

Die Standards MPEG-4 und MPEG-7 in den Multimedia-Diensten



Der Autor
Dr. Jürgen Lohr ist seit 1993 bei der Deutschen Telekom und deren Tochter-Unternehmen tätig. Seit 1998 ist er als Projektmanager für Informations- und Mediendienste bei der T-Nova Deutschen Telekom Innovationsgesellschaft, Berkom, beschäftigt.

Die Multimedia-Dienste erhalten durch die Datenreduktion bei der Kompressionstechnologie eine Wirtschaftlichkeit, die den breiteren Einsatz von breitbandigen Diensten erlaubt. Die Dienste benötigen für die verschiedenen Medien nicht mehr so große Übertragungs- und Speicherleistungen. Bei den entwickelten Verfahren, den so genannten MPEG-(Motion Picture Experts Group-)Standards, werden die Video- und Tonsignale in die digitale Ebene überführt und anschließend unrelevante Signale entfernt. Der daraus resultierende Datenstrom benötigt weniger Bandbreite bei der Übertragung zum Endkunden.

Die MPEG-Organisation wurde bereits im Jahre 1988 ins Leben gerufen und ist ein gemeinsames Gremium der beiden Organisationen ISO (International Standard Organization) und IEC (International Electrotechnical Commission), welches sich mit der Standardisierung von Kodier- und Kompressionsverfahren für die digitalen Bild-, Video- und Audioformate befasst.

Mittlerweile sind vier wichtige Standards mit MPEG-1, MPEG-2 und MPEG-4 verabschiedet worden sowie mit MPEG-7 in Vorbereitung. Da die Grundlagen zu MPEG-1, -2 und -Audio bereits in anderen Beiträgen¹ behandelt wurden, werden hier ausschließlich die neuen bzw. aktuellen MPEG-Standards vorgestellt.

1 Ausgangslage

Die Globalisierung sowie die Konvergenz² in den so genannten TIMES³-Märkten (Telecommunications, Informationstechnology, Multimedia, Entertainment and Security) führen zu starken Strukturänderungen bei den Providern von Multimedia-Diensten. Den aktuel-

len Entwicklungen am Markt kann nur mit grundlegenden Strukturän-

¹ Siehe hierzu den Beitrag „Business TV – Fernsehen für Unternehmen“, Unterrichtsblätter, Heft 7/1997, S. 392–406.

² Siehe hierzu den Beitrag „Konvergenz – das Zusammenwachsen von Festnetz, Mobilfunk und Internet“, Unterrichtsblätter, Heft 4/2000, S. 170–175.

³ Abkürzungen werden in einem Verzeichnis auf Seite 328 erläutert.

Das Thema im Überblick

Für die immer umfangreicher werdenden digitalen Multimedia-Dienste sind Kodier- und Kompressionsverfahren nötig, weil die dabei entstehenden Datenmengen nicht mehr in einem annehmbaren Verhältnis zur verfügbaren Übertragungsrate und der Speichermenge stehen. Inzwischen sind bei der Festlegung derartiger Standards neue Wege beschritten worden, die einen globalen Zugang, die Navigation und die intelligenten Inhalte im Fokus haben.

Das Ziel von MPEG-4 ist, ein Konzept zu objektorientierten Multimedia-Diensten zu beschreiben und die Audio- und Videodaten derart zu komprimieren, dass Multimedia-Dienste über schmalbandige Leitungen übertragbar werden. Damit liefert MPEG-4 die Basistechnik u. a. für die Bildtelefonie und für Videokonferenzsysteme, so dass gute Bildqualität erreicht werden kann. MPEG-7 beschäftigt sich damit, den Inhalt von multimedialen Diensten in einer effizienten und zweckmäßigen Beschreibungssprache zu charakterisieren. Aufgrund dieser Beschreibungssprache ist eine komfortable Suche nach einem bestimmten audiovisuellen Inhalt erst möglich.

derungen begegnet werden, damit weiterhin innovative Produkte und Dienstleistungen angeboten werden können. Die wichtigsten Faktoren hierfür liegen in:

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

Im Hinblick auf diese Provider-Strategien und das Entstehen eines integrierten TIMES-Sektors der Wirtschaft ist die Entwicklung neuer Formen der Informationsbeschaffung und -präsentation von außerordentlicher Wichtigkeit (Bild 1).

1.1 Globaler Zugang

Mehr und mehr kommt es darauf an, wichtige Informationen weltweit und unabhängig vom verwendeten Endgerät zugänglich zu machen (Bild 2). Dies ist bislang in der Regel nur innerhalb einer geschlossenen Nutzergruppe, wie z. B. in Unternehmen, möglich. In Zukunft muss der Zugang zu Informationen über folgende Endgeräte, basierend auf unterschiedlichen Standards („Web-standards“, H. 320,...), erfolgen:

- Fernsehgerät,
- PC,
- tragbare mobile Geräte (Laptop, Personal Data Assistant⁴ = PDA),
- in Fahrzeuge integrierte Geräte (z. B. Boardcomputer, Leitsystem),
- Sprachtelefon, Bildtelefon, „Screenphone“, Fax,
- Mobilfunk-Display (Handy), Pager.

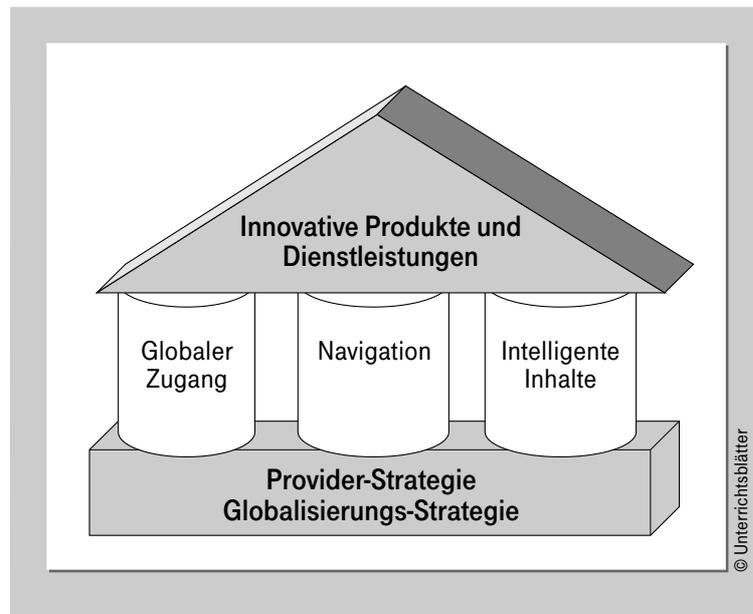


Bild 1: Faktoren für innovative Produkte und Dienstleistungen

Darüber hinaus sollte eine Kommunikation über beliebige Übertragungswege möglich sein:

- schmal- und breitbandige terrestrische Netze (z. B. PSTN, ISDN, ADSL, ATM, BVN),
- schmal- und breitbandige mobile Netze (z. B. GSM, UMTS),
- Satelliten- und Richtfunknetze (z. B. DVB, MMDS, LMDS).

Die Aufbereitung der Inhalte stellt einen weiteren Aspekt für den globalen Zugang dar. Je nach Displaygröße der Endgeräte und dem zu Grunde liegenden technischen Standard müssen gleiche Inhalte in unterschiedlicher Form aufbereitet werden. Dies sollte automatisch erfolgen und ist nur möglich, wenn

für die Gestaltung von Inhalten Grundregeln geschaffen werden, die deren Einsatz als eine Art „modularer Inhalt“ sicherstellen.

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 ist ein Schlüsselproblem. Hier müssen neue Lösungsansätze der Naviga-

⁴ Fachbegriffe werden in einem Glossar auf Seite 328 erklärt.

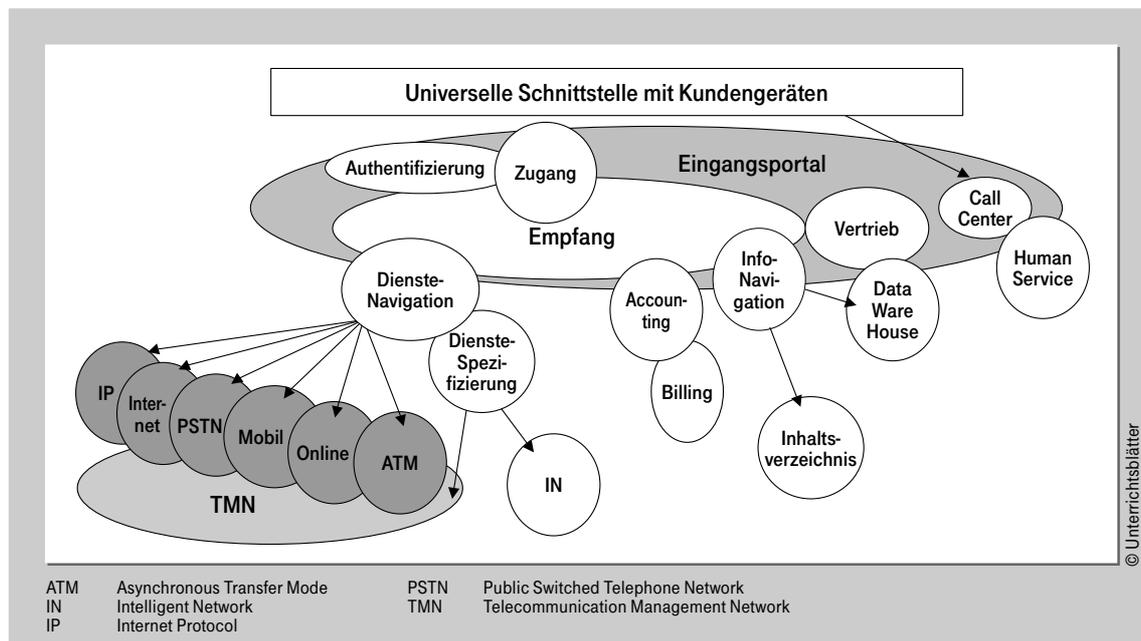


Bild 2: Darstellung eines Globalen Zugangs

ATM Asynchronous Transfer Mode
 IN Intelligent Network
 IP Internet Protocol
 PSTN Public Switched Telephone Network
 TMN Telecommunication Management Network

tion, etwa durch Einsatz von folgenden Aspekten, entwickelt werden:

- Sprach-Ein-/Ausgabe,
- dreidimensionale (3D), virtuelle, grafische Darstellung (Virtual Reality Modelling Language = VRML),
- interaktive Videos/Animationen mit Objekterkennung,
- Integration von Videoobjekten in 3D-Räumen,
- Layer⁵-Techniken.

