

Audiocodierung: Vom Hörfunkstandard zum Advanced Audio Coding

Peter Noll, Institut für Nachrichtentechnik und Theoretische Elektrotechnik der TU Berlin

Prof. Dr.-Ing. Peter Noll, Promotion 1969, Habilitation 1973, beides an der TU Berlin, war von 1976 bis 1980 o. Professor für Statistische Nachrichtentheorie an der Universität Bremen und ist seit 1980 Professor für Fernmeldetechnik an der Technischen Universität Berlin. Er hat fünf Jahre im Rahmen der MPEG-Standardisierungsaktivitäten die Audio-Arbeitsgruppe geleitet. Lehr- und Forschungsschwerpunkte: Digitale Nachrichtenverarbeitung und -übertragung, Audio- und Bildcodierung für ATM-, Internet- und Mobilfunksysteme.

Wer eine Audio-CD (CD = Compact Disc) abspielt, kennt die hohe Wiedergabequalität dieses Speichermediums. Allerdings begrenzt die zur digitalen Darstellung des Audiosignals erforderliche hohe Bitrate die Spieldauer einer CD auf etwa 70 Minuten. Die Bitrate des digitalisierten Audiosignals liegt bei etwa 1,4 Mbit/s, zusätzlich sind fast 3 Mbit/s erforderlich, um das zu speichernde Binärsignal an die Eigenschaften des Speichermediums anzupassen und den notwendigen Fehler-schutz sicherzustellen.

Verfahren zur Datenkompression müssen eingesetzt werden, wenn bei gleichbleibendem Speichermedium höhere Spieldauern erreicht werden sollen, oder wenn in Übertragungssystemen nur geringe Bitraten zur Verfügung stehen, wie z.B. im terrestrischen oder satellitengestützten digitalen Hörfunk (DAB = digital audio broadcast) oder in zukünftigen mobilen Multimedia-Anwendungen.

Im Rahmen der internationalen MPEG-Standardisierungsaktivitäten (MPEG = moving picture experts group) wurden Verfahren der Datenkompression vorgeschlagen, optimiert und standardisiert, die eine Datenkompression um den Faktor 6 bis 12 zulassen, ohne daß gegenüber der als Referenz dienenden CD-Qualität wahrnehmbare Unterschiede entstehen. Im Folgenden werden die Prinzipien dieser Verfahren erläutert. Als Anwendungsbeispiele werden der MPEG-Standard für den digitalen Hörfunk (MPEG-1) und der MPEG Advanced Audio Coding (AAC) Standard dargestellt. AAC erzielt die höchsten Kompressionsgewinne. Schließlich werden die weite Bitratenbereiche erfassenden Möglichkeiten einer Sprach- und Audiocodierung für Multimedia-Anwendungen benannt (MPEG-4).

Audio Coding: From Broadcast Standard(s) to Advanced Audio Coding

When playing an Audio-CD (Compact Disc), we notice the high quality of audio representation of this storage medium. The digital representation of the audio signal demands a high bit rate which limits the replay time ?? of a CD to around 70 minutes. The bit rate of the digitized audio signal is around 1.4 Mbit/s, additionally, almost 3 Mbit/s are needed to adapt the binary signal to be stored to the properties of the storage medium, and to ensure the necessary error protection.

Data compression techniques have to be implemented to achieve higher replay times ?? or to allow for transmitting over channels which accept only low bit rates. Examples are terrestrial and satellite-based digital audio broadcast services or future mobile multimedia applications.

Within the framework of the international MPEG standardization activities (MPEG = moving picture experts group) data compression techniques have been proposed, optimized, and standardized, which allow for a compression gain between 6 and 12, without causing acoustically noticeable

Autor(en)	Titel	Zeichen	Dateiname	Datum	Seite
Peter Noll, Institut für Nachrichtentechnik und Theoretische Elektrotechnik der TU Berlin	Audiocodierung: Vom Hörfunkstandard zum Advanced Audio Coding	29279	it+ti - Aufsatz_neu	02.07.1999	1 (14)

differences in comparison to CD quality. In the following, the principles of the techniques will be explained. The MPEG standards for the digital audio broadcast (MPEG-1) and the MPEG Advanced Audio Coding (AAC) will serve as application examples. AAC has the highest compression gains. Finally, the MPEG-4 standard offers flexible means and a wide range of bit rates for coding audio and speech signals for multimedia applications.

1 Einleitung

Populäre Speicher speichern Audiosignale in binärer Form im sog. PCM-Format. Dazu werden der linke und der rechte Stereokanal mit 44,1 kHz (Audio-CD) oder 48 kHz (Digital Audio Tape) abgetastet, und jeder Abtastwert wird mit einer Auflösung von 16 bit in binärer Form dargestellt. Es stehen also für jeden abgetasteten Amplitudenwert $2^{16} = 65536$ diskrete Amplitudenstufen zur Verfügung (Bildsignale benötigen hingegen nur $2^8 = 256$ Helligkeit repräsentierende Amplitudenstufen). Die Bitrate für die Quellencodierung des Stereosignals beträgt somit etwa 1,4 bis 1,5 Mbit/s. Die folgende Tabelle vergleicht die wesentlichen Parameter für PCM-Formate von Audiosignalen mit denen von Breitband- und Telefonsprachsignalen. Die Bandbreite von Breitbandsprachsignalen ist doppelt so hoch wie die von herkömmlichen Telefonsignalen. Als Vorteil ergibt sich eine höhere subjektive Qualität der Sprachwiedergabe, insbesondere bei Freisprecheinrichtungen und bei Telekonferenzschaltungen, sowie eine verringerte Fehlerrate bei automatischer Sprecheridentifizierung bzw. Spracherkennung. Weiterhin lassen Übertragungskanäle für Breitbandsprache eine Übertragung von Audiosignalen mit eingeschränkter Qualität zu.

Tabelle 1: Wesentliche Parameter bei einer PCM -Codierung von Sprach- und Audiosignalen

	Frequenzbereich in Hz	Abtastfrequenz in kHz	PCM-Bits je Abtastwert	PCM-Bitrate in kbit/s
Telefonsprachsignale	300 - 3.400	8	8	64
Breitbandsprachsignale	50 - 7.000	16	8	128
Breitbandaudiosignale (mono)	10 - 20.000	48 ¹	16	768

Es waren große Anstrengungen erforderlich, um die für Breitbandaudiosignale erforderlichen Bitraten bei gleichbleibender subjektiver Wiedergabequalität so zu reduzieren, daß sie für die drahtlose und mobile Übertragung sowie für begleitende Audiosignale in Videokonferenz- und anderen Multimedia-Anwendungen geeignet waren [1].

