

Universität Hannover

**Institut für
Theoretische Nachrichtentechnik
und Informationsverarbeitung**

Heiko Purnhagen

Studienarbeit

**Implementation und Untersuchung eines Verfahrens
zur Reduktion der Stereo-Redundanz
bei der Codierung von Audiosignalen**

Hannover, Dezember 1992

Studienarbeit für

Herrn cand. el. Heiko Purnhagen

"Implementation und Untersuchung eines Verfahrens zur Reduktion der Stereo-Redundanz bei der Codierung von Audiosignalen"

Bei der Übertragung und Speicherung von digitalen Audiosignalen ist eines der Kernprobleme die effiziente Nutzung der begrenzt verfügbaren Kanal- bzw. Speicherkapazität. Aus diesem Grund werden Verfahren der Quellencodierung zur Reduktion der zu übertragenden bzw. zu speichernden Datenrate eingesetzt.

Innerhalb der letzten Jahre wurde im Rahmen der ISO ein Audio-Codierungsstandard erarbeitet, der ein breites Spektrum von Anwendungen ermöglicht und den heutigen Stand der Technik darstellt. Eine der Compact Disc (CD) vergleichbare Audioqualität wurde mit einer von 2×706 kbit/s auf 2×128 kbit/s reduzierten Datenrate erreicht. Bei der Codierung mit einer Datenrate von 2×64 kbit/s traten bei einigen Testsignalen noch hörbare Quantisierungsfehler auf, so daß weitere Verbesserungen erforderlich sind.

Ein Verbesserungsansatz liegt in der Optimierung des sog. "Joint Stereo Coding" Modus, bei dem die beiden Kanäle eines Stereosignals gemeinsam codiert werden. Gegenüber der unabhängigen Codierung der Kanäle läßt sich die zur Codierung erforderliche Datenrate durch Ausnutzung der Stereo-Redundanz und der Stereo-Irrelevanz weiter reduzieren. Bisherige Verfahren zur Reduktion der Stereo-Redundanz nutzen lediglich die Abhängigkeiten zwischen zeitgleichen Abtastwerten der beiden Kanäle eines Stereosignals. In dieser Arbeit soll ein als Interkanal-Prädiktion bezeichnetes Verfahren untersucht werden, bei dem die zwischen den Signalen im linken und rechten Kanal bestehenden Kreuzkorrelationen unter Berücksichtigung von Laufzeitunterschieden ausgenutzt werden.

Herr Purnhagen erhält die Aufgabe, die Interkanal-Prädiktion zu implementieren und anhand von verschiedenen Testsignalen hinsichtlich der erzielbaren Prädiktionsgewinne in Abhängigkeit von den freien Parametern des Prädiktors zu untersuchen. Der als C-Programm vorliegende ISO MPEG Layer II Codec soll um die Interkanal-Prädiktion erweitert werden. Insbesondere soll dabei auf eine effiziente Codierung der erforderlichen Seiteninformation geachtet werden. Mit Hilfe des modifizierten Codec soll untersucht werden, inwieweit objektive und subjektive Qualitätsverbesserungen erreicht werden können. Notwendige Simulationsprogramme sollen in der Programmiersprache "C" erstellt werden.

Die eingereichten Exemplare der Arbeit bleiben Eigentum des Instituts.



Ich versichere, diese Arbeit selbständig durchgeführt und alle benutzten Unterlagen angegeben zu haben.