Hierbei kommt insbesondere der Einsatz von MPEG-4 in Betracht. Im Gegensatz zu MPEG-1 und MPEG-2⁶ will MPEG-4 keine speziellen Anwendungen adressieren, sondern möglichst viele Anwendungsgebiete abdecken. Eine detaillierte Beschreibung des MPEG-4-Standards erfolgt in Abschnitt 3.

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 Agenten⁷ sowie
- intelligenter Inhalt (MPEG-7)

erforderlich. Zurzeit haben sich die ersten Internet-Dienste in Richtung persönliche Profile entwickelt. Als nächster Schritt ist der Einsatz intelligenter Agenten anzustreben, die automatisiert gezielte Suchvorgänge durchführen und aus Zustimmung bzw. Ablehnung der Sucher-

gebnisse 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.

⁵ Layer: dt. Schicht; nach der Aufgabe abgegrenzte und zusammengehörige Funktion zur Ermöglichung von Kommunikation.

⁶ Siehe hierzu den Beitrag „Datenreduktion digitalisierter Fernsehbilder“, Unterrichtsbücher, Heft 8/1999, S. 444-457.

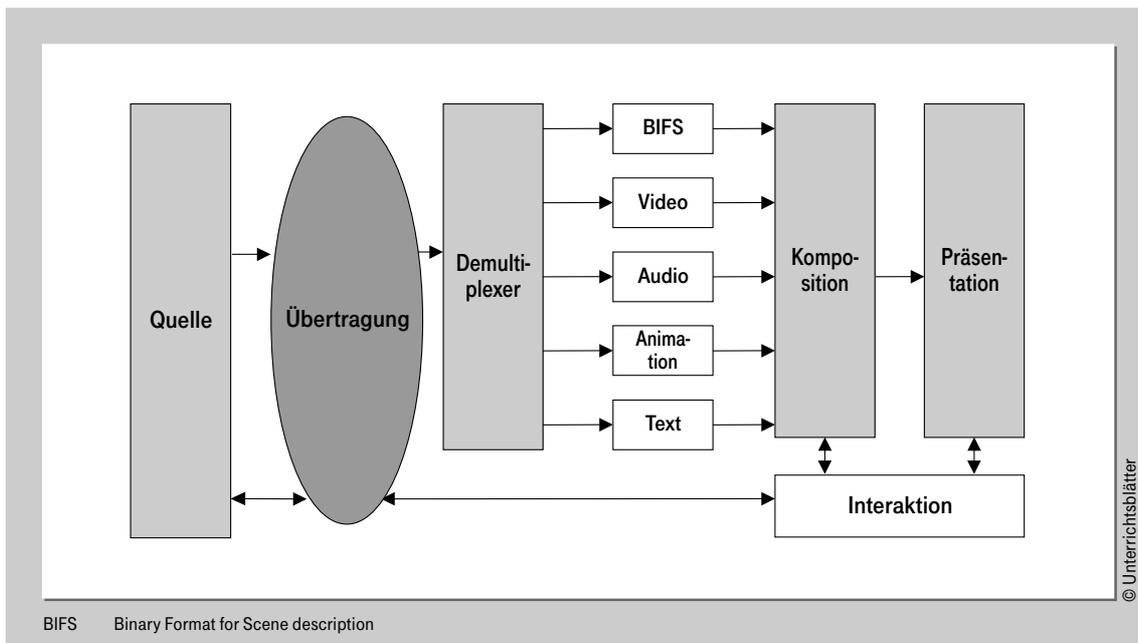
⁷ Agenten: Allgemein eine Software, die auf Anfrage bestimmte Informationen zurückliefert, die sie selbstständig gesammelt hat.

Verwendete Abkürzungen

AAC	Advanced Audio Coding
ADPCM	Adaptive Differential Pulse Code Modulation
ADSL	Asymmetric Digital Subscriber Line
AM	Amplitudenmodulation
ATM	Asynchronous Transfer Mode
BIFS	Binary Format for Scene description
BVN	Breitbandverteilstrom
CELP	Code-Excited Linear Prediction
DAI	DMIF Application Interface
DCT	Diskrete Cosinus-Transformation
DDL	Data Description Language
DMIF	Delivery Multimedia Integration Framework
DS	Description Scheme
DVB	Digital Video Broadcasting
DVD	Digital Versatile Disc
ETSI	European Telecommunications Standards Institute
FM	Frequenzmodulation
GSM	Global System for Mobile Communication
G.726	ITU-T-Empfehlung: Für die Telefonübertragung über drahtgebundene Verbindungen, Satelliten- und Funkverbindungen sowie digitale Netze.
H.261	ITU-T-Empfehlung: Standard für Grundprinzipien der digitalen Bewegtbildübertragung.
H.263	ITU-T-Empfehlung: Für die digitale Videokonferenz; spezifiziert ein Kodierverfahren für Videodaten für die Bildtelefonie.
H.320	ITU-T-Empfehlung: Für den Videokonferenzdienst über ISDN-Leitungen (aber auch für LAN); legt mehrere Normen für die Audio- und Videoübertragung fest.
H.324	ITU-T-Empfehlung: Für Multimediaübertragungen bei geringen Bitraten; beinhaltet verschiedene weitere Empfehlungen wie einen Video-Codec für V.34-Modemübertragungen über analoge Telefonleitungen sowie ATM und ISDN.

HDTV	High Definition Television
HILN	Harmonic and Individual Line plus Noise
HVXC	Harmonic Vector eXcitation Coding
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IP	Internet Protocol
ISO	International Standard Organization
ITU	International Telecommunications Union
IWU	Internetworking Units
JPEG	Joint Photographic Experts Group
LLC	Logical Link Control
LMDS	Local Multipoint Distribution System/Service
LPC	Linear Predictive Coding
MIDI	Musical Instruments Digital Interface
MMDS	Multichannel Microwave Distribution System/Service
MPEG	Motion Picture Experts Group
PCM	Pulse Code Modulation
PDA	Personal Data Assistant
PSTN	Public Switched Telephone Network
QCIF	Quarter Common Intermediate Format
SA	Structured Audio-Format
SAOL	SA Orchestra Language
SASL	SA Score Language
subQCIF	sub Quarter Common Intermediate Format
TIMES	Telecommunications, Information-technology, Multimedia, Entertainment and Security
TMN	Telecommunication Management Network
TTS	Text To Speech (Conversion)
UMTS	Universal Mobile Telecommunications System
VRML	Virtual Reality Modelling Language
V.34	ITU-T-Empfehlung (Standard) für die Datenübertragung über das Telefonnetz; beschreibt ein Modemübertragungsverfahren für die synchrone/asynchrone Datenübertragung von 2 400 bit/s bis 33 600 bit/s auf Zweidrahtleitungen im Duplex- und Halbduplexbetrieb.

Bild 3:
MPEG-4
Standard



Die zunehmende Bereitstellung digitaler, audiovisueller Inhalte macht es erforderlich, den potenziellen Anwendern ein Instrument für eine effizientere und effektivere Suche nach audiovisuellen Inhalten zu geben. Um Suchvorgänge entscheidend vereinfachen zu können, ist ein intelligenter Inhalt erforderlich. Dies gilt besonders für audiovisuelle 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 zurzeit in der Definitionsphase und wird in Abschnitt 4 erläutert.

2 MPEG-Standards

MPEG-1 und -2 wurden für die Kompression audiovisueller Gesamtszenen entwickelt und erlauben Interaktivität nur auf der Programmebene. Sie sind jedoch störungsanfällig bei fehlerhafter Übertragung. MPEG-3 war ursprünglich für die Anwendungen im Bereich HDTV (High Definition Television) gedacht. Dieser Teil der MPEG-3 wurde in MPEG-2 integriert, so dass MPEG-3 heute keine Bedeutung mehr besitzt. MPEG-5 und MPEG-6 wurden nicht definiert und sind daher nicht relevant.

MPEG-4 wurde für die Kompression audiovisueller 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, verbes-

serte Skalierung und Kodiereffizienz definiert.