Im Rahmen der internationalen ISO-Standardisierungsaktivitäten der Moving Picture Experts Group (MPEG) wurden seit 1988 Verfahren zur Datenkompression vorgeschlagen, optimiert und standardisiert. Da an diesen weltweiten Aktivitäten Wissenschaftler aus der Industrie, aus Universitäten und aus Forschungseinrichtungen der Rundfunkanstalten teilnahmen, entstanden Standards, die wegen ihrer hohen Leistungsfähigkeit viele Anwendungen fanden, vor allem im Digitalen Hörfunk (DAB = digital audio broadcast), im Digitalen Fernsehen (DVB = digital video broadcast) und in Internetanwendungen.

Die standardisierten Verfahren lassen eine Datenkompression um den Faktor 6 bis 12 zu, ohne daß gegenüber der als Referenz dienenden CD-Qualität wahrnehmbare Unterschiede entstehen. Im Folgenden werden die Prinzipien dieser Verfahren erläutert. Als Anwendungsbeispiel wird die

¹ Die CD verwendet eine Abtastfrequenz von 44,1 kHz, in Anwendungen mit sog. UKW-Qualität /Frequenzbereich 50 - 15.000 Hz) wird auch 32 kHz verwendet.

Verwendung eines MPEG-Standards im digitalen Hörfunk dargestellt (MPEG-1), weiterhin Gründe dafür angegeben, daß der kürzlich standardisierte MPEG Advanced Audio Coding (AAC) Standard die höchsten Kompressionsgewinne erzielt. Schließlich werden die flexiblen Möglichkeiten des MPEG-4 Standards zur Sprach- und Audiocodierung im Bitratenbereich von 2 kbit/s bis 64 kbit/s für Multimedia-Anwendungen benannt. Damit steht jetzt eine große Zahl von MPEG-Standards für Sprach- und Audiosignale im Bitratenbereich von 2 kbit/s (Vocoder-Sprache) bis zu mehreren hundert kbit/s für Mehrkanalübertragungen zur Verfügung.

Wir merken schließlich noch an, daß höchsten Qualitätsansprüchen Abtastfrequenzen bis zu 96 kHz und amplitudenmäßige Auflösungen von 24 Bits und mehr vorgesehen werden müssen, um zu verhindern, daß Codierverzerrungen akkumulieren. Sollen Audiosignale archiviert werden, so wird häufig gefordert, daß die PCM-Formate nicht einer Datenkompression unterworfen werden dürfen, die zusätzliche Verzerrungen erzeugt. In solchen Fällen lassen sich Verfahren der verlustlosen Datenkompression einsetzen, der Kompressionsgewinn liegt dann aber nur in der Größenordnung von 2.

2 Anwendungen der MPEG-Audio-Codierungsstandards

Anwendungen umfassen alle Bereiche der digitalen Speicherung und Übertragung, von denen einige hier aufgeführt werden sollen:

Digitaler Hörfunk und Internet. Der digitale Hörfunk wird in Deutschland zur Zeit flächendeckend eingeführt. Er verwendet zur Quellencodierung 256 kbit/s (Stereo), läßt aber auch andere Bitraten zu. Niedrigere Bitraten spielen vor allem in Satellitenrundfunksystemen und in Internet Übertragungen eine große Rolle. Dabei sind bei der Verteilung von Rundfunksignalen im Internet bei Echtzeitwiedergabe (sogenanntem *streaming*) zur Zeit noch Qualitätseinbußen durch die unvermeidlichen Paketverluste hinzunehmen, die durch schwankende Belastungen des Netzes und dadurch eintretende stochastische Verzögerungen verursacht werden.

Von aktuellem Interesse sind Untersuchungen, in denen die Möglichkeiten einer digitalen Übertragung von Audiosignalen über amplitudenmodulierte Übertragungstrecken (Mittel- und Langwelle) erkundet werden.

Audioübertragung in ISDN-Netzen. MPEG-Verfahren erlauben inzwischen eine hochwertige Digitalisierung von stereophonen Audiosignalen bei Bitraten im Bereich von 64 bis 128 kbit/s. Es besteht damit die Möglichkeit, Audiosignale über ISDN-Telefonleitungen zu übertragen. Anwendungen ergeben sich beispielsweise im Rundfunk- und Fernsehbereich, bei der Zuspielung solcher Signale oder bei individuellem Abruf von Audiosignalen aus Tondatenarchiven.

Multimedia-Systeme. Die MPEG-Aktivitäten haben von Anfang an auf Multimedia- und Computer-Anwendungsmöglichkeiten gezielt, daher stehen heute viele Algorithmen zur effizienten Audio- und Videocodierung zur Verfügung, insbesondere auch PC-gestützte Systeme zur Aufnahme und Wiedergabe von MPEG-Audio- und Bildsequenzen. Die MPEG-Standards erlauben dabei auch durch geeignete Datenformate eine sehr flexible Handhabung von Multimediasignalen.

Digitale Speichermedien. Viele Einsatzmöglichkeiten bieten sich auch bei der digitalen Speicherung von Tonsignalen auf magnetischen oder optischen Medien. Von zunehmendem Interesse ist die Realisierung digitaler Aufnahmegeräte ohne bewegte Teile mittels Speicherchips (64 oder 256 Mbit/Chip). Bei einer Bitrate von 64 oder 128 kbit/s je Stereosignal sind inzwischen mit solchen *audio cards* Spieldauern realisierbar, die denen der Audio-CD entsprechen.

Digitaler Fernsehfunk. Im Digitalen Fernsehfunk (DVB = *Digital Video Broadcast*) werden auch an die entsprechende Tonkomponenten hohe Anforderungen bezüglich Qualität und geringer Bitrate gestellt. Mit der Annäherung des Fernsehens an Eigenschaften des Kinofilms (Breitwandformat 16:9, Surround-Klang) wird eine effiziente Datenreduktion vor allem des Mehrkanaltons unumgänglich. Der MPEG-2 Standard stellt entsprechende Verfahren zur Verfügung.