Hannover, Dezember 1992

0 Inhaltsverzeichnis

0	Inhaltsverzeichnis	1
1	Einleitung	3
2	Grundlagen der digitalen Audiocodierung	6
2.1	Prinzipielle Funktionsweise des ISO Audio-Codecs	9
2.2	Der ISO Audio-Codec für Layer II	13
2.2.1	Struktur des Bitstroms bei Layer II	13
2.2.2	Struktur des Audio-Coders für Layer II	15
2.2.3	Struktur des Audio-Decoders für Layer II	19
3	Interkanal-Prädiktion	21
3.1	Prinzipielle Funktionsweise von Prädiktoren	21
3.2	Grundlagen der Interkanal-Prädiktion	22
3.3	Quantisierungseffekte bei der Interkanal-Prädiktion	29
3.3.1	Quantisierung der Prädiktorkoeffizienten	30
3.3.2	Prädiktion bei quantisierten Signalen	34
3.4	Ermittlung des erzielbaren Prädiktionsgewinns	40
3.4.1	Prädiktionsgewinn bei theoretischen Signalmodellen	41
3.4.2	Prädiktionsgewinn bei realen Tonsignalen	46
3.4.3	Verfahren zur Codierung der Prädiktorparameter	51
3.4.4	Prädiktionsgewinn bei Berücksichtigung der Seiteninformation	53
4	Modifizierter Audio-Codec mit Interkanal-Prädiktion	56
4.1	Struktur des modifizierten Audio-Codecs	56
4.1.1	Struktur des erweiterten Bitstroms	58
4.1.2	Struktur des modifizierten Audio-Coders	60
4.1.3	Struktur des modifizierten Audio-Decoders	67
4.2	Untersuchung der codierten Tonsignale	68
4.2.1	Objektive Qualität der codierten Tonsignale	68
4.2.2	Subjektive Qualität der codierten Tonsignale	73

5	Ausblick auf mögliche Weiterentwicklungen	78
5.1	Optimierungsmöglichkeiten beim ISO Audio-Codec	78
5.2	Optimierung der Codierung von Stereosignalen	82
5.3	Interkanal-Prädiktion bei Mehrkanalsignalen	84
6	Zusammenfassung	86
7	Literaturverzeichnis	89
A	Anhang	A 1

1 Einleitung

Zusammen mit dem Erfolg der CD (Compact Disc) hat die digitale Übertragung, Speicherung und Verarbeitung von Audiosignalen in den vergangenen Jahren erheblich an Bedeutung gewonnen. Seit einiger Zeit wird intensiv an Verfahren zur Quellencodierung gearbeitet, die bei möglichst hoher subjektiver Tonqualität eine Reduzierung der für die Übertragung benötigten Datenrate ermöglichen. Durch diese Quellencodierung werden neue Anwendungen der digitalen Audiotechnik wie DCC (Digital Compact Cassette) und DAB (Digital Audio Broadcasting) möglich. Die entwickelten Codierungsverfahren werden zur Zeit durch die ISO standardisiert /1/ /5/ /9/.

Der ISO Audio-Codec nutzt zur Verringerung der Datenrate hauptsächlich psychoakustische Eigenschaften des menschlichen Gehörs aus. Dabei wird vorwiegend von Verdeckungseffekten (Maskierungseffekten) Gebrauch gemacht. Diese Effekte bezeichnen die Eigenschaft des Gehörs, daß ein lautes Geräusch in einem Frequenzbereich dazu führt, daß Geräusche in benachbarten Frequenzbereichen nicht mehr wahrgenommen werden, wenn sie leiser als eine entsprechende Mithörschwelle (Maskierungsschwelle) sind.

Um die Maskierungseffekte bei der Quellencodierung zur Irrelevanzreduktion auszunutzen, wird das digitale Tonsignal zunächst spektral in mehrere Frequenzbereiche, sogenannte Teilbänder, zerlegt und in kurze Zeitabschnitte (Blöcke) unterteilt. Entsprechend dem Abtasttheorem wird die Abtastrate Teilbandsignale reduziert. Mit Hilfe eines psychoakustischen Modells werden die durch das zu codierende Tonsignal selbst hervorgerufenen Maskierungsschwellen für die Teilbänder berechnet. Die einzelnen Teilbandsignale werden nun so quantisiert, daß das dabei entstehende Quantisierungsgeräusch möglichst gerade unterhalb der Maskierungsschwelle liegt und damit unhörbar bleibt. Die Maskierungsschwellen werden für jeden Zeitabschnitt neu berechnet, so daß auf diese Weise auch Vor- und Nachverdeckungseffekte in der Größenordnung einer Blocklänge ausgenutzt werden.

Der Entwurf des ISO-Standards beschreibt einen Audio-Codec mit drei möglichen Ausbaustufen (Layer I, II und III), die sich in Komplexität und erzielbarer Datenratenreduktion unterscheiden, aber in ihrer Struktur aufeinander aufbauen. Der Layer I ermöglicht bei relativ niedriger Komplexität des Codecs eine Reduktion der

Datenrate etwa um den Faktor 4. Die Codecs für Layer II und III bieten jeweils eine höhere Datenratenreduktion und sind entsprechend komplexer.