Mit MPEG-7 wurde eine multimediale inhaltsbeschreibende Schnittstelle definiert, welche eine standardisierte Beschreibung bzw. Attribuierung verschiedener Arten multimedialer Informationen erreicht. Diese Attribute werden mit dem Inhalt selbst verknüpft und ermöglichen dem Anwender eine schnelle, effiziente Suche nach den gewünschten Inhalten. Im Folgenden werden nun die neuen bzw. aktuellen MPEG-Standards beschrieben:

- MPEG-4 – Audiovisuelle Szenen
- MPEG-4 – Videokompression
- MPEG-4 – Audiokompression
- MPEG-7 – Beschreibungsschnittstelle für Inhalte
- MPEG-7 – Informationssuche von Anwendungen

3 MPEG 4

Im Jahre 1997 begann das Gremium, MPEG-4 zu spezifizieren. Ende 1998 wurde der neue internationale Multimedia-Standard MPEG-4 in der Version 1 verabschiedet. MPEG-4 will keine spezielle Anwendung adressieren, sondern möglichst viele Anwendungsbereiche abdecken (Bild 3). Die charakteristischen Neuerungen waren die Integration des Standards H.263 zur Videokodierung, die inhaltsorientierte Interaktivität und ein universeller Zugriffsmechanismus. Die Funktionalitäten, die MPEG-4 zusätzlich unterstützen soll, können in die drei Teilbereiche:

- inhaltsbezogene Interaktivität,
- Kompression und
- Skalierbarkeit

unterteilt werden.

Bei der **inhaltsbezogenen Interaktivität** soll:

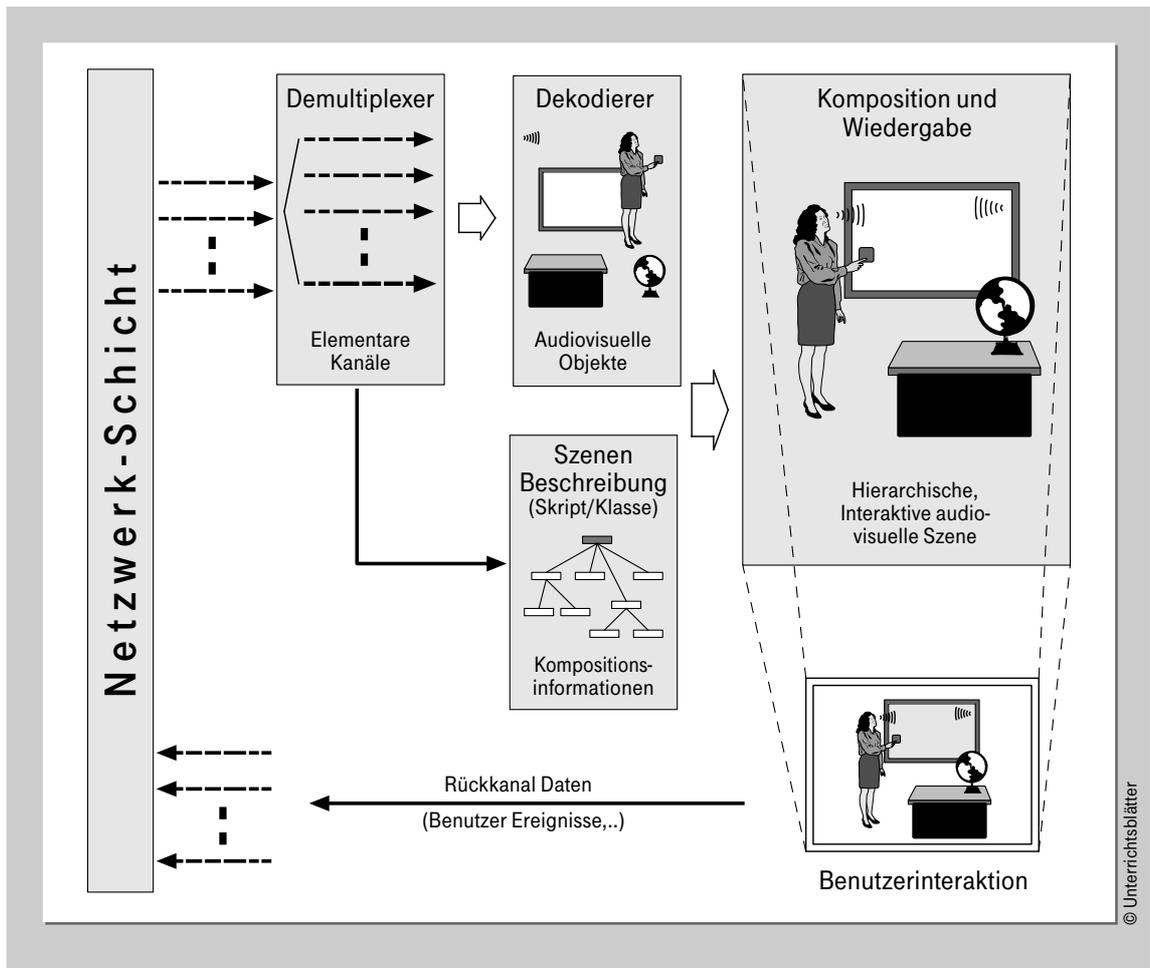
- ein schneller Zugriff auf die audiovisuellen Daten von audiovisuellen Objekten (z. B. bei Indizierung, „Hyperlinking⁸“, Laden bzw. Herunterladen und Löschen) garantiert werden,
- ein Syntax- und Kodierungsschema bereitgestellt werden, so dass spezielle Objekte einer Szene verändert werden können, ohne dass die digitale Darstellung derselben bekannt ist,
- eine Methode bereitgestellt werden, um künstliche mit natürlichen Szenen zu verknüpfen – dies wäre ein erster Schritt zur Vereinheitlichung jeglicher audiovisueller Information – sowie
- der zufällige Zugriff auf Teile einer audiovisuellen Sequenz ermöglicht werden.

Bei der **Kompression** soll

- eine bessere Qualität bei vergleichbarer Bitrate gegenüber bereits bestehenden Standards erreicht werden sowie
- mehrere Blickrichtungen oder Tonspuren einer festen Szene kodiert und später wieder synchronisiert werden können, so

⁸ **Hyperlink:** Querverweis (Hypertext-Link), der als hervorgehobene Stelle im Hauptfenster des Benutzers erscheint.

Bild 4:
MPEG-4
Terminalmodell



dass es möglich wird, auch 3D-Objekte darzustellen. MPEG-4 soll dabei die Überschneidungen der unterschiedlichen Blickrichtungen bei der Kodierung ausnutzen.

Bei der **Skalierbarkeit** ist sowohl eine räumliche als auch eine zeitliche Auflösung gemeint. Es sollen auch einzelne Objekte einer Szene skaliert werden können, das heißt, dass bestimmte Objekte stärker hervorgehoben bzw. mit stärkerer Auflösung dargestellt werden können.

3.1 MPEG 4 – Audiovisuelle Szenen

Der MPEG-4-Standard sieht die Möglichkeit zur Interaktion mit audiovisuellen Szenen-Inhalten vor und führt die Komposition von audiovisuellen Objekten in eine kohärente⁹ Präsentation nun nicht mehr beim Sender, sondern beim Empfänger durch. 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 „Vir-

tual Reality Modelling Language (VRML)“ basiert. Für den Prozess der Komposition selbst wurden bisher keine normativen¹⁰ Festlegungen getroffen.

Das Bild 4 zeigt ein Modell des MPEG-4-Terminals. Die Elemente der Systemspezifikation sind die Objektbeschreibung, die Szenenbeschreibung, die Synchronisationsebene und die Multiplexebene. Sie bilden einen Rahmen, in den die Dekodierer, die einzelne Elementarströme mit Daten für audiovisuelle 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 audiovisuelle Objekte (AVO) eingeteilt, denen ein „Video Object Plan“ (VOP) zu Grunde 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 Eigenschaften. 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 audiovisuellen Objekten. Zusätzlich erlaubt BIFS die Dar-

⁹ **kohärent:** zusammenhängend.

¹⁰ **normativ:** als Norm geltend, maßgebend, als Richtschnur dienend.

stellung von grafischen Elementen, angefangen von Linien bis hin zu komplexen Polygonzügen¹¹.

Architektonisch ist BIFS insofern flexibler als sein Vorgänger, als die Beschreibungsdaten grundsätzlich 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 so genanntes BIFS-Animationsformat bestimmt.

Die Struktur eines Szenengrafen ist in Bild 5 beispielhaft dargestellt. Es enthält ein natürliches audiovisuelles Objekt (Person), bestehend aus einer animierten synthetischen Sprecherin, einem aus Text generierten Sprachereignis (Text To Speech = TTS) und weiteren synthetischen grafischen Elementen.

Durch die Aufspaltung von audiovisuellen Präsentationen in einzelne Objekte können die verschiedenartigsten Terminals in heterogenen Netzen effizient erreicht werden. Zum einen kann für jedes Terminal ausgewählt werden, welche Objekte dargestellt werden, das heißt 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

andererseits bereits während der Übertragung der Präsentation, die 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 audiovisuellen Objekten in Raum und Zeit wiedergibt, erlauben zusätzliche Objektdeskriptoren¹² 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 Internetworking 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 Transkodierung¹³ zwischen verschiedenen Videokonferenzstandards (z. B. H.320 und H.324) eingesetzt werden, ist, dass die Inhalte der Datenströme nicht dekodiert und wieder rekodiert¹⁴ werden müssen. Eine IWU für ein rein auf MPEG-4 beruhendes au-

divisuelles Kommunikationssystem in einem heterogenen Netz kann also weniger aufwendig konzipiert werden als bei den genannten bereits bestehenden Standards.