3 Nutzbare psychoakustische Eigenschaften des Gehörs

Das von einer Schallquelle in Form von Luftdruckschwankungen ausgehende Signal wird im Gehör einer vielschichtigen Verarbeitung unterzogen. Die Schallwelle wird zuerst als mechanische Schwingung durch das Außen- und Mittelohr an die Cochlea weitergeleitet, in der, vereinfacht ausgedrückt, eine Kurzzeit-Frequenzanalyse durchgeführt wird. Dabei entstehen entlang der Basilar-membran Aktionspotentiale, die über Nervenfasern zum Gehirn weitergeleitet werden. Die Wahrnehmbarkeit von akustischen Reizen wird durch dieses komplexe System bestimmt. Einige Eigenschaften des Gehörs können zur Datenkompression genutzt werden. So müssen akustische Reize zumindest partiell die sog. Ruheshwelle überschreiten, damit sie wahrnehmbar werden. Bild 1 zeigt den Verlauf dieser Ruheshwelle über der Frequenz. Es wird deutlich, daß akustische Reize bei sehr tiefen und sehr hohen Frequenzen einen hohen Schalldruckpegel besitzen müssen, damit sie überhaupt hörbar werden. Signale, die in der hellgrau unterlegten Fläche unterhalb der Ruheshwelle liegen, sind nicht wahrnehmbar und können bei einer Audiocodierung vernachlässigt werden.

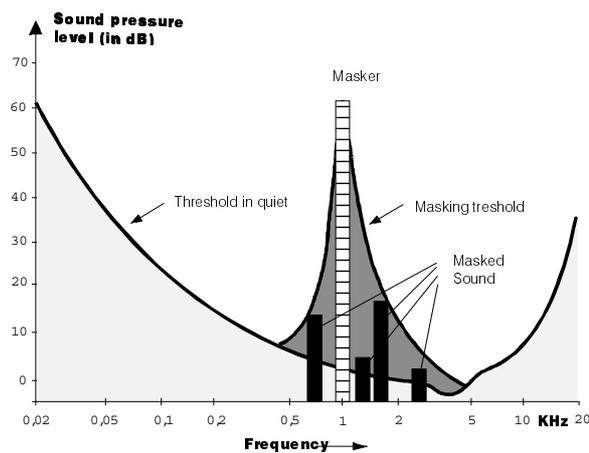


Bild 1: Ruheshwelle und Beispiel einer Verdeckung

Die Eigenschaft der akustischen Verdeckung (auch Maskierung genannt) hat eine noch größere Bedeutung für die Datenkompression [2]. Betrachten wir einen in der Umgebung von der Frequenz 1 kHz liegenden akustischen Reiz, den sog. Maskierer. In der frequenzmäßigen Umgebung dieses Maskierers tritt eine Hörschwelle auf, die im Bild 1 dunkelgrau hinterlegt ist. Solange der Maskierer vorhanden ist, können akustische Reize mit einem Schalldruckpegel unterhalb der Verdeckungsschwelle nicht wahrgenommen werden (schwarze Balken im Bild 1). Das gilt sowohl für andere, gleichzeitig auftretende akustische Reize wie auch für Verzerrungen, die z.B. durch Quantisierungsrauschen oder durch Kanalfehler entstanden sind.

Wird ein Audiosignal digitalisiert, so entsteht ein Quantisierungsrauschen, dessen Schalldruckpegel bei hoher Amplitudenauflösung weit unter dem Pegel des maskierenden Audiosignals liegt. Der Schalldruckpegel des Quantisierungsrauschens erhöht sich jeweils um 6 dB, wenn die Bitanzahl für die Quantisierung jeweils um ein Bit verringert wird (dadurch halbiert sich die Amplitudenauflösung). Es ist jetzt möglich, diesem Quantisierer weitere Bits zu entziehen, bis das Quantisierungsrauschen gerade noch unterhalb der Hörschwelle liegt, es wird dann nicht wahrnehmbar sein. Bei komplexen, über den ganzen Frequenzbereich verteilten akustischen Ereignissen können viele Maskierer identifiziert werden, deren individuelle Verdeckungsschwellen zusammen mit der Ruhehörschwelle zu einer globalen Verdeckungsschwelle zusammengefaßt werden können.

Die globale Verdeckungsschwelle kann in Verfahren der *wahrnehmungsangepaßten Audiocodierung* zur Datenkompression genutzt werden [3]. Dazu wird zuerst das Leistungsdichtespektrum des zu codierenden Audiosignals blockweise mittels einer Kurzzeit-Frequenzanalyse bestimmt. Aus dem Leistungsdichtespektrum ergibt sich die korrespondierende globale Verdeckungsschwelle, dabei werden alle nutzbaren Kenntnisse über akustische Verdeckungseffekte ausgenutzt. Aus dem Leistungsdichtespektrum und der globalen Verdeckungsschwelle ergibt sich die erforderliche Anzahl von Bits pro Abtastwert in unterschiedlichen Frequenzbereichen des Audiosignals,

4 MPEG-basierte Audiocodierung

Die MPEG Audiocodierungsstandards bieten Algorithmen unterschiedlicher Komplexität und Qualität an. Wir beschränken uns auf die sog. MPEG-1 Layer 2 - Codierung, die u.a. im Digitalen Hörfunk verwendet wird, auf die MPEG-1 Layer 3 - Codierung, die u.a. in Satellitenübertragungssystemen und im Internet (hier unter der Bezeichnung MP 3) verwendet wird, und auf den MPEG-2 AAC - Algorithmus. Alle MPEG - Codierer führen die Codierung der Audiosignale im Frequenzbereich vor, entweder durch eine Teilbandaufteilung auf 32 Bandpaßsignale (Layer 2) oder durch eine hochauflösende lineare Frequenztransformation (AAC) [4 - 7]. Der MP3 - Codierer benutzt eine hybride Filterbank, d.h. eine Kaskadierung von Teilbandaufteilung und Frequenztransformation.

4.1 MPEG-1 Layer 2 - Codierung [5, 6]

Der MPEG-1 Layer 2 - Codierer verwendet eine Teilbandcodierung, bei der das Audiosignal in 32 äquidistante, jeweils 750 Hz breite Teilbandsignale aufgeteilt wird (diese und folgende Werte gelten für eine Abtastfrequenz von 48 kHz). Die frequenzmäßige Aufteilung wird über eine digitale Filterbank mit Polyphasenstruktur vorgenommen. Bild 2 zeigt die Struktur des Layer 2 - Codecs.

