen im direkten Zusammenhang mit dem Verfahren, wie die Merkmale bestimmt werden: Viele Merkmale aus der untere Ebene können in vollautomatischen Verfahren gewonnen werden. Dagegen wird bei der Merkmalsgewinnung der oberen Ebene bedeutend mehr menschliche Interaktion benötigt.

Als nächstes folgt eine Beschreibung der Inhalte, die gebraucht werden, um weiter Typen von Information über Multimedia-Daten einbeziehen zu können:

- Das Format: Ein Beispiel für das Format ist das genutzte Kodierverfahren (JPEG, MPEG-2) oder das gesamte Datenvolumen. Diese Information hilft festzustellen, ob das Material beim Nutzer gelesen bzw. dargestellt werden kann.
- Zugangsbedingungen zum Inhalt: Dieses könnte Informationen über Urheberrechte bzw. Lizenzrechte und Preise integrieren.
- Einstufung: Dieses könnten elterliche Vorgaben, wie Einschaltquoten, und Inhaltsklassifikationen in vordefinierten Kategorien sein.
- Verweise zu weiteren relevanten Inhalten: Diese Information hilft dem Nutzer bedeutend schneller gesuchtes Material zu finden.
- Kontext: Im Fall von nicht fiktiven Inhalten ist es sehr wichtig Informationen über Autor, Aufnahmedatum, Aufnahmeort, wie gesellschaftliche Ereignisse und Herkunft, zu wissen (z.B.: Olympische Spiele 1996, Finale des 200 Meter, Hürdenlauf der Männer)

In vielen Fällen ist es wünschenswert textuelle Informationen für die Beschreibung zu nutzen. Vorsicht ist angebracht immer dann, wenn die Gebrauch der Beschreibungen so weit wie möglich unabhängig von der sprachlichen Region sein soll. Ein gutes Beispiel bei einer Kinoauskunft wäre: Ein Text auf einem mobilen Telefon stellt die Namen des Autors, des Films und der Aufführungsorte dar.

MPEG-7 Daten können physikalisch fest mit seinem assoziativen audio-visuellen Material in dem gleichen Datenstrom oder auf dem gleichen Speichermedium verbunden sein. Aber die Beschreibungs-Elemente können genauso gut sich auch an einem anderen Ort auf der Erde befinden. Wenn der Inhalt und seine Beschreibungs-Elemente nicht gemeinsam vom Nutzer ausfindig gemacht werden kann, existieren Mechanismen, die das AV-Material und die MPEG-7 Beschreibungsdaten automatisch verbinden. Diese Verweise arbeiten in beiden Richtungen.

2.5.1 MPEG-7 Beschreibungsschnittstelle für Inhalte

Ein Gesamtsystem zur Informationssuche besteht prinzipiell aus drei Teilen: Analysewerkzeug(e), Datenbank (enthält die Beschreibung) und Suchwerkzeug(e). Analyse- und Suchwerkzeuge werden von MPEG-7 nicht spezifiziert, sondern lediglich das Format der in der Datenbank enthaltenen Informationen.

MPEG -7 zielt auf Anwendungen, die Speicher-orientiert (Online oder Off-line) oder in der Streaming-Technologie (u.a. Rundfunk-Verteilung, Push-Technologie im Internet) arbeiten. Es kann für die beiden Anwendungsarten Echtzeit und Nicht-Echtzeit eingesetzt werden. Unter Echtzeit-Umgebung ist gemeint, daß Informationen mit dem Inhalt in Verbindung gebracht wird, während gleichzeitig die Informationen bzw. Daten beim Anwender erfaßt werden.

Abbildung 13 zeigt ein grob vereinfachtes Blockdiagramm einer möglichen MPEG-7 Prozeßkette, um den Umfang des MPEG-7 Standards zu erklären. Die Prozeßkette beinhaltet die Merkmalsextraktion (Analyse), die Beschreibung selbst und die Suchmaschine (Applikation). In der Tat werden bei der MPEG-7 Beschreibung die automatische Extraktion von Merkmalen (Beschreinungs-Elemente) äußerst wichtig werden. Beispiele hierfür wären Bildanalyse und Spracherkennung. Es ist ebenfalls klar, daß die automatische Extraktion nicht immer möglich sein wird. Es ist noch anzumerken, daß je höher die Abstraktionsebene gewählt wird, um so schwieriger automatische Extraktionen durchgeführt werden können und um so notwendiger werden die interaktiven Extraktionswerkzeuge eingesetzt. Obwohl die automatischen Extraktionswerkzeuge sehr wichtig sind, werden weder automatische noch semi-automatische Merkmals-Extraktionsalgorithmen Bestandteil des Standards sein. Auch die Suchmaschine wird nicht spezifiziert. Der Hauptgedanke dieses nicht zu standardisieren liegt darin begründet, das die Interoperabilität zugelassen werden soll, die nur im industriellen Wettkampf befriedigende Ergebnisse bringt. Auch möchte man die zukünftigen innovativen Weiterentwicklungen und technologischen Steigerungen in diesem Gebiet mit berücksichtigen.