Für Codierung eines Stereosignals sind zwei verschiedene Modi vorgesehen. Im Modus "Independent Stereo Coding" werden die beiden Kanäle eines Stereosignals vom ISO Audio-Coder praktisch unabhängig voneinander codiert. Im Modus "Joint Stereo Coding" werden dagegen beide Stereokanäle gemeinsam codiert. Dabei kommen relativ einfache Verfahren zur Ausnutzung der Stereo-Redundanz und der Stereo-Irrelevanz zum Einsatz, die häufig eine gewisse Qualitätsverbesserung ermöglichen /5/.

In der Regel wird mit solchen einfachen Verfahren, die die Abhängigkeiten zwischen zeitgleichen Abtastwerten der beiden Kanäle eines Stereosignals ausnutzen, nur eine geringe Reduktion der Stereo-Redundanz erzielt /3/ /13/.

Ziel dieser Arbeit ist es daher, die in einem Tonsignal enthaltene Stereo-Redundanz genauer zu untersuchen und anschließend bei der Codierung auszunutzen. Dabei soll ein Prädiktor verwendet werden, der von den Abtastwerten in einem Kanal des Stereosignals ausgehend möglichst gute Schätzwerte für die Abtastwerte im zweiten Kanal berechnet. Ein solcher Prädiktor wird hier als Interkanal-Prädiktor bezeichnet. Bei der anschließenden Codierung braucht dann anstelle des zweiten Stereokanals nur ein Prädiktionsfehlersignal übertragen zu werden. Dieses Fehlersignal hat eine um den Prädiktionsgewinn kleinere Amplitude als das Signal des zweiten Stereokanal und kann daher entsprechend gröber quantisiert werden. Es ist jedoch zu berücksichtigen, daß die verwendeten Prädiktorparameter als Seiteninformation zu übertragen sind.

Im Rahmen dieser Arbeit wird zunächst ein Verfahren zur Bestimmung der optimalen Parameter von Interkanal-Prädiktoren hergeleitet. Es werden dann die damit erzielbaren Prädiktionsgewinne in Abhängigkeit von den freien Parametern des Prädiktors anhand von verschiedenen theoretischen Signalmodellen und unterschiedlichen realen Tonsignalen untersucht. Dabei wird von der Teilbandzerlegung und der zeitlichen Blockstruktur des ISO Audio-Codecs für Layer II ausgegangen, so daß für jedes Teilband und jeden Block ein eigener Interkanal-Prädiktor eingesetzt wird. Es werden auch die Auswirkungen der Quantisierung der Prädiktorparameter und Teilbandsignale auf den Prädiktionsgewinn betrachtet.

Auf diese Untersuchungen aufbauend wird ein als C-Programm vorliegender Audio-Codec für Layer II um die Interkanal-Prädiktion erweitert. Dazu muß ein geeignetes Verfahren zur Codierung der Prädiktorparameter entwickelt werden. Es soll eine Anpassung der Komplexität der Prädiktoren an die im Tonsignal enthaltene Stereo-Redundanz ermöglichen und so auch eine Optimierung der erforderlichen Seiteninformation erlauben. Außerdem muß der erzielte Prädiktionsgewinn in geeigneter Weise bei der Quantisierung der Teilbandsignale berücksichtigt werden. Mit Hilfe des so modifizierten Codecs wird dann untersucht, inwieweit eine Verbesserung der objektiven und subjektiven Tonqualität der codierten Stereosignale möglich ist.

Der vorliegende Bericht ist in folgender Weise gegliedert: Im Kapitel 2 wird die prinzipielle Funktionsweise des ISO Audio-Codecs beschrieben, von dem bei den Untersuchungen ausgegangen wurde. In Kapitel 3 werden dann die Grundlagen der Interkanal-Prädiktion erläutert und die damit möglichen Prädiktionsgewinne betrachtet. Anschließend wird in Kapitel 4 der modifizierte Audio-Codec mit Interkanal-Prädiktion beschrieben und es wird die damit erzielte objektive und subjektive Tonqualität aufgezeigt. In Kapitel 5 wird ein Ausblick auf mögliche Weiterentwicklungen und andere Anwendungen der Interkanal-Prädiktion gegeben. In Kapitel 6 werden die Ergebnisse der Arbeit dann abschließend zusammengefaßt.