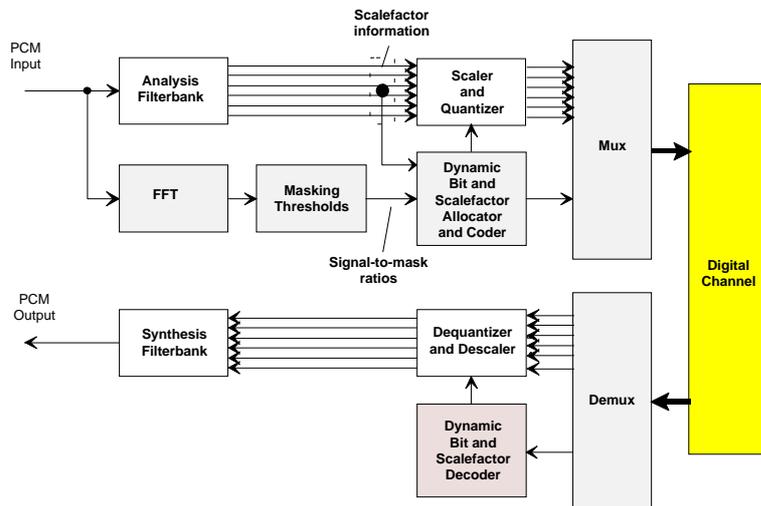


Bild 2: Struktur eines MPEG-1 Layer 2 - Codecs

Die kritisch, d.h. in ihrer Abtastfrequenz um den Faktor 32 reduzierten Teilbandsignale werden quantisiert, wobei eine dynamische Bitzuweisung dafür sorgt, daß in jedem Teilband ein Quantisierer mit einer gerade ausreichenden Amplitudenaufösung zur Verfügung steht, so daß das Quantisierungsrauschen in dem jeweiligen Band maskiert wird. Für die Quantisierung stehen Quantisierer mit zwischen 3 und 65535 Stufen zur Verfügung. Die Bitzuweisung gilt jeweils für 36 Teilbandwerte, also für $36 \times 32 = 1152$ PCM-Eingangswerte des Audiosignals. Der Codierer arbeitet also blockweise mit Blocklängen von $1152 / 48 = 24$ ms. Der Standard unterstützt Abtastfrequenzen von 32, 44,1 und 48 kHz und Bitraten zwischen 32 kbit/s (mono) und 384 kbit/s (stereo).

Durch die begrenzte Zahl der Teilbänder ist die Filterbank nicht in der Lage, die spektrale Genauigkeit des menschlichen Gehörs im Bereich tiefer Frequenzen nachzubilden. Im Gehör werden nämlich akustische Reize, die sich in diesem Bereich frequenzmäßig nur um 50 Hz unterscheiden, unterschiedlich verarbeitet. Um die mangelnde Auflösung der Filterbank zumindest bei der Festlegung der dynamischen Bitzuweisung zu kompensieren, wird parallel zur Teilbandfilterung eine Kurzzeit-Frequenzanalyse mittels einer schnellen Fouriertransformation des Zeitsignals vorgenommen. Aus diesem Spektrum wird die globale Verdeckungsschwelle berechnet, dafür stehen im nicht normativen Teil des MPEG-Standards zwei psychoakustische Modelle zur Verfügung.

Es sei darauf hingewiesen, daß die psychoakustische Modellbildung und Bestimmung der globalen Verdeckungsschwelle und dynamischen Bitzuordnung nur im Codierer notwendig sind. Daher ist die Komplexität des Decodierers geringer als die des Codierers. Durch diese Asymmetrie können reine Wiedergabesysteme im Consumerbereich, beispielsweise DAB-Empfänger und Abspielgeräte, vergleichsweise preisgünstig realisiert werden. Weiterhin ergibt sich dadurch die Möglichkeit, die Qualität der MPEG-Codierer in evolutionärer Weise zu verbessern, z. B. durch Einbeziehung neuer Kenntnisse über die akustische Wahrnehmung, ohne die Empfänger (Decodierer) ändern zu müssen. In der Tat sind nur die MPEG-Decodierer standardisiert.

Das folgende Bild 3 zeigt exemplarisch die dynamische Bitzuordnung für zwei Blöcke ($\hat{=}$ 1152 Abtastwerten) eines Sprachsignals. Block 10 zeigt einen Block mit vorwiegend tieffrequenten Anteilen, der zu einer entsprechend hohen Bitzuordnung in den unteren Teilbändern führt. Block 18 zeigt eine Mischung aus einem schwächer ausgeprägten tieffrequenten Anteil und einem höherfrequenten Anteil. Die Bitzuordnung ist daher eher ausgeglichen.

Wie aufwendige subjektive Tests zeigten, haben MPEG-1 Layer 2 - Codierer eine hervorragende Wiedergabequalität bei einer Bitrate von 256 kbit/s für das Stereosignal, daher wird dieser Standard im digitalen Hörfunk verwendet.

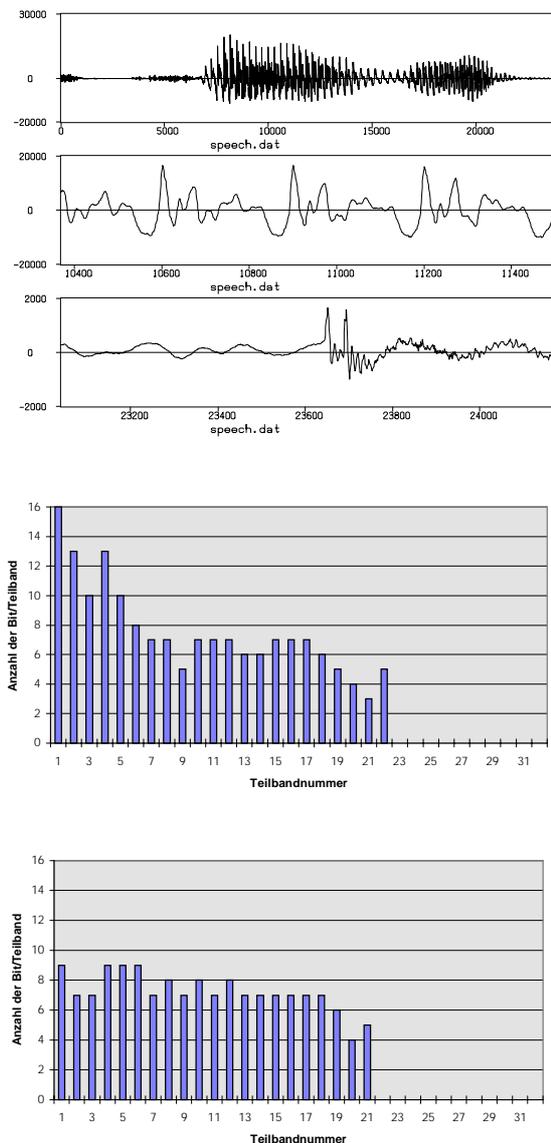


Bild 3: MPEG-1 Layer 2 - Codierung von zwei Blöcken eines Sprachsignals. a) Zeitsignal (20 Blöcke à 1152 Abtastwerten sowie Zeitverläufe der Blöcke 10 und 18 (b) Bitzuordnung für Block 10 (c) Bitzuordnung für Block 18.