3.1.1 Schichtenstruktur 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, so genannte 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 Multiplexdaten-

¹¹ **Polygon:** Das Vieleck; geometrisches Gebilde aus n Punkten P_1, P_2, \dots, P_n , den Ecken, und n Verbindungsstrecken (geschlossenen ebenen Polygonzügen), den Seiten des Polygons.

¹² **Deskriptor:** Kenn- oder Schlüsselwort, durch das der Inhalt einer Information charakterisiert wird und das zur Bestimmung von Daten im Speicher eines Computers dient.

¹³ **Transkodierung:** Umsetzung eines Standards in einen anderen.

¹⁴ **Rekodierung:** Die nach der Dekodierung erfolgende Umsetzung in den Kode der Zielsprache.

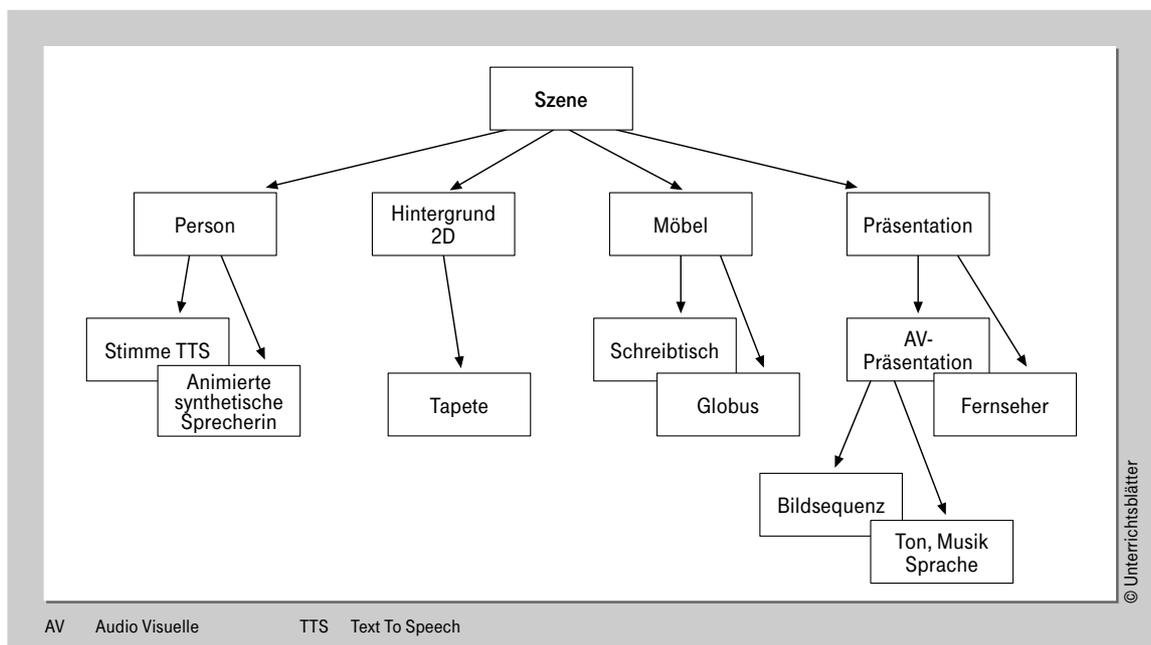
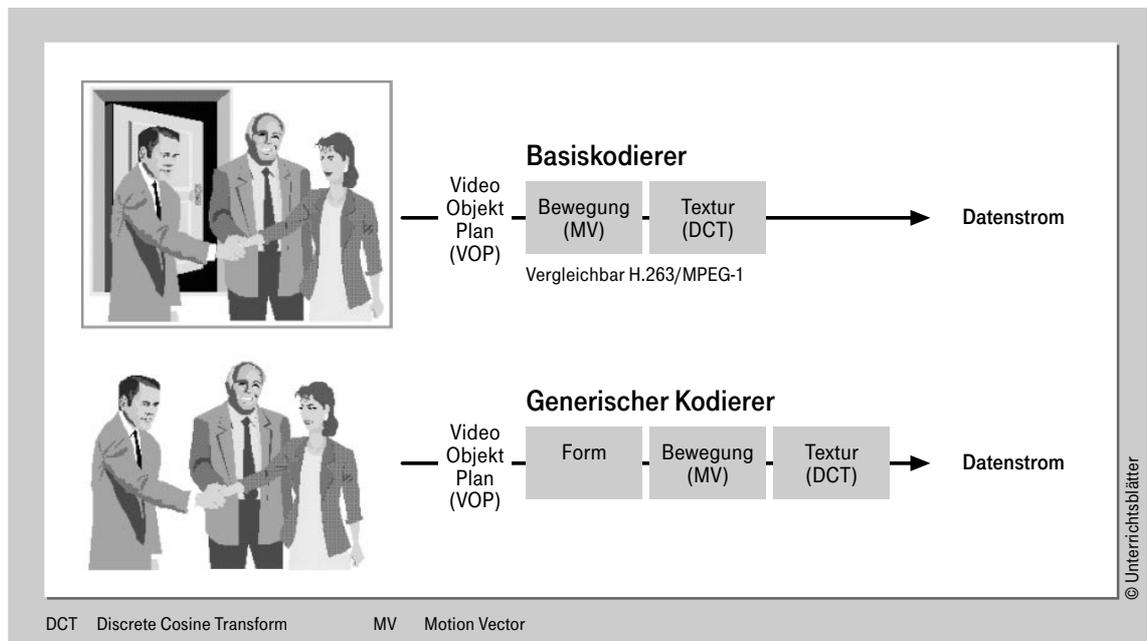


Bild 5: Szenengraf in einer BIFS-Szenenbeschreibung

Bild 6:
MPEG-4 Stufen-
konzept des
konventionellen
Hybridkodierers



ströme 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 FlexMux¹⁵, 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). Das Delivery Multimedia Integration Framework (DMIF) ist ein weiterer Teil der MPEG-4-Standardserie, der sich mit der konkreten Spezifikation dieser Schicht befasst. Für jedes mögliche Transportnetzwerk, sei es ein IP Netzwerk, ein DVB-TV-Netz oder ein ATM-Netz, ist eine Adaption zu spezifizieren, um tatsächlich Kommunikation in heterogenen Netzwerken zu ermöglichen.

3.2 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 Beschrei-

bung eines Kopf- und Gesichtsmodells nebst Animationsparametern sowie die Definition von zweidimensionalen Gittern zur Animation von Texturen¹⁶.

3.2.1 Funktionalitäten des Videoteils

Der MPEG-4-Standard ist ähnlich wie bereits die Vorgänger MPEG-1/2 generisch ausgelegt, das heißt 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 progressiv¹⁷ als auch im klassischen Zeilensprungverfahren (interlaced¹⁸) erfolgen. Um sich an unterschiedliche Kanalkapazitäten anpassen zu können, bietet MPEG-4 die Möglichkeit der Skalierbarkeit, das heißt, Videodaten können hierarchisch in einer Weise kodiert werden, die es dem Dekodierer 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 des „Video Object Plan (VOP)“ einge-

führt. 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 der VOP beliebig 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 beispielsweise 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.

3.2.2 Struktur des MPEG-4-Videokodierers

Der MPEG-4-Video-Standard setzt sich zusammen aus einem Basiskodierer und einem darauf aufsetzenden generischen Videokodierer. Bild 6 zeigt das grundlegende Konzept, mit dem – ausgehend von

¹⁵ **FlexMux:** Abk. Flexibler Multiplexer; bei der Datakom Austria eingesetzte Datenübertragungseinrichtung, die auf einer 2-Mbit/s-Anschlussleitung mehrere Simultankanäle für beliebige Dienste im synchronen Zeitmultiplexverfahren bereitstellen kann.

¹⁶ **Textur:** Gewebe, Faserung.

¹⁷ **progressiv:** stufenweise fortschreitend, sich entwickelnd.

¹⁸ **Interlaced:** dt. zusammengefügt, miteinander verbunden.

bekannten Kodieretechniken – eine objektbasierte Videokodierung ermöglicht wird.

Der Basiskodierer entspricht im Wesentlichen der konventionellen Hybridkodierung, bei der die beiden Komponenten Bewegungsinformation und Textur bzw. Prädiktionsfehler¹⁹ 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.

3.2.2.1 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 approximative²⁰ Formbeschreibung durch eine Unterabtastung²¹ der Binärmaske vor der eigentlichen Kodierung realisiert wird. Zur Kodierung der Maske wird ein kontextadaptives²² arithmetisches Verfahren verwendet, bei dem auch zwischen Intraframe-²³ und zeitlich prädiktiver²⁴ 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.