2 Grundlagen der digitalen Audiocodierung

Die in den letzten Jahren rasant fortschreitende Entwicklung der Halbleitertechnologie ermöglicht die kostengünstige Herstellung von Chips mit einer immer höheren Integrationsdichte. Dadurch wird es möglich, komplexe Algorithmen der digitalen Signalverarbeitung auch bei Echtzeitanwendungen einzusetzen. In der Nachrichtentechnik ist so der Einsatz neuer leistungsfähiger Verfahren zur Quellen- und Kanalcodierung möglich geworden. Beispiele dafür sind das D-Netz für Mobiltelefone oder die Bildtelefonie über einen 64 kbit/s ISDN-Kanal. Ein weiteres Einsatzgebiet ist die digitale Tontechnik.

Um ein Audiosignal digital in Studioqualität zu übertragen, wird eine Abtastrate von 48 kHz und eine gleichförmige Quantisierung mit 16 bit pro Abtastwert verwendet. Diese Codierung wird als PCM (Pulse Code Modulation) bezeichnet. Ein Stereosignal besitzt somit eine Datenrate von $2 \cdot 768 \text{ kbit/s} = 1,536 \text{ Mbit/s}$. Bei der CD ergibt sich aufgrund der Abtastrate von 44,1 kHz ein Wert von $2 \cdot 705,6 \text{ kbit/s} = 1,4112 \text{ Mbit/s}$. Für die Speicherung von digitalen Audiosignalen bei diesen Datenraten sind spezielle Medien wie die CD oder DAT (Digital Audio Tape) entwickelt worden, die inzwischen weit verbreitet sind.

Seit einigen Jahren ist die MPEG (Moving Pictures Expert Group) der ISO damit beschäftigt, Verfahren zu entwickeln und zu standardisieren, die es ermöglichen sollen, auf den oben genannten Speichermedien Bild- und Tonsignale gleichzeitig zu speichern [9] [1]. Dazu wird die zur Verfügung stehende Datenrate von etwa 1,4 Mbit/s aufgeteilt in 1,15 Mbit/s für das Bild und 256 kbit/s für den Ton. Bei der Toncodierung ist also eine Reduzierung der Bitrate um den Faktor 5,5 notwendig. Trotzdem soll sich die subjektive Tonqualität gegenüber der CD nicht verschlechtern.

Der Einsatz solcher Toncodierungsverfahren ermöglicht aber auch neue Anwendungen der digitalen Tonübertragung und -speicherung. Dazu gehören z.B. DCC (Digital Compact Cassette), die aufgrund des verwendeten feststehenden Kopfes nur eine Datenrate von 384 kbit/s speichern kann, sowie DAB (Digital Audio Broadcasting), eine terrestrische Rundfunkübertragung, die Datenraten von etwa 256 kbit/s für ein Stereosignal zur Verfügung stellt. Eine Reduzierung der Datenrate auf 128 kbit/s ermöglicht die Übertragung eines Stereosignals über einen ISDN-Basisanschluß (2 B-Kanäle).

Bei der direkten Übertragung digitalisierter Tonsignale mit $2 \cdot 768 \text{ kbit/s}$ wird keine Redundanzreduktion eingesetzt. Zur Irrelevanzreduktion dient nur die (unumgängliche) Abtastung und Quantisierung des analogen Audiosignals. Dabei wird davon ausgegangen, daß für die Audioübertragung eine Bandbreite von 20 kHz bei einem Signal/Rausch-Abstand (SNR, Signal Noise Ratio) von rund 95 dB benötigt wird. Aufgrund der sehr vielfältigen und komplexen Struktur von Tonsignalen ist im Gegensatz zu Sprach- oder Bildsignalen eine Redundanzreduktion relativ schwierig.

Da das menschliche Gehör aber die bei 20 kHz Bandbreite und 95 dB SNR übertragbare Information nie vollständig wahrnehmen kann, ist eine deutliche Irrelevanzreduktion möglich. Hierbei werden hauptsächlich Verdeckungseffekte des Gehörs ausgenutzt [15]. Da diese Effekte im Englischen mit "masking" bezeichnet werden, wird auch von Maskierungseffekten gesprochen.