4.2 MP3 - Audiocodierung [6, 8]

Die MPEG-1 Layer 3 - Codierung hat unter der Kurzbezeichnung MP3 insbesondere in Internetanwendungen eine große Bedeutung erlangt. Es wurde bereits erwähnt, daß die MPEG-Polyphasenfilterbank durch die begrenzte Zahl der Teilbänder nicht in der Lage ist, die spektrale Genauigkeit des menschlichen Gehörs im Bereich tiefer Frequenzen nachzubilden. Im MP3 - Codierer wird den Teilbandsignalen der Filterbank daher jeweils eine diskrete lineare Transformation mit Blocklängen von 18 oder 6 Werten nachgeschaltet. Als Transformation wird die modifizierte

diskrete Cosinustransformation (MDCT) eingesetzt. Es entsteht damit eine Frequenzauflösung, die um den Faktor 18 bzw. 6 gegenüber dem Layer 2 - Codierer erhöht ist (siehe Bild 4). Dadurch verringert sich die Bandbreite der Frequenzkomponenten im günstigsten Fall auf $24000/(32 \times 18) = 41,67$ Hz. Eine Umschaltung auf kurze Blocklängen und damit auf eine bessere Zeitauflösung wird vorgenommen, wenn in Teilbandsignalen nichtstationäres Verhalten auftritt. Die Umschaltung wird dabei nicht zwangsweise auf alle Teilbandsignale angewendet.

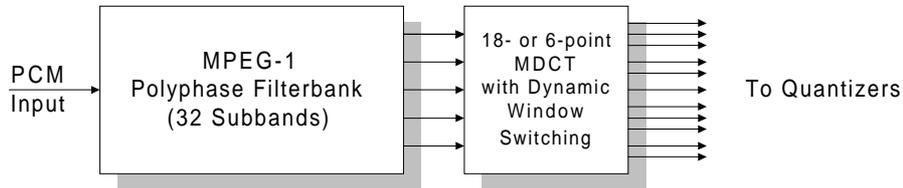


Bild 4: Hybride Filterbank der MP3 - Codierer.

Die Qualität des MP3-Codierers erhöht sich nicht nur wegen der besseren Frequenzauflösung, sondern auch wegen der Quantisierung mit ungleichförmigen Abstufungen. Ein weiterer Beitrag zur Datenkompression ergibt sich, weil den Quantisiererindizes gemäß ihrer Auftretenswahrscheinlichkeiten binäre Codewörter unterschiedlicher Länge zugeordnet werden. Durch diese verlustfreie Entropiecodierung verringert sich die Bitrate, zusätzlich werden Folgen von gleichen Quantisiererindizes durch eine nachgeschaltete Lauflängencodierung reduziert.

MP3 - Codierer bieten eine hervorragende Wiedergabequalität bei Stereosignal-Bitraten von 192 kbit/s, für viele Audiosignale auch bei 128 kbit/s. Sie lassen weiterhin eine Codierung mit variabler Bitrate zu, für die es vielseitige Anwendungen gibt.

5 Codierung für den Digitalen Hörfunk (DAB)

Der standardisierte digitale Hörfunk verwendet zur Datenkompression den MPEG-1 Layer 2 - Algorithmus, ein Stereosignal wird mit einer Bitrate von 256 kbit/s übertragen. Für Kommentarkanäle steht auch der MP3 - Codierer mit niedrigeren Bitraten zur Verfügung. Um eine robuste Übertragung der quellencodierten und damit fehleranfälligen Audiosignale über Funkkanäle sicherzustellen, die nicht allein durch additives Rauschen, sondern vor allem durch Mehrwegeempfang gestört werden, wird in DAB-Systemen ein Mehrträgersystem eingesetzt. Die individuellen Bits eines jeden Audioblocks werden in Abhängigkeit von ihrer Fehlerempfindlichkeit durch eine Faltungscodierung gegen Kanalfehler geschützt (siehe unten). Die Fehlerempfindlichkeiten der einzelnen Bits eines Audioblockes wurden durch subjektive Messungen bestimmt [9].

Bild 5 zeigt die erreichbare subjektive Qualität der empfangenen und decodierten Audiosignale in Abhängigkeit vom Signal-Rausch-Abstand C/N auf dem Kanal bei mobilem Empfang mit einer Fahrzeuggeschwindigkeit von 50 km/h. Die hohe Güte des DAB-Verfahrens wird von einem analogen FM-Rundfunkverfahren auch bei sehr störungsarmen Kanälen nicht erreicht, sie ist zudem fast unabhängig von dem aktuellen C/N-Wert. Erst bei geringen SNR-Werten unterhalb von 15 dB, wie sie an Rändern von Versorgungsgebieten auftreten, kann keine akzeptable Qualität mehr erreicht werden. Es kommt dann zu einem totalen Ausfall der Fehlerkorrekturverfahren. Solche Situationen muß der DAB-Empfänger durch Verfahren der Fehlerverschleierung oder durch Stummschaltung auffangen.