3.2.2.2 Bewegungsschätzung und Kompensation

MPEG-4 teilt das Videoobjekt in ein regelmäßiges Blockraster und verwendet dann blockbasierte Bewegungsschätzung und -kompensation, um zeitliche Korrelationen zwischen aufeinanderfolgenden VOP auszunutzen. Die geschätzten Bewegungsvektoren und der verbleibende Prädiktionsfehler werden ähnlich wie in MPEG-1/-2 oder H.261/263 kodiert, wobei zwischen 8×8-Blöcken und 16×16-Makroblöcken umgeschaltet werden kann, und auch eine Bewegungskompensation mit überlappenden Nachbarblöcken (overlapped block motion compensation) möglich ist. Auf Grund 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 so genanntes Polygon-Matching²⁵ 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 Intraframe-Kodierung und vorwärts- oder bidirektional-prädizierten²⁶ VOP umgeschaltet werden.

3.2.2.3 Texturkodierung

Die Intraframe-kodierten VOP – wie auch der Prädiktionsfehler nach der Bewegungskompensation – werden unter Verwendung der DCT (Diskreten Cosinus Transformation) für 8×8-Blöcke kodiert. In einem Makroblock werden dabei vier benachbarte 8×8-Luminanzblöcke²⁷ und zwei Chrominanzblöcke²⁸ zusammengefasst. Ähnlich zur Bewegungskompensation ist auch hier eine besondere Anpassung für die Randblöcke eines Videoobjekts 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.

3.2.2.4 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, das heißt, 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.

3.3 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

¹⁹ **Prädiktion:** Vorhersage, Voraussage.

²⁰ **approximativ:** angenähert.

²¹ **Unterabtastung:** Abtastung unterhalb der minimalen Rate.

²² **kontextadaptiv:** auf der Anpassung an die Umgebung beruhend.

²³ **Intraframe-Kodierung:** Kodierung innerhalb des Rahmens.

²⁴ **prädiktiv:** die Möglichkeit einer Prädiktion enthaltend, vorhersagbar.

²⁵ **Polygon-Matching:** Durchsuchen nach charakteristischen Polygonen.

²⁶ **prädiziert:** ein Prädikat beigelegt, einen Begriff durch ein Prädikat bestimmt.

²⁷ **Luminanzsignal:** Signalanteil (z. B. beim Farbfernsehen) zur Übertragung der Helligkeitswerte.

²⁸ **Chrominanzsignal:** Signalanteil (z. B. beim Farbfernsehen) zur Übertragung der Farbwerte.

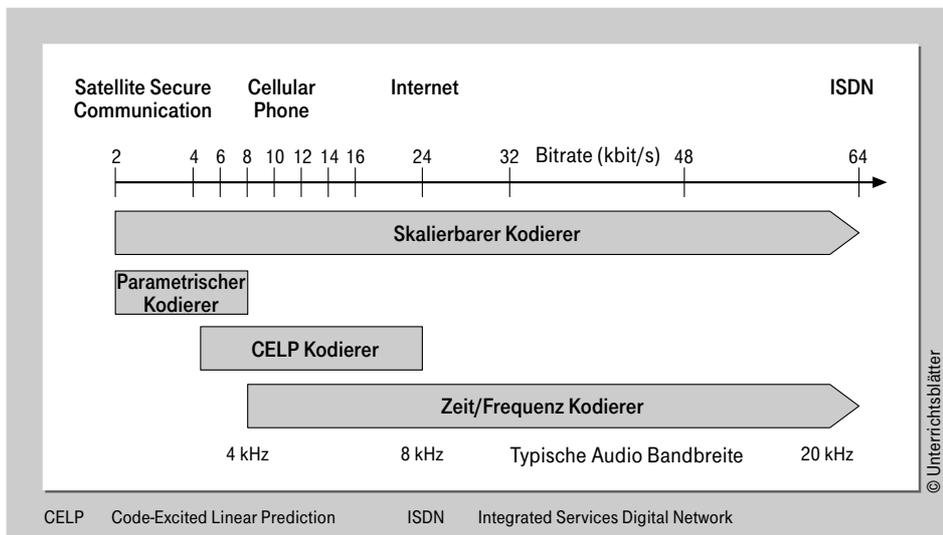


Bild 7: Übersicht über MPEG-4-Audio-Kodierungsverfahren

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.

3.3.1 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-Standards möglich war. Parallel zu diesen Aktivitäten wurde noch am MPEG-2-Standard, Teil 7 (Advanced Audio Coding = AAC), gearbeitet. Dieser kann bis zu 48 Kanäle mit 64 kbit/s/ch²⁹ kodieren, ohne dass hörbare Unterschiede zum Originalsignal (etwa 700 kbit/s/ch) wahrnehmbar sind. Für eine vergleichbare Qualität gegenüber MPEG-1 und MPEG-2/Layer-2 braucht AAC damit etwa eine halb so große Datenrate.

Als hochwertiges Kodierungsverfahren wurde AAC dann 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-Kodie-

rung), zu denen auch AAC gehört, mit Datenraten von etwa 24 kbit/s/ch 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 kbit/s bis 24 kbit/s arbeiten, oder sprach-optimierte parametrische Verfahren für sehr niedrige Datenraten von 2 kbit/s 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. Hierbei erlaubt MPEG-4 eine feine Skalierung der Datenrate (Fine Rate Control) und der 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 musik-optimierte, parametrische Verfahren (ab 4 kbit/s).
- Beim parametrischen HILN-Kodierer (Harmonic and Individual Line plus Noise) werden harmonische und dominierende Frequenzanteile sowie Rausch-

eigenschaften extrahiert und übertragen.

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

In Bild 7 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 Basisbitstrom 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.

3.3.2 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, prosodische³⁰ 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 grafisch animierten Gesichtern benutzt werden. Die Umsetzung des Textes in die entsprechende Sprache bleibt dem Implementierer³¹ überlassen.

²⁹ kbit/s/ch: kbit/s je Kanal.

³⁰ prosodisch: die Prosodie (Lehre von der metrisch-rhythmischen Behandlung der Sprache) betreffend, die Silben messend.

³¹ implementieren: Software, Hardware o. Ä. in ein bestehendes Computersystem einsetzen, einbauen und so ein funktionsfähiges Programm erstellen.

Als synthetisches Musikformat wurde unter anderem MIDI (Musical Instruments Digital Interface) mit der DLS-Erweiterung (Downloadable Sounds) gewählt. Zusätzlich wird das SA-Format (Structured Audio) eingesetzt. Es besteht aus einer Instrumenten-Beschreibungssprache 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 etwa 100 vorinstallierten Audio-Funktionen. Der besondere Vorteil ist, dass hiermit ein Musiker den Klang seiner Instrumente ganz genau bestimmen kann, was mit MIDI allein nicht möglich ist.

3.3.3 Komposition der Audioobjekte zu einem Klangbild

Audio-Szenekomposition 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 beispielsweise von möblierten Räumen oder Konzerthallen nachzubilden.

3.4 MPEG-4-Profil

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 nach-

geordneten Ebenen (Levels) feste Rahmen für die freien Kodierparameter wie Datenrate, Bildgröße, Audio-Abtastfrequenz, Anzahl der Szenenobjekte festlegt. Da die Vielzahl 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 hauptsächlich für zukünftige mobile Netzwerke, wie UMTS, geeignet.
- 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 Dekodierkomplexität mehr als eine Qualitätsstufe benötigen. Dies ist beispielsweise bei einem softwarebasierten Videoabruf der Fall.
- Das „Core Visual Profile“ ergänzt zusätzlich noch die Funktionalität der Formkodierung und eignet sich damit im Besonderen für interaktive Anwendungen wie Multimediakommunikation im Internet.
- Das „Main Visual Profile“ erlaubt ergänzend die Kodierung von semitransparenten Videoobjekten im Zeilensprungverfahren und zielt damit insbesondere auf hochqualitative Fernsehdienste und DVD-Anwendungen ab.

4 MPEG-7

Der Standard MPEG-7 wurde formal mit „Multimediale Inhaltsbeschreibende Schnittstelle“ (Multimedia Content Description Interface) definiert. Im Oktober 1998 erfolgte der „Call for Proposals“, im Dezember 1999 wurde der erste Entwurf erstellt. Der internationale Standard ist für September 2001 geplant.

Der Standard befasst sich damit, den Inhalt von audiovisuellen Sequenzen, Bildern und Grafiken in einer effizienten und zweckmäßigen Darstellung zu repräsentieren, so dass damit eine Informationssuche möglich wird. Das Ziel von

MPEG-7 ist es, die zurzeit bestehenden eingeschränkten Möglichkeiten proprietärer³² Lösungen für die Identifizierung von Inhalt zu erweitern, insbesondere durch die Einbeziehung einer größeren Menge von Datentypen. Mit MPEG-7 soll eine standardisierte Beschreibung bzw. Attribuierung verschiedener Arten multimedialer Informationen erreicht werden. Diese Attribute werden mit dem Inhalt selbst verknüpft und ermöglichen dem Anwender eine schnelle, effiziente Suche nach den gewünschten Inhalten.

MPEG-7 spezifiziert einen Standardsatz von Beschreibungs-Elementen, die zur Beschreibung von unterschiedlichen Typen von Multimedia-Informationen benutzt werden. MPEG-7 legt neben den Beschreibungs-Elementen auch eine Struktur (Beschreibungs-Schemen = Description Schemes/DS) und deren Beziehungen untereinander fest. Diese Beschreibungen, die eine Kombination aus den Beschreibungs-Elementen und den Beschreibungs-Schemen sind, können mit dem eigentlichen Inhalt direkt verbunden werden. Dieses erlaubt eine schnelle und effektive Suche der Inhalte nach den Vorstellungen des Anwenders. In MPEG-7 wird außerdem eine Sprache zur detaillierten Definition der Beschreibungs-Schemen standardisiert (Beschreibungs-Definitionssprache = Description Definition Language/DDL).