Die Verdeckungseffekte beschreiben die Eigenschaft des Gehörs, daß ein lautes Geräusch in einem Frequenzbereich dazu führt, daß ein anderes Geräusch in einem benachbarten Frequenzbereich nicht wahrgenommen wird, wenn es leiser als eine bestimmte Maskierungsschwelle (Mithörschwelle) ist. Diese Maskierungsschwellen sind frequenz- und pegelabhängig. Außerdem hängen sie von der Art des verdeckenden und des verdeckten Geräusches ab. Für die Verdeckung eines Tones durch ein Schmalbandrauschen sind einige exemplarische Mithörschwellen in den Bildern 2.1 und 2.2 dargestellt. Die gestrichelte Linie stellt dabei die Ruheshörschwelle dar.

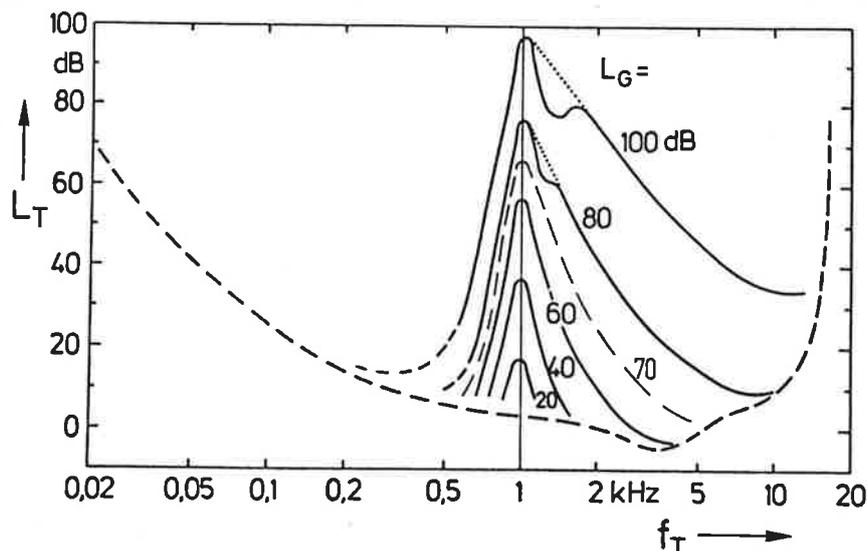


Bild 2.1 Mithörschwelle L_T für Töne verdeckt durch Schmalbandrauschen der Mittenfrequenz $f_m = 1 \text{ kHz}$ mit verschiedenen Pegeln L_G [15]

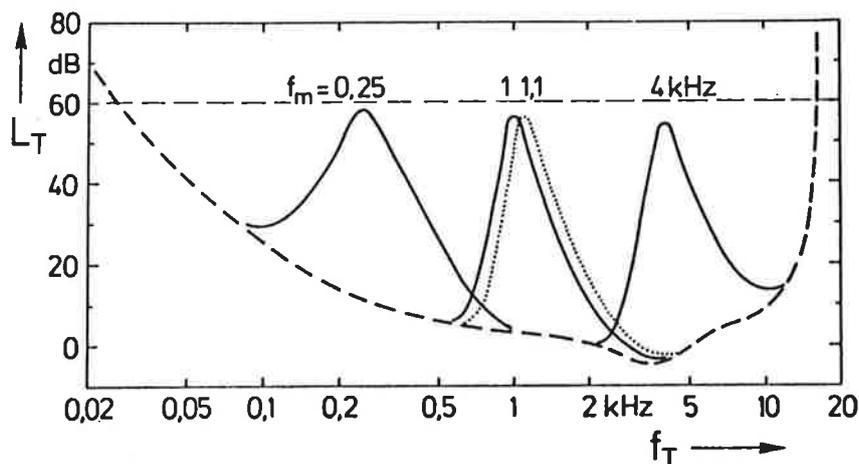


Bild 2.2 Mithörschwelle L_T für Töne verdeckt durch Schmalbandrauschen des Pegels $L_G = 60$ dB mit verschiedenen Mittenfrequenzen f_m [15]

Wenn anstelle des Schmalbandrauschens ein Ton als Maskierer verwendet wird, ergeben sich relativ ähnliche Kurven. Dabei muß jedoch auch die Wahrnehmung von Schwebungseffekten berücksichtigt werden. Auch bei der Verdeckung eines Schmalbandrauschens durch ein Schmalbandrauschen oder Ton als Maskierer ergeben sich prinzipiell ähnliche Mithörschwellen.