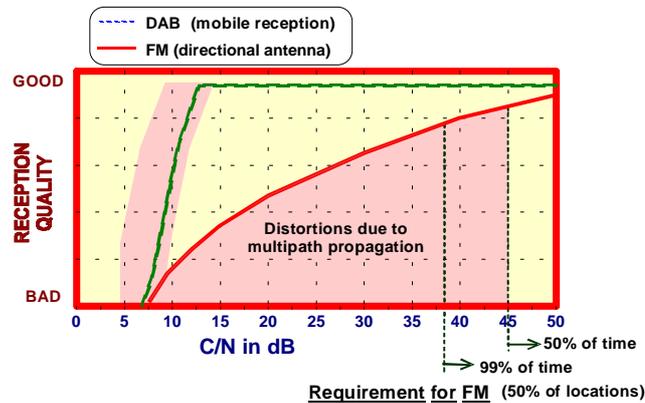


Bild 5: Vergleich der Empfangsqualität von analogem FM- und digitalem Hörfunk. C/N = Kanal-SNR bei mobilem Empfang (50 km/h) und Raleigh-Modellkanal [9]

Bei der zur Fehlerkorrektur eingesetzten Faltungscodierung wird der stärkste Schutz gegen Kanalfehler mit einem Coder der Coderate $\frac{1}{4}$ erreicht, bei der zum Schutz jedes Bits des Quellensignals drei weitere Bits zur Verfügung stehen. Durch sog. Punktieren der durch die Kanalcodierung entstandenen Codebits werden einzelne Bits vor ihrer Übertragung unterdrückt, so daß sich die Coderate erhöht. Die DAB-Spezifikationen lassen Coderaten zwischen $\frac{1}{4}$ und $\frac{8}{9}$ zu. Der soft decision - Viterbidecodierer interpretiert die punktierten Codebits als Auslöschungen, er muß daher nicht an die unterschiedlichen Coderaten angepaßt werden. Das ermöglicht ein dynamisches Umschalten zwischen den Coderaten. So werden in jedem Audiodatenblock die Kopfinformation und die Information über die dynamische Bitzuweisung sehr stark, die Teilbandwerte hingegen nur schwach geschützt. Zusätzlich können die Coderaten an Anwendungen mit unterschiedlichen Übertragungskanälen, wie terrestrische Ausstrahlung, Satellitenfunk und Kabelverteilsysteme, angepaßt werden, ohne Änderungen im Decodierer vornehmen zu müssen.

Bild 6 zeigt beispielhaft die Abhängigkeit der nach der Kanaldecodierung verbliebenen Rest-Bitfehlerwahrscheinlichkeit vom Kanal-SNR bei mobilem Empfang (50 km/h) und Raleigh-Kanal. Soll beispielsweise eine Rest-Bitfehlerwahrscheinlichkeit von 10^{-5} nicht überschritten werden, so muß ein Kanal-SNR von mindestens 22 dB sichergestellt sein, wenn die mittlere Coderate $R = \frac{2}{3}$ vorgegeben ist. Bei stärkerem Fehlerschutz mit einer mittleren Coderate $R = \frac{1}{3}$ reicht ein Kanal-SNR von 12 dB.

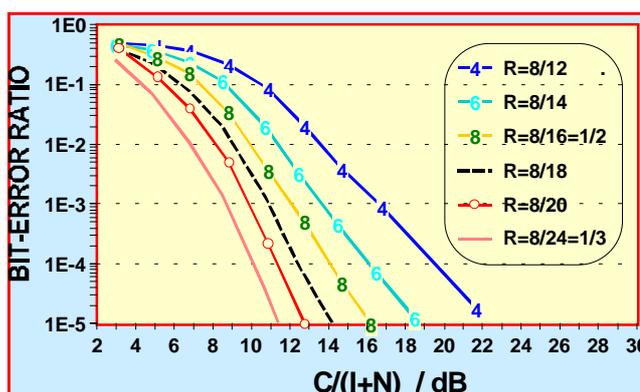


Bild 6: Rest-Bitfehlerwahrscheinlichkeit von dem Kanal-SNR. C/(I+N) = Kanal-SNR bei mobilem Empfang (50 km/h) und Raleigh-Kanal [9]

6 MPEG Advanced Audio Coding [10, 11]

Der MPEG Advanced Audio Coding (AAC)-Standard wurde zuerst für Mehrkanal- (insbesondere 5-Kanal-) Anwendungen entwickelt, er bietet aber auch für Mono- und Stereosignale eine deutlich bessere Qualität als andere MPEG-Audiocodierverfahren. Dazu tragen verschiedene neue Techniken bei: Der AAC-Codierer verwendet eine hochauflösende Filterbank und verzichtet dabei auf suboptimale Hybridlösungen, wie sie der MP3-Codierer benutzt, er benutzt weiterhin Prädiktionsverfahren, schließlich werden die Codebits noch einer sehr effizienten verlustfreien Codierung unterworfen.

Die Filterbank ist eine 2048-Werte-MDCT, d.h. die Transformation wird bei einer 50%igen Überlappung mit Blöcken von 2048 Audioabtastwerten vorgenommen, die durch eine Fensterfunktion gewichtet sind. Die Filterbank kann dynamisch auf acht 256-Werte-MDCT's umgeschaltet werden, um eine bessere zeitliche Auflösung zu erhalten. Eine wesentliche Rolle spielt die Möglichkeit, auf die bei der linearen Transformation entstehenden Transformationskoeffizienten Verfahren der Datenkompression anzuwenden. So wird eine lineare Prädiktion zwischen den Transformationskoeffizienten eines jeden Audioblocks vorgenommen, nur der Prädiktionsfehler wird quantisiert und übertragen. Es wird also eine lineare Prädiktion in der Spektralebene vorgenommen. Eine gute Vorhersage ergibt sich vor allem dann, wenn der Zeitverlauf der Audioabtastwerte des Blocks ausgeprägte Maxima hat, wie sie z.B. beim Anschlagen einer Triangel auftreten. Transiente Signalverläufe würden ohne Prädiktion große Quantisierungsfehler liefern. Da solche Signalverläufe andererseits in der Spektralebene gut vorhersagbar sind, treten dabei mit Prädiktion nur kleine Quantisierungsfehler auf. Im Decodierer wird der übertragene Prädiktionsfehler durch ein inverses Prädiktionsfehlerfilter wieder in Transformationskoeffizienten umgesetzt. Gleichzeitig erfahren die Quantisierungsfehler durch das inverse Prädiktionsfehlerfilter eine Färbung, die dazu führt, daß der *zeitliche* Quantisierungsfehler an den Signalverlauf des Audiosignals angepaßt werden kann. Im MPEG-Standard wird dieses Verfahren *Temporal Noise Shaping Tool* genannt. Abschließend werden die Transformationskoeffizienten im Decodierer über eine inverse MDCT in einen Audioblock rücktransformiert.

Insbesondere für stationäre Anteile in Audiosignalen verspricht eine zusätzliche *zeitliche* Prädiktion zwischen Transformationskoeffizienten in aufeinander folgenden Audioblöcken eine verbesserte Datenkompression (siehe wiederum Bild 7). Es werden dafür rückwärtsgesteuerte adaptive Prädiktoren verwendet, um keine Verzögerungen entstehen zu lassen, d.h. zur Prädiktion werden ausschließlich vorhergehende, bereits übertragene und damit dem Decodierer auch zugängliche Werte benutzt.