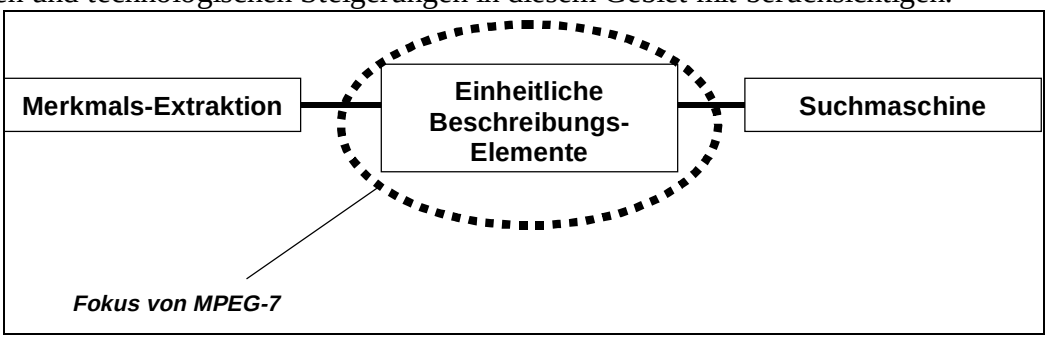

Abbildung 13: Die Prozeßkette und der Fokus von MPEG-7

Um einen besseren Eindruck der vorgestellten Technologie zu erhalten, sind die Beziehungen von den Beschreibungselementen, den Beschreibungsschemen und der Beschreibungs-Definitionssprache in Abbildung 14 dargestellt. Das gepunktete Rechteck in der Abbildung umfaßt die normierten Komponenten des MPEG-7-Standards. Die Pfeile von der Bescheibungs-Definitionssprache zu den Beschreibungsschemen zeigen, das die Beschreibungsschemen aus der Bescheibungs-Definitionssprache generiert werden. Darüberhinaus beschreibt die Zeichnung die Tatsache, das ein neues Beschreibungsschema aus einem bestehenden Beschreibungsschema abgeleitet werden kann.

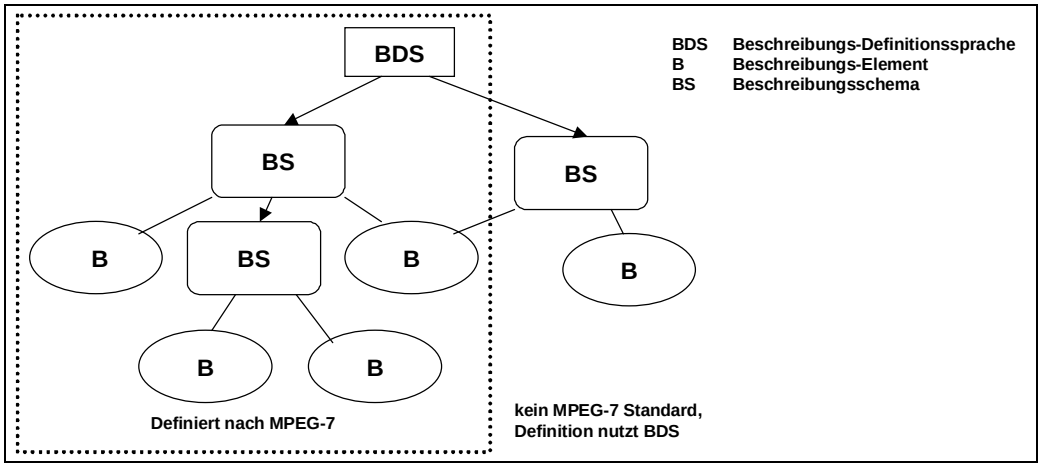

Abbildung 14: Beziehungen zwischen Beschreibungs-Elementen und Beschreibungsschemen bei MPEG-7

Die Abbildung 15 zeigt, daß die Beschreibungs-Definitionssprache den Mechanismus zum Aufbau von Beschreibungsschemen besitzt, welche die Grundlage für den Beschreibungs-Generator bildet. Die Instanzen der Beschreibungsschemen sind in Abbildung 4 zu entnehmen.