Die Verdeckungseffekte treten auch im Zeitbereich auf. Neben der oben beschriebenen Simultanverdeckung gibt es noch die Vorverdeckung, die schon etwa 10 ms vor Einsetzen des Maskierers wirksam ist, sowie die Nachverdeckung, die noch etwa 50 ms nach Ende des Maskierers andauert. Die Stärke des Verdeckungseffektes hängt vom zeitlichen Abstand zum Anfang beziehungsweise Ende des Maskierers ab.

Die Ausnutzung dieser Verdeckungseigenschaften ermöglichen nun eine wesentlich effizientere Quellencodierung von Tonsignalen. Dabei wird zunächst die durch das zu codierende Signal selbst hervorgerufene Mithörschwelle als Funktion der Frequenz berechnet. Hierzu ist eine spektrale Analyse des Signals sowie die geeignete Überlagerung der durch die verschiedenen ton- und rauschartigen Signalanteile hervorgerufenen Maskierungsschwellen notwendig.

Anschließend müssen die verschiedenen Frequenzbereiche des Audiosignals individuell codiert werden. Die dabei notwendige spektrale Zerlegung kann entweder durch eine Filterbank mit anschließender Abtastratenreduktion oder mit einer Transformation wie z.B. einer FFT (Fast Fourier Transform) erreicht werden. Die

spektral zerlegten Signale werden dann so quantisiert, daß das entstehende Quantisierungsgeräusch unterhalb der für den jeweiligen Frequenzbereich ermittelten Maskierungsschwelle liegt. Ist das zu codierende Signal selbst leiser als die Maskierungsschwelle, braucht es gar nicht übertragen zu werden. Die zeitlichen Verdeckungseffekte werden dadurch ausgenutzt, daß die Codierung blockweise erfolgt. Die verwendeten Blocklängen liegen in der Größenordnung von 10 ms bis 20 ms.

Um das so codierte Audiosignal wieder entschlüsseln zu können, ist neben der Übertragung der spektral zerlegten und individuell quantisierten Signalanteile auch die Übertragung von Seiteninformation notwendig. Sie beschreibt hauptsächlich die für die verschiedenen Spektralbereiche benutzten Quantisierer. Für die quantisierten Signalanteile und die Seiteninformation ist teilweise auch eine Entropiecodierung zur Redundanzreduktion sinnvoll.

Ausgehend von den hier beschriebenen Grundlagen wurde eine Reihe verschiedener Audio-Codecs entwickelt. Dazu gehören unter anderem PASC (Precision Adaptive Subband Coding) für die DCC, ATRAC (Adaptive Transform Acoustic Coding) für die MD (Music Disc), MUSICAM (Masking Pattern Universal Subband Integrated Coding And Multiplexing) für DAB sowie ASPEC (Adaptive Spectral Perceptual Entropy Coding).

Im Rahmen der oben erwähnten ISO-Standardisierung wurde nun ein vielfältig einsetzbarer Audio-Codec entwickelt, der PASC und MUSICAM mit einigen Eigenschaften von ASPEC kombiniert.

2.1 Prinzipielle Funktionsweise des ISO Audio-Codecs

Der Entwurf des ISO-Standard für die Toncodierung /1/ beschreibt einen Audio-Codec mit drei möglichen Ausbaustufen (Layer I, II und III). Die einzelnen Stufen bauen von Layer I ausgehend aufeinander auf und ermöglichen durch die zunehmende Komplexität des Codecs jeweils eine höhere Datenratenreduktion als die darunter liegende Stufe.

Layer I entspricht etwa PASC und erlaubt bei einem Stereosignal eine Datenratenreduktion auf 384 kbit/s ohne wahrnehmbare Qualitätsverschlechterung. Layer II entspricht MUSICAM und ermöglicht eine Codierung mit 256 kbit/s. Layer III

enthält zusätzlich Elemente von ASPEC und bietet bei 128 kbit/s noch eine hohe subjektive Tonqualität.