Das Material, welches mit MPEG-7-Daten ergänzt wurde, kann indexiert und gesucht werden. Es kann 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 (Information der szenischen Komposition = scenarios composition information).

Der MPEG-7-Standard baut auf einzelnen Definitionen der analogen Signalverarbeitung, der Pulse Code Modulation (PCM), der Standards MPEG-1, -2 und -4 auf. Eine der Funktionen 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

³² proprietär: eigentümlich.

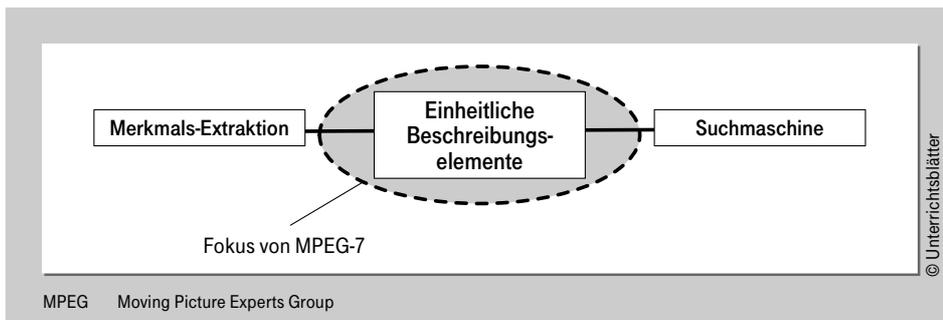


Bild 8:
Die Prozesskette
und der Fokus
von MPEG-7

die 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. Ein MPEG-7-Beschreibungs-Element kann an ein analoges Video oder an ein Bild, das zum Druck auf Papier vorgesehen ist, angehängt werden. Gerade deswegen sind MPEG-7-Beschreibungs-Elemente nicht abhängig von der kodierten Repräsentation der Inhalte, weil MPEG-7 im Zusammenhang mit MPEG-4 steht.

Der MPEG-4-Standard ist mit Werkzeugen ausgestattet, die die audiovisuellen Inhalte als Objekte mit bestimmter Zuordnung zu Zeit (Synchronisation) und Raum (auf dem Bildschirm für Video, in dem akustischen Raum bei Audio) enkodieren. Wird der MPEG-4-Enkodierer genutzt, ist es möglich, Beschreibungs-Elemente zu Objekten innerhalb einer Szene hinzuzufügen. Das können Audio- oder visuelle Objekte sein. MPEG-7 erlaubt unterschiedliche Detaillierungsstufen der Beschreibungs-Elemente, um die Möglichkeit von verschiedenen Ebenen von Unterscheidungen anbieten zu können.

Die beschriebenen Merkmale sind unterschiedlich für verschiedene kulturelle Regionen und bei verschiedenen Anwendungen. Dies besagt, 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 gegeben:

Auf der **unteren Ebene** (Abstraktionsebene) werden Beschreibungs-Elemente wie Form, Größe, Textur, Farbe, Bewegungsrichtung und Position benötigt. 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 wie: „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 unteren Ebene können in vollautomatischen Verfahren gewonnen werden. Dagegen wird bei der Merkmalsgewinnung der oberen Ebene bedeutend mehr menschliche Interaktion benötigt.

Die Inhalte, die benötigt werden, um weitere Typen von Information über Multimedia-Daten einbeziehen zu können, sind:

- Das Format: Beispielsweise 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 können Informationen über Urheberrechte bzw. Lizenzrechte und Preise sein.
- Einstufung: Dieses können 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: Bei nicht fiktiven Inhalten ist es sehr wichtig, Informationen über Autor, Aufnahmezeitpunkt, Aufnahmeort sowie gesellschaftliche Ereignisse und Herkunft zu bekommen (z. B.: Olympische Spiele 1996, Finale des 200-Meter-Hürdenlaufs der Männer).

In vielen Fällen ist es wünschenswert, textuelle³³ Informationen für die Beschreibung zu nutzen. Vorsicht ist immer dann angebracht, wenn der Gebrauch der Beschreibungen weitestgehend unabhängig von der sprachlichen Region sein soll. (Ein Beispiel einer Kinokunft wäre: Ein Text auf einem mobilen Telefon stellt die Namen des Autors, des Films und der Aufnahmestellen dar.)

Die MPEG-7-Daten können physikalisch fest mit seinem audio-visuellen Material in demselben Datenstrom oder auf demselben Speichermedium verbunden sein. Aber genauso gut können die Beschreibungs-Elemente sich auch auf einem anderen Speichermedium befinden. Wenn der Inhalt und seine Beschreibungs-Elemente nicht gemeinsam durch den Nutzer auffindbar gemacht werden können, existieren Mechanismen, die das audiovisuelle Material und die MPEG-7-Beschreibungsdaten automatisch verbinden. Diese Verweise arbeiten in beiden Richtungen.

4.1 MPEG-7 Beschreibungsschnittstelle für Inhalte

Ein Gesamtsystem zur Informationssuche besteht grundsätzlich 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 ab, die speicherorientiert (Online oder Offline) oder in der Streaming³⁴-Technologie (u. a. Rundfunk-Verteilung, Push³⁵-Technologie im Internet) arbeiten. Es kann

³³ **textuell:** den Text betreffend.

³⁴ **Streaming:** In der Datenkommunikation Verfahrensprinzip, bei dem die Daten als stetiger, kontinuierlicher Strom verarbeitet werden.

³⁵ **Push:** Allgemein ein Vorgang, bei dem Daten von einem anderen Programm oder Rechner spontan gesendet werden, ohne abgerufen worden zu sein.

für die beiden Anwendungsarten Echtzeit und Nicht-Echtzeit eingesetzt werden. Unter Echtzeit-Umgebung ist gemeint, dass Informationen mit dem Inhalt in Verbindung gebracht werden, während gleichzeitig die Informationen bzw. Daten beim Anwender erfasst werden.

Das Bild 8 zeigt ein grob vereinfachtes Blockdiagramm einer 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). Die automatische Extraktion von Merkmalen (Beschreibungs-Elementen) bei der MPEG-7-Beschreibung ist äußerst wichtig. Beispiele hierfür sind Bildanalyse und Spracherkennung, jedoch ist die automatische Extraktion nicht immer möglich. Je höher die Abstraktionsebene gewählt wird, um so schwieriger automatische Extraktionen durchgeführt werden können, desto 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, diese nicht zu standardisieren, liegt darin begründet, dass die Interoperabilität³⁶ zugelassen werden soll. Man möchte die zukünftigen innovativen Weiterentwicklungen und technologischen Steigerungen in diesem Gebiet mit berücksichtigen.

Um einen besseren Eindruck der vorgestellten Technologie zu erhalten, sind die Beziehungen von den Beschreibungs-Elementen, den Beschreibungs-Schemen und der Beschreibungs-Definitionssprache in Bild 9 dargestellt. Das gepunktete Rechteck im Bild umfasst die normierten Komponenten des MPEG-7-Standards. Die Pfeile von der Beschreibungs-Definitionssprache zu den Beschreibungs-Schemen zeigen, dass die Beschreibungs-Schemen aus der Beschreibungs-Definitionssprache generiert werden. Darüber hinaus beschreibt die Zeichnung die Tatsache, dass ein neues Beschreibungs-Schema aus einem bestehenden Beschreibungs-Schema abgeleitet werden kann.

In Bild 10 wird angedeutet, dass die Beschreibungs-Definitionssprache den Mechanismus zum Aufbau von Beschreibungs-Schemen be-

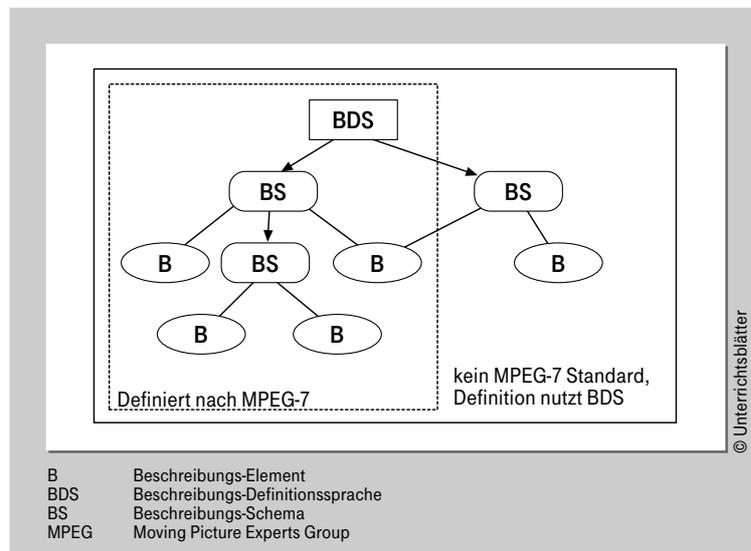


Bild 9: Beziehungen zwischen Beschreibungs-Elementen und Beschreibungs-Schemen bei MPEG-7

sitzt, welche die Grundlage für den Beschreibungs-Generator bildet. (Die Instanzen der Beschreibungs-Schemen sind in Bild 4 zu entnehmen.)