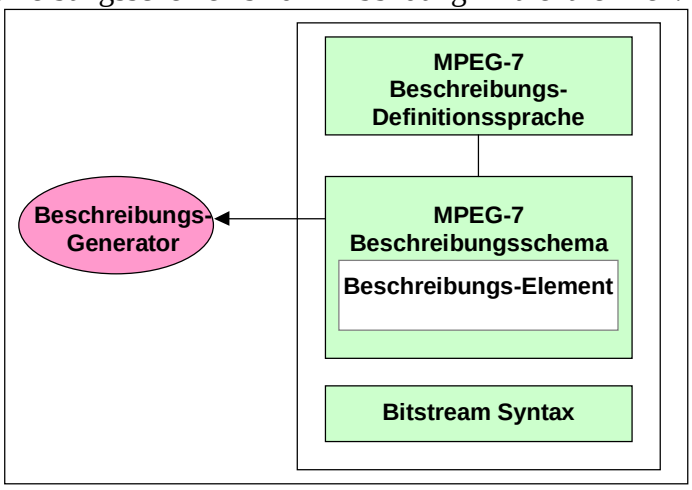

Abbildung 15: Bechreibungs-Elemente und Beschreibungsschema in MPEG-7

Abbildung 16 erklärt, wie MPEG-7 zukünftig praktisch arbeiten wird. Es zeigt die Arbeitsweise ausgehend vom multimedialen Material über den Enkodierer, Dekodierer bis zum Nutzer. Es könnten auch weitere Datenströme direkt zum Nutzer geleitet werden. Die Nutzung eines Enkodierers und Dekodierers sind optional.

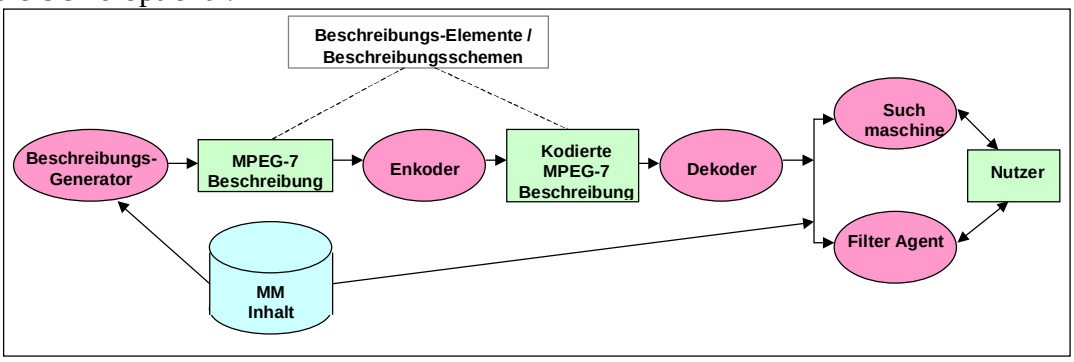

Abbildung 16: Grobes Konzept der Arbeitsweise von Anwendungen bei MPEG-7

Die Betonung bei MPEG-7 liegt auf der Bereitstellung von neuartigen Lösungen für audio-visuelle Inhaltsbeschreibungen. Deshalb sind die reinen text-orientierten Dokumente nicht das Ziel von MPEG-7. Die audio-visuellen Inhalte können jedoch im Zusammenhang mit audio-visuellen Informationen textuelle Dokumente beinhalten oder auf diese verweisen. Deswegen wird erwägt, die bestehenden Standards für "nur-Text-Dokumente" von anderen Organisationen zu unterstützen und anzupassen.

Und nun zu den Beschreibungs-Elementen selbst, Die Datenbankstruktur spielt eine kritische Rolle in der abschließenden informationsfindenden Performance. Ein schneller Such-Prozeß, ob ein Inhalte von Interesse und somit zutreffend wäre, erlaubt erst die Indexierung der Information, die damit strukturiert wird. Hier gibt es hierarchische und assoziative Verfahren.

2.5.2 MPEG-7 Informationssuche von Anwendungen

Es gibt viele Anwendungen und Anwendungsfelder, die einen Vorteil aus der einheitlichen Informationssuche des MPEG-7 Standards ziehen. Einige der wichtigen Anwendungen wären:

- Digitale Bibliotheken (Bild-Kataloge, Musik-Wörterbücher)
- Multimedia Verzeichnis-Dienste (Gelbe Seiten)
- Programmführer für Rundfunk bzw. Landesmedien (Radiokanäle, TV-Sender)
- Multimedia Bearbeitung (personalisierter elektronischer Nachrichten Dienst, Erstellung von Media-Diensten)

Die leistungsfähigen Anwendungen, die eine Informationssuche benötigen, erstrecken sich über die folgenden Anwendungsfelder:

- Bildung, (Erziehung, Berufsbildung, Virtuelle-Universität, Erwachsenenbildung)
- Journalismus (Nachrichten, Top-News, Wirtschaftsinformationen)
- Touristische Informationen (Verkehrstelematik, Reisebüro)
- Kulturelle Dienste (Historisches Museum, Bilder-Galerie)
- Unterhaltung (Spielen und Karaoke)
- Untersuchungs- und Retrieval-Dienste (Wiedererkennung von menschlichen Eigenschaften, Gerichtsmedizin)
- Geographische Informationssysteme (Regionale Touristik-Informationen, universitäre Auswertungen von Ökologische Systeme)
- Entfernte Aufzeichnungs-Dienste (Landkarten, Ökologie, Kontrolle der Naturschätze)
- Überwachung (Verkehr, Sicherheit, oberirdische Beförderung, militärische Operationen)
- Bio-Medizinische Anwendungen (Altenversorgung, Teletherapie)
- Einkauf (Kleidung)
- Haus- und Garten-Dienste (Architektur, Immobilien, Wohnungsdesign)
- Soziale Dienste (Single-Club, Treffpunkte, Vereine)
- Medien-Archive (Film, Video, Radio)

Wie bzw. mit welchem Verfahren die MPEG-7 Daten genutzt werden, um Antworten auf Nutzeranfragen zu finden, liegt nicht im Fokus des Standards. Im Prinzip kann gesagt werden, daß jede Art von audio-visuellen Materialien gefunden werden könnte, wenn jede Art von Datenanfrage unterstützt wird. Das heißt, daß zum Beispiel ein Video-Material erfragt werden könnte, wenn Video, Audio oder Sprache einzeln genannt werden. Es ist eine Angelegenheit der Suchmaschine, die passende Daten der Suchanfrage und der MPEG-7 Beschreibungs-Elemente zusammenzubringen. Einige Beispiele für Suchanfragen wären:

- Musik: Spiel einige wenige Noten auf einer Klaviertastatur und erhalte eine Liste von musikalischen Kompositionen zurück oder bekomme die vollständige Melodie oder hole ein irgendwie passendes Bild (z. B. nach dem Ausdruck der Emotionen)
- Grafik: Zeichne einige wenige Linien auf dem Bildschirm und erhalte einen Satz von Bildern zurück, die ähnliche Grafiken, Logos oder Begriffszeichen besitzen.
- Bilder: Definiere Objekte, wie farbige Flächen und Texturen, und erhalte einige Beispiele zurück. Diese durchsuche nach den interessantesten Objekten und ergänze damit dein eigenes Bild.
- Film und Video: Gib einen Satz von Objekten an, die die Bewegungen und Beziehungen der Objekte beschreiben. Dann erhalte eine Liste der zutreffenden Animationen zurück, die temporär und im direkt Zugriff stehen.
- Sprache: Nehme eine Sprachprobe von Pavarotti's Stimme und erhalte eine Liste von Pavarotti's Audioaufnahmen und Video-Clips zurück, in denen Pavarotti singt oder auftritt.
- Szenarien: Beschreiben die Handlungen einer Szene, z.B.: ein lachender Clown, und erhalte eine Liste von ähnlichen Szenen mit ähnlichen Handlungen zurück, wie ein weinender Clown und ein lachender Zauberer.

3 Zusammenfassung

Innovative multimediale Dienste werden durch die Globalisierung und Konvergenz der Märkte, als auch durch Provider-Strategien ausgerichtet. Grundlegende Innovationsfelder sind: Globaler Zugang, Navigation und Intelligenter Inhalt. Die MPEG-Standards - im besonderen MPEG-4 und MPEG-7 - helfen, die oben genannten Forderungen zu erfüllen.

Weiterhin ermöglichen sie auch für die Provider und den Kunden eine Zukunftssicherheit zu geben und einen zeitlichen Bestand für innovative Produkte zu sichern. Die Aufwärtkompabilität der MPEG-Standards ermöglicht die Vermeidung von Überschneidung und die Erschließung neuer Dimensionen.

4 Verwendete Abkürzungen

MPEG	Motion Picture Experts Group
JPEG	Joint Photographic Experts Group
ADPCM	Apaptive Delta Pulce Code Modulation
VRML	Virtual Reality Modeling Language
BIFS	Binary Format for Scene description
AVO	Audio Visual Object
VOP	Video Object Plan
DDL	Description Definition Language
DS	Description Schemes

5 Schrifttum

[1] International Organisation For Standardisation – Organisation Internationale de Normalisation, ISO/IEC JTC1/SC29/WG11, ISO/IEC JTC1/SC29/WG11 N2323, MPEG-4 Overview – (Dublin Version), Final-Status, Dublin, März 1999

[2] International Organisation For Standardisation – Organisation Internationale de Normalisation ISO/IEC JTC1/SC29/WG11, MPEG Draft: MPEG-7 Context and Objectives (Version 10) International Organisation For Standardisation, ISO/IEC JTC1/SC29/WG11N2329, Approved-Status, Dublin, Oktober 1998