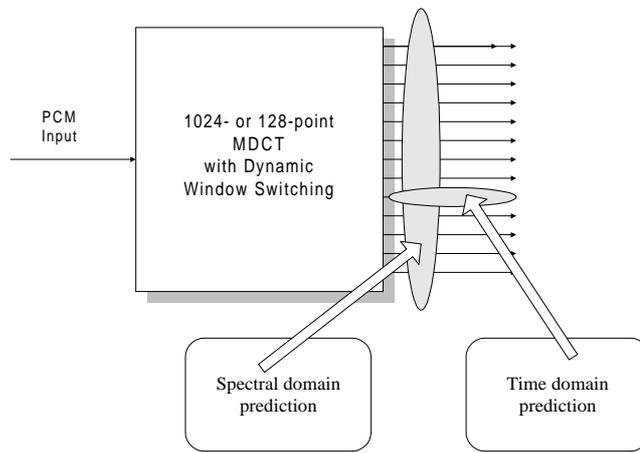


Bild 7: Spektrale und zeitliche Prädiktion von Transformationskoeffizienten in MPEG-2 AAC-Verfahren

Die erreichbare subjektive Qualität der AAC-Codierer geht aus Hörversuchen hervor, in denen Versuchspersonen eine Benotung zwischen 5 (nicht unterscheidbar) bis 1 (sehr störend) vorgenommen haben [12]. Diese in aufwendigen subjektiven Tests erzielten Benotungen wurden mit denen der als Referenz herangezogenen Compact Disc-Qualität verglichen, die Differenzwerte sind im folgenden Bild als Diffscores für verschiedene MPEG - Codierer graphisch dargestellt worden. Beispielsweise bietet der verwendete AAC - Codierer bei einer Stereobitrate von 128 kbit/s (AAC Main 128) eine der CD nahekommende Qualität, die Differenz (Diffscore) liegt (statistisch) bei -0,18. Die gezeigten Abweichungen geben das 95%-Vertrauensintervall wieder. Bei der gleichen Bitrate liegt der MP3 - Codierer (MP3 128) fast um einen ganzen Wert unter der Qualität des Referenzsignals. Weiterhin hat der verwendete AAC-Codierer AAC Main 96 bei seiner Bitrate von 96 kbit/s eine bessere Qualität als der im digitalen Hörfunk verwendete Codierer MPEG-1 Layer 2 (MP2 196). Weitere Vergleiche können in dem Bild mit low complexity - bzw. skalierbaren AAC-Codierern (LC bzw. SSR) gezogen werden.

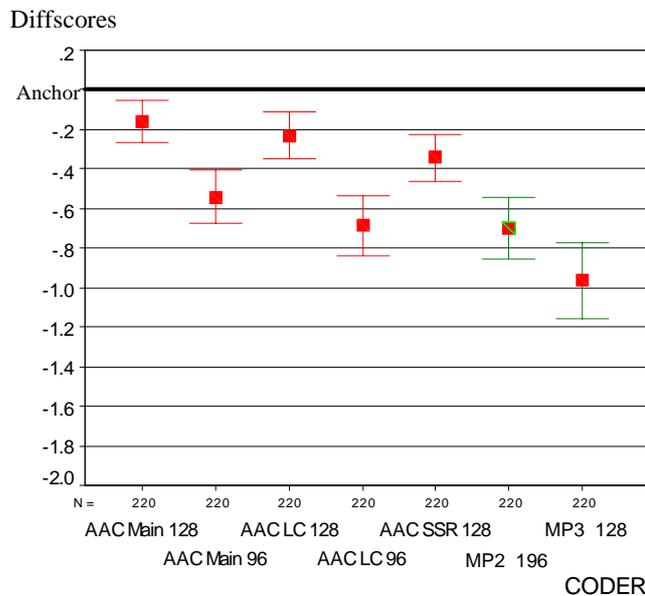


Bild 8: Abweichung der subjektiven Qualität ausgewählter Audiocodierer von der Referenzqualität der CD [12]

7 MPEG-4 Audio- und Sprachcodierung [13 - 15]

Der MPEG-4 Standard zielt auf Multimedia-Anwendungen. Er interpretiert eine audiovisuelle Szene als eine Kombination von Objekten, die auf unterschiedliche Weise repräsentiert werden können. Beispielsweise sind im folgenden Bild 9 das Zuggeräusch, die Unterhaltung innerhalb der Personengruppe, die Lautsprecherdurchsage und die Hintergrundmusik getrennte Audio-Objekte, die unterschiedlich codiert werden können. So könnte die Hintergrundmusik mit acht Kanälen wiedergegeben werden, bei geringer Kanalkapazität oder einfachem Empfänger auch mit einem Kanal, die Unterhaltung könnte dreikanalig und für verschiedene Sprachen zur Verfügung stehen, etc.

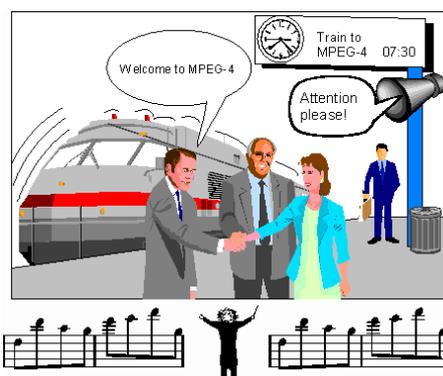


Bild 9: Audiovisuelle Szene [15]

MPEG-4 Audio bietet derzeit Datenkompressionen in unterschiedlichen Formaten von Audio- und Sprachsignalen im Bereich von 2 - 64 kbit/s an, dazu gehören parametrische Codierer für eine Vocoder-ähnliche Codierung, CELP-Codierer, die eine Vektorquantisierung von Prädiktionsfehler-signalen vornehmen, und AAC-Codierer. Daneben unterstützt MPEG-4 auch automatische Umsetzungen von Text nach Sprache (*text-to-speech tools*) und zu *talking heads* (Gesichtsanimationen). Sehr niedrige Bitraten ergeben sich bei der Verwendung von Beschreibungssprachen für die Synthese von Musik (*structured audio*). Schließlich können Sender und Empfänger auch die Verwendung anderer Algorithmen vereinbaren. Die folgende Grafik erläutert den Bereich der für eine Codierung zur Verfügung stehenden Bitraten und dafür geeignete Codieralgorithmen.