Das Bild 11 stellt die Funktionsweise dar, wie MPEG-7 zukünftig praktisch arbeiten wird. Es zeigt die Arbeitsweise ausgehend vom multimedialen Material (Inhalt) 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.

Die Betonung bei MPEG-7 liegt auf der Bereitstellung von neuartigen Lösungen für audiovisuelle Inhaltsbeschreibungen. Deshalb sind die reinen textorientierten Dokumente nicht das Ziel von MPEG-7. Die audiovisuellen Inhalte können jedoch im Zusammenhang mit audiovisuellen Informationen textuelle Dokumente beinhalten oder auf diese verweisen. Daher wird erwo-

gen, die bestehenden Standards für „nur-Text-Dokumente“ von anderen Organisationen zu unterstützen und anzupassen.

Bei den Beschreibungs-Elementen selbst spielt die Datenbankstruktur eine kritische Rolle in der abschließenden informationsfindenden Ausführung. Ein schneller Suchprozess, ob ein Inhalt von Interesse und somit zutreffend wäre, erlaubt erst die Indexierung der Information, die damit strukturiert wird. Hier gibt es hierarchische und assoziative Verfahren.

4.2 MPEG-7 Informationssuche von Anwendungen

Es gibt viele Anwendungen und Anwendungsfelder, die einen Vorteil aus der einheitlichen Informations-

³⁶ Interoperabilität: Fähigkeit, zur Erfüllung von Aufträgen gemeinsam im nationalen und internationalen Einsatz zusammenzuwirken.

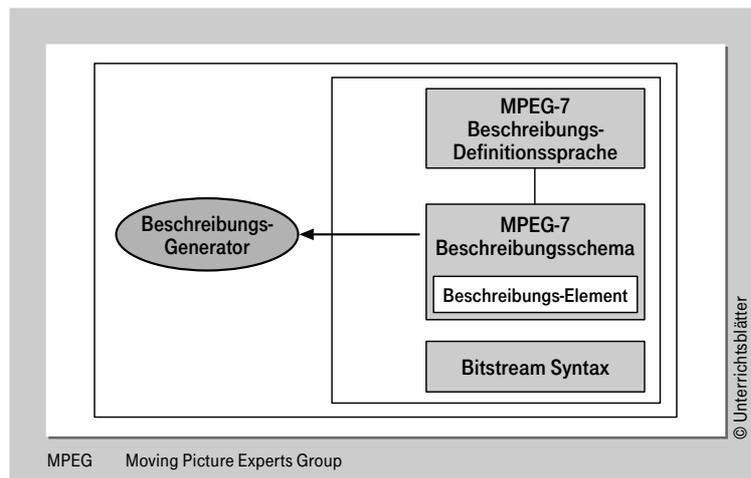
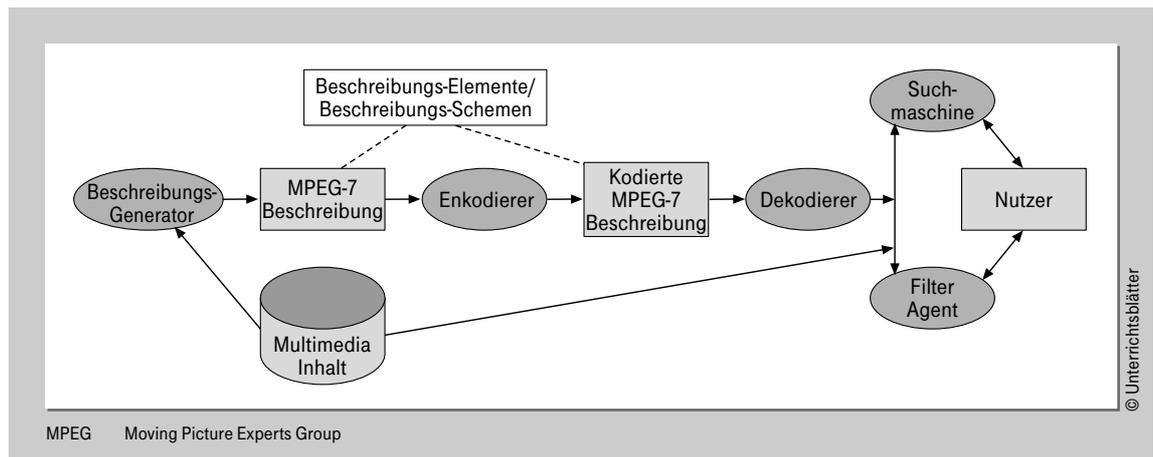


Bild 10: Beschreibungs-Definitionssprache, Beschreibungs-Schema und Beschreibungs-Generator

Bild 11:
Grobes Konzept
der Arbeitsweise
von Anwendungen
bei MPEG-7



suche des MPEG-7-Standards ziehen. Einige der wichtigsten Anwendungen wären:

- digitale Bibliotheken (Bild-Kataloge, Musik-Wörterbücher),
- Multimedia-Verzeichnisdienste (z. B. Gelbe Seiten)
- Programmführer für Rundfunk bzw. Landesmedien (Radio- und TV-Sender),
- Multimedia-Bearbeitung (personalisierter elektronischer Nachrichtendienst, Erstellung von Media-Diensten).

Leistungsfähige 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 Retrieval³⁷-Dienste (Wiedererkennung von menschlichen Eigenschaften, Gerichtsmedizin),
- geografische Informationssysteme (Regionale Touristik-Informationen, universitäre Auswertungen Ökologischer Systeme),
- entfernte Aufzeichnungs-Dienste (Landkarten, Ökologie, Kontrolle der Naturschätze),
- Überwachung (Verkehr, Sicherheit, oberirdische Beförderung, militärische Operationen),
- biomedizinische Anwendungen (Altenversorgung, Teletherapie),
- Einkauf (Kleidung)

³⁷ Retrieval: Das Suchen und Auffinden gespeicherter Daten in einer Datenbank.

- Haus- und Garten-Dienste (Architektur, Immobilien, Wohnungsdesign),
- soziale Dienste (Single-Club, Treffpunkte, Vereine),
- Medien-Archive (Film, Video, Radio).

Mit welchem Verfahren die MPEG-7-Daten genutzt werden, um Antworten auf Nutzeranfragen zu finden, liegt nicht im Fokus des Standards. Grundsätzlich kann gesagt werden, dass jede Art von audiovisuellen Materialien gefunden werden kann, wenn jede Art der Datenanfrage unterstützt wird. Das heißt, dass zum Beispiel ein Video-Material erfragt werden könnte, wenn Video, Audio oder Sprache einzeln genannt werden. Es ist die Aufgabe der Suchmaschine, die passenden Daten der Suchanfrage und der MPEG-7-Beschreibungs-Elemente zusammenzubringen. Einige Beispiele für Suchanfragen werden hier angeführt:

- 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. Durchsuche diese nach den interessantesten Objekten und ergänze damit dein eigenes Bild.
- Film und Video: Gib einen Satz von Objekten an, welche die Bewegungen und Beziehungen der Objekte beschreiben. Dann er-

halte eine Liste der zutreffenden Animationen zurück, die temporär und im direkten Zugriff stehen.

- Sprache: Nimm eine Sprachprobe der Stimme eines Sängers und erhalte eine Liste seiner Audioaufnahmen und Video-Clips zurück, in denen er singt oder auftritt.
- Szenarien: Beschreibe die Handlungen einer Szene (z. B. eines lachenden Clowns) und erhalte eine Liste von ähnlichen Szenen mit ähnlichen Handlungen zurück (z. B. ein weinender Clown und ein lachender Zauberer).

5 Zusammenfassung

Innovative multimediale Dienste werden sowohl 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.

Darüber hinaus ermöglichen sie für die Provider und auch den Kunden eine Zukunftssicherheit für innovative Produkte. Die Aufwärtskompatibilität der MPEG-Standards ermöglicht die Vermeidung von Überschneidung und die Erschließung neuer Dimensionen.

(Ge)

Literaturhinweise

- [1] International Organisation For Standardisation – ISO/IEC JTC1/SC29/WG11, ISO/IEC JTC1/SC29/WG11 N2323, MPEG-4 Overview, Final-Status, Dublin, März 1999.
- [2] International Organisation For Standardisation – ISO/IEC JTC1/SC29/WG11, MPEG Draft: MPEG-7 Context and Objectives (Version 10) ISO/IEC JTC1/SC29/WG11 N2329, Approved-Status, Dublin, Oktober 1998.

Glossar

Advanced Audio Coding (AAC)

Datenkompressionsverfahren, welches speziell für akustische Signale entwickelt wurde.

Adaptive Differential Pulse Code Modulation (ADPCM)

dt. adaptive Delta-Pulsmodulation. Zunächst in ITU-T G.721 standardisiertes Verfahren zur komprimierten Übertragung telefonischer Sprachsignale.

Asymmetric Digital Subscriber Line (ADSL)

Asymmetrisches DSL-System zur Realisierung breitbandiger Datenanschlüsse über das herkömmliche Kupferkabel-Anschlussnetz.

Amplitudenmodulation (AM)

Modulationsverfahren, bei dem die zu übertragende Nachricht der Amplitude der meist sinusförmigen Trägerschwingung aufgeprägt wird.

Asynchronous Transfer Mode (ATM)

Weitgehend standardisierte Übertragungs- und Vermittlungstechnologie, um den Breitbandbereich von 2 Mbit/s bis zu derzeit 622 Mbit/s abzudecken.