Die drei Layer sind untereinander in der Weise kompatibel, daß der Decoder einer Ausbaustufe auch den vom Coder einer niedrigeren Ausbaustufe erzeugten Bitstrom entschlüsseln können muß.

Die Codecs erlauben jeweils die Übertragung von Mono- oder Stereosignalen. Diese müssen in digitaler Form mit einer Abtastrate von 48 kHz, 44,1 kHz oder 32 kHz und gleichförmiger Quantisierung mit 16 bit vorliegen. Die Datenraten des codierten Bitstroms können zwischen 32 kbit/s bei Monosignalen und 384 kbit/s bei Stereosignalen liegen.

In den Bildern 2.3 und 2.4 ist die grundlegende Struktur des Audio-Coders und des Audio-Decoders dargestellt. Anschließend werden die einzelnen Layer genauer beschrieben.

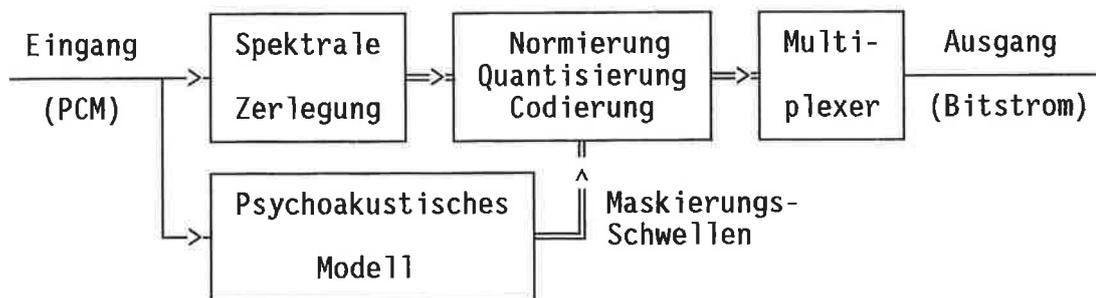


Bild 2.3 Blockdiagramm des Audio-Coders

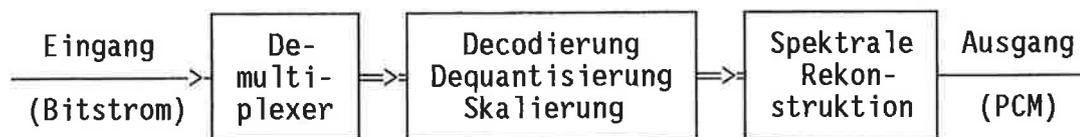


Bild 2.4 Blockdiagramm des Audio-Decoders

Layer I: Für die spektrale Zerlegung wird bei Layer I eine Polyphasenfilterbank verwendet, die das Tonsignal in 32 gleichbreite Teilbänder aufteilt. Die Abtastrate der einzelnen Teilbandsignale wird dann entsprechend um den Faktor 32 reduziert. Jeweils 12 der auf diese Weise für ein Teilband berechneten Teilbandsample wer-

den nun zu einem Block zusammengefaßt. Für eine Abtastrate von 48 kHz ergibt sich so eine Teilbandbreite von 750 Hz und eine Blocklänge von 8 ms. Die Sample eines Blocks werden nun normiert, wobei der für die Rekonstruktion notwendige Skalierungsfaktor als Seiteninformation übertragen wird.

Anschließend wird von einem psychoakustischen Modell das zu codierende Signal spektral analysiert und es werden für die einzelnen Teilbänder die Maskierungsschwellen ermittelt. Für jedes Teilband ergibt sich der geforderte Signal/Masken-Abstand (SMR, Signal Mask Ratio) als Differenz zwischen Signalpegel und Maskierungsschwelle.

Die bei der gewählten Bitrate für die Quantisierung der Teilbandsample zur Verfügung stehenden Bits werden von einem Bitzuteilungsalgorithmus auf die einzelnen Teilbänder verteilt. Abhängig von der Anzahl der zugeteilten Bits ergibt sich so für jedes Teilband ein gewisses Quantisierungsgeräusch - in der Regel ein Quantisierungsrauschen - und damit ein Signal/Rausch-Abstand (SNR, Signal Noise Ratio). Der Bitzuteilungsalgorithmus hat das Ziel, in jedem Teilband mit dem Signal/Rausch-Abstand den geforderten Signal/Masken-Abstand zu erreichen oder zu übertreffen. Die gewählte Bitzuteilung wird als Seiteninformation übertragen.