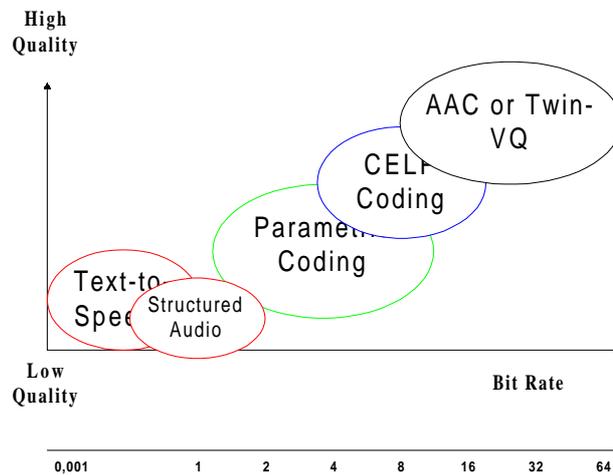


Bild 10: Qualitative Darstellung der mit MPEG-4 Audio angebotenen subjektiven Qualitäten und Bitraten.

Zur Zeit werden subjektive Tests für drei Anwendungsfelder durchgeführt: (a) Übertragung von Sprachsignalen mit Bitraten unterhalb von 10 kbit/s, (b) Übertragung von Audiosignalen über das Internet, (c) digitaler Hörfunk über AM-Mittel- und Langwellenkanäle. Ergebnisse liegen noch nicht vor.

8 Zusammenfassung und Ausblick

Der klassische Hörfunk (DAB) wird durch digitale Übertragungsverfahren ergänzt und langfristig ersetzt werden. DAB bietet höhere Wiedergabequalitäten als der analoge FM-Rundfunk (UKW), insbesondere durch die Möglichkeiten, Auftretende Störungen durch algebraische Fehlerkorrekturverfahren zu beseitigen oder zu verdecken, so daß sie nicht oder kaum wahrnehmbar sind. In diesem Beitrag wurden die internationalen MPEG-Standards (MPEG = moving picture experts group) zur Datenkompression vorgestellt. Sie lassen eine Datenkompression um den Faktor 6 bis 12 zu, ohne daß gegenüber der als Referenz dienenden CD-Qualität wahrnehmbare Unterschiede entstehen. Das Beispiel des DAB-Hörfunks hat gezeigt, daß mit der MPEG-1 - Codierung eine hochwertige und fehlerresistente Digitalisierung stereophoner Audiosignale möglich ist. Zusätzliche Kompressionsgewinne lassen sich durch den neuen MPEG - AAC - Standard erzielen, der daher in zukünftigen Anwendungen eine große Rolle spielen wird. Da AAC-Bitströme mit Bitraten bis zu 64 kbit/s über ISDN-Kanäle übertragen werden können, ist eine Echtzeitübertragung von Audiosignalen aus Tonarchiven zum Teilnehmer ohne Änderung der Infrastruktur möglich. Schnelle Überspielungen sind damit auch über das Internet/Web realisierbar. Desweiteren öffnet MPEG-4 die Tür zu komplexen Multimedia-Anwendungen, insbesondere werden unterschiedliche Formen audiovisueller Kommunikation unterstützt.

Der herkömmliche (Verteil-) Hör- und Fernsehfunke werden um interaktive Komponenten ergänzt, die es Teilnehmern schließlich ermöglichen, Programme selbst zusammenzustellen (personalized broadcast).

Literatur

- [1] P. Noll, "Digital Audio Coding for Visual Communications," Proc. of the IEEE, vol. 83, No. 6, June 1995.
 [2] E. Zwicker and R. Feldtkeller, Das Ohr als Nachrichtenempfänger. Stuttgart: S. Hirzel Verlag, 1967.

Autor(en)	Titel	Zeichen	Dateiname	Datum	Seite
Peter Noll, Institut für Nachrichtentechnik und Theoretische Elektrotechnik der TU Berlin	Audiocodierung: Vom Hörfunkstandard zum Advanced Audio Coding	29279	it+ti - Aufsatz_neu	02.07.1999	13 (14)

- [3] N. S. Jayant, J.D. Johnston, and R. Safranek, "Signal compression based on models of human perception, ", *Proc. of the IEEE*, vol. 81, No. 10, pp. 1385-1422, 1993.
- [4] R. Zelinski and P. Noll, "Adaptive Transform Coding of Speech Signals," *IEEE Trans. on Acoustics, Speech and Signal Proc.*, Vol. ASSP-25, pp. 299-309, August 1977.
- [5] ISO/IEC JTC1/SC29, "Information Technology - Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5 Mbit/s - IS 11172 (Part 3, Audio)", 1992.
- [6] K. Brandenburg and G. Stoll, "The ISO/MPEG-Audio Codec: A Generic Standard for Coding of High Quality Digital Audio," *Journal of the Audio Engineering Society (AES)*, Vol. 42, No. 10, pp. 780 - 792, Oct. 1994.
- [7] P. Noll, "MPEG Audio Coding Standards", *IEEE Signal Processing Magazine*, Sept. 1997.
- [8] ISO/IEC JTC1/SC29, "Information Technology - Generic Coding of Moving Pictures and Associated Audio Information - IS 13818 (Part 3, Audio)" , 1994.
- [9] C. Weck "The Error protection of DAB," AES Conference "DAB - The Future of Radio", London, 1995
- [10] ISO/IEC JTC1/SC29, "Information Technology - Generic Coding of Moving Pictures and Associated Audio Information - IS 13818 (Part 7, AAC)" , 1997.
- [11] M. Bosi et al, "ISO/IEC MPEG-2 advanced audio coding," *J. Audio Eng. Soc.*, Vol 45, No. 10, S. 789 - 814, 1997.
- [12] ISO/IEC/JTC1/SC29, "Report on the formal subjective listening tests of MPEG-2 NBC multichannel audio coding," Document N1371, Oct. 1996.
- [13] ISO/IEC JTC1/SC29, "Information Technology - Generic Coding of Moving Pictures and Associated Audio Information - IS 14496 (Part 3, Audio)" , 1999.
- [14] ISO/IEC/JTC1/SC29, Overview of the MPEG-4 Version 1 Standard, MPEG 97/N 19009.
- [15] ISO/IEC/JTC1/SC29, MPEG-4 Applications Document, MPEG 97/2304.