Balance

Beim Audiosignal die Position im Stereofeld.

Binary Format for Scene description (BIFS)

Im Rahmen des Kodier- und Kompressionsverfahrens MPEG-4 für die digitale Video- und Audiokommunikation eine neu entwickelte Beschreibungssprache zur Ergänzung der kodierten Bilddaten.

Code-Excited Linear Prediction (CELP)

Modernes Sprach-Kompressionsverfahren für Bandbreiten zwischen 2,4 kbit/s und 8 kbit/s, das hinsichtlich seiner Sprachqualität ADPCM (ITU-T G.726) entspricht.

Chorus

Beim Audiosignal die Modulation des Ausgangssignals mit einem Zeitversatz.

DMIF Application Interface (DAI)

Eine Systemspezifikation, die ein generisches Interface zur Transportschicht beschreibt.

Diskrete Cosinus-Transformation (DCT)

Verfahren der linearen Transformation, das meistens als erster Schritt bei der digitalen Bildkompression angewendet wird.

Data Description Language (DDL)

Dateibesprechungssprache, welche bei Datenbanksystemen die Sprache, die dem Datenbankverwalter die Definition von Datenstrukturen erlaubt.

Delivery Multimedia Integration Framework (DMIF)

Für jedes mögliche Transportnetzwerk, sei es ein IP-Netzwerk, ein DVB-TV-Netz oder ein ATM-Netz, ist eine Anpassung zu definieren.

Digital Video Broadcasting (DVB)

Europäischer Standard digitaler Fernsehverteilungsnetze, die über Satelliten und auch terrestrisch ausgestrahlt werden.

Digital Versatile Disc (DVD)

(von engl. versatile = vielseitig) Ursprünglich Digital Video Disc.

Echo

Durch Reflexion der elektromagnetischen Wellen auf dem Übertragungsweg verursachte, dem akustischen Echo vergleichbare Erscheinung.

Equalizer

Klangregleinrichtung an Verstärkern zur Korrektur, Entzerrung oder sonstigen Beeinflussung des Klangbildes, wobei der gesamte Tonfrequenzbereich in mehrere Bänder aufgeteilt wird.

Flanger

Beim Audiosignal die Modulation mit einem sich verändernden Frequenzbereich.

Frequenzmodulation (FM)

Elektrisches Modulationsverfahren, bei dem die zu übertragende Nachricht (Signal) der Frequenz der meist sinusförmigen Trägerschwingung aufgeprägt wird.

ITU-T-Empfehlung H.261

Standard für Grundprinzipien der digitalen Bewegtbildübertragung.

ITU-T-Empfehlung H.263

Für die digitale Videokonferenz. Sie spezifiziert ein Kodierverfahren für Videodaten für die Bildtelefonie.

ITU-T-Empfehlung H.320

Für den Videokonferenzdienst über ISDN-Leitungen (aber auch für LAN); legt mehrere Normen für die Audio- und Videoübertragung fest.

ITU-T-Empfehlung H.324

Für Multimediaübertragungen bei geringen Bitraten; beinhaltet verschiedene weitere Empfehlungen wie einen Video-Codec für V.34-Modemübertragungen über analoge Telefonleitungen sowie ATM und ISDN.

Hall

Beim Audiosignal die Anreicherung des Signals mit Raumsimulationsanteilen (Reflexionen).

High Definition Television (HDTV)

Hochauflösendes Fernsehkonzept, vorzugsweise für Satellitenübertragung, im Kinoformat 16:9 mit 1250 horizontalen Zeilen und hoher Bildwechselfrequenz.

Harmonic and Individual Line plus Noise (HILN)

Der Kodierer identifiziert die harmonischen und dominierenden Frequenzanteile und lässt weiterhin die Rauschanteile bestehen.

Harmonic Vector eXcitation Coding (HVXC)

Der Kodierer identifiziert die dominierenden Signalanteile. Die sekundären Signalanteile werden zusätzlich übertragen. Diese werden in eine digitale Form umgewandelt und anschließend in parametrische Werte verschlüsselt.

Hyperlink

Im Internet für Hypertext Link. Darunter versteht man besonders hervorgehobene Strings oder Symbole in Texten, hinter denen sich logische Verbindungen zu anderen Dokumenten verbergen.

Internet Protocol (IP)

Ein wichtiges Internetprotokoll (RFC 791) der TCP/IP-Protokollfamilie der Internet-Schicht (entsprechend Schicht 3 des OSI-Referenzmodells) mit Funktionen zur Leitweg- und Flusskontrolle.

International Electrotechnical Commission (IEC)

Seit ihrer Gründung im Jahre 1906 erstellt und veröffentlicht die Internationale Elektrotechnische Kommission Internationale Standards (IS) für die Elektrotechnik, die Elektronik und andere verwandte Gebiete.

International Standard Organization (ISO)

Die Internationale Organisation für Standardisierung ist ein weltweit zugelassener Zusammenschluss von nationalen Standardisierungsgremien aus etwa 130 Ländern.

International Telecommunications Union (ITU)

Die ITU ist der älteste heute noch bestehende Zusammenschluss auf dem Gebiet der Telekommunikation. Ziele und Aufgaben der ITU sind im Internationalen Fernmeldevertrag niedergelegt.

Internetworking Units (IWU)

Allgemein für Gateway zur Unterstützung der netz-überschreitenden Kommunikation.

Joint Photographic Experts Group (JPEG)

Standard zum Speichern und Komprimieren von digitalisierten Fotografien oder Standbildern, der auch „JPEG File Interchange Format (JFIF)“ genannt wird.

Kompression

Reduzierung digitaler Daten (Audio- und Videodaten), um sie schneller übertragen zu können, wobei redundante Daten zusammengefasst werden. Die Daten werden beim Empfänger wieder dekomprimiert.

Logical Link Control (LLC)

Teilschicht der Schicht 2 im OSI-Referenzmodell, welche für diverse LAN-Protokolle der Serie IEEE 802 spezifiziert ist.

Local Multipoint Distribution System/Service (LMDS)

Ein in den USA im 28 GHz-Band mit 1,3 GHz Bandbreite vorgesehenes drahtloses Teilnehmer-Anschlussystem mit einer Punkt-zu-Mehrpunkt-Technik.

Linear Predictive Coding (LPC)

Digitales Quellkodierungsverfahren für Sprache, welches digitale Sprachübertragung mit 2,4 kbit/s bis 4,8 kbit/s mittels MODEM über analoge Telefonkanäle ermöglicht.

Musical Instruments Digital Interface (MIDI)

Standard, der die Kommunikation zwischen elektronischen Musikinstrumenten untereinander und mit Rechnern über digitale Steuerdaten im 8-Bit-Format spezifiziert.

Multichannel Microwave Distribution System/Service (MMDS)

Allgemein für ein breitbandiges mehrkanaliges Funksystem bzw. Funkdienst für TV- und Multimedia-Anwendungen, beispielsweise im 40-GHz-Bereich.

Motion Picture Experts Group (MPEG)

Gemeinsames Gremium von ISO und IEC, das sich mit der Standardisierung von Kodier- und Kompressionsverfahren für die digitale Video- und Audiokommunikation befasst.

Pulse Code Modulation (PCM)

Verfahren zur Digitalisierung und Mehrkanalübertragung analoger Quellsignale, insbesondere von Telefonsignalen.

Personal Data Assistant (PDA)

Auch als Handheld PC bezeichnete Rechner, die so klein sind, dass sie in einer Hand gehalten werden können, während die andere Hand Daten eingibt.

Public Switched Telephone Network (PSTN)

In der ITU gebräuchliche Bezeichnung für verbindungsorientierte/leitungsvermittelte analoge Telefonnetze, über die das Versenden digitaler Daten im Modemverfahren erfolgt.

Quarter Common Intermediate Format (QCIF)

Vom CIF abgeleitetes Format (H.263) für die Übertragung von Bildern in der Bildtelefonie mit 176×144 Pixeln.

Structured Audio-Format (SA)

Es besteht aus einer Instrumenten-Beschreibungssprache SAOL und einem Steuercode SASL.

SA Orchestra Language (SAOL)

Neben einer Programmiersprache enthält sie eine umfassende Bibliothek von etwa 100 vorinstallierten Audio-Funktionen, um den Klang eines Instrumentes zu bestimmen.

SA Score Language (SASL)

Ein Steuercode, der den Eigenschaften einer Partitur nachempfunden wurde. Hier können u. a. Tremolo und Glissando umgesetzt werden.

subQCIF

Wie QCIF, jedoch mit 128×96 Pixeln.

TIMES

Telecommunications, Informationstechnology, Multimedia, Entertainment and Security: Telekommunikation, Informationstechnik, Multimedia, Unterhaltung und Sicherheitstechnik.

Telecommunication Management Network (TMN)

Architekturrahmen für die Standardisierung von Betreiberaufgaben und technischen Strukturen von Telekommunikationsnetzen.

Text To Speech (Conversion) (TTS)

Bezeichnung für die Umwandlung von Text in Sprache mittels Sprachsynthese.

Universal Mobile Telecommunications System (UMTS)

ETSI-Projekt eines künftigen, universell nutzbaren Mobilfunknetzes der so genannten dritten Generation im Frequenzband von 2 GHz.

Virtual Reality Modelling Language (VRML)

Eine Beschreibungssprache, die eine räumliche, virtuelle, grafische Darstellung ermöglicht.