Anschließend werden die Teilbandsample entsprechend der gewählten Bitzuteilung gleichförmig quantisiert. Eine weitere Codierung findet nicht statt. Zusammen mit einem Kopf, der Informationen über den verwendete Layer und die benutzte Abtastrate und ähnliches enthält, werden die einzelnen Bestandteile der Seiteninformation und die quantisierten Teilbandsample in einem Multiplexer zu einem Bitstrom zusammengefügt.

Im Decoder werden die jeweils inversen Operationen in umgekehrter Reihenfolge ausgeführt. Zunächst wird also der Bitstrom vom Demultiplexer in seine Bestandteile zerlegt. Dann werden die Teilbandsample gemäß der Seiteninformation dequantisiert und skaliert. Nachdem die Abtastrate der Teilbandsignale um den Faktor 32 erhöht wurde, wird das Tonsignal durch eine Polyphasenfilterbank wieder rekonstruiert.

Layer II: Der Layer II ist an verschiedenen Stellen gegenüber dem Layer I verbessert und optimiert. So werden jeweils 3 Blöcke mit 12 Teilbandsample zu einem größeren Block zusammengefaßt und gegebenenfalls gemeinsam normiert. Das

psychoakustische Modell arbeitet jetzt mit einer verbesserten spektralen Auflösung. Die möglichen Bitzuteilungen hängen vom jeweiligen Teilband ab, da bei hohen Frequenzen oft nur eine grobe Quantisierung notwendig ist. Außerdem werden die quantisierten Teilbandsample zum Teil auch blockweise codiert. Somit ist auch eine Quantisierung mit zum Beispiel 5 oder 9 Stufen pro Abtastwert sinnvoll möglich.

Layer III: Bei Layer III wird zunächst auch die Polyphasenfilterbank von Layer I und II verwendet. Anschließend werden die unterabgetastetem Teilbandsignale dann mit einer MDCT (Modifizierte Diskrete Cosinus Transformation) spektral noch feiner zergliedert. Die MDCT wird auf sich überlappende Blöcke mit einer Länge von 36 oder 12 Teilbandsample angewendet. Bei einer Blocklänge von 36 ergibt sich so eine Auflösung von 576 Spektrallinien. Bei einer Abtastrate von 48 kHz entspricht dieses einer Linienbreite von 41,7 Hz.

Die Normierung erfolgt nun getrennt für verschiedene Spektralbereiche. Für dem Wahrnehmungsvermögen des Gehörs angepaßte Frequenzbereiche werden die Maskierungsschwellen von einem entsprechend optimierten psychoakustischen Modell bestimmt. Die Quantisierung der einzelnen Spektrallinien erfolgt ungleichförmig (kompanzierend). Jeweils 2 oder 4 Linien werden dann gemeinsam codiert. Dabei stehen verschiedene Tabellen für die verwendete Huffmancodierung zur Auswahl. Der Layer III ist also wesentlich komplexer als die Layer I und II, bietet aber speziell für tonale Signale eine bessere Codierung.

Layer I..III: Bei allen 3 Layern können die beiden Kanäle eines Stereosignals mit "Independent Stereo Coding" praktisch unabhängig voneinander codiert werden. Außerdem ist bei allen Layern mit "Joint Stereo Coding" auch eine gemeinsamen Codierung der beiden Kanäle eines Stereosignals möglich. Die Mithörschwellen für einen Kanal werden aber immer ohne Berücksichtigung des jeweils anderen Kanals ermittelt. Bei Layer I und II wird bei "Joint Stereo Coding" ein Intensitätsstereo-Verfahren (Intensity Stereo) eingesetzt. Dabei wird nur das Summensignal aus beiden Stereokanälen quantisiert und codiert. Die für die Skalierung notwendige Information wird jedoch weiter für beide Kanäle getrennt übertragen. Dieses Verfahren wird nur für die höherfrequenten Signalanteile (zum Beispiel für den Frequenzbereich oberhalb von 3kHz) angewendet. Bei Layer III ist es zusätzlich dazu möglich, statt des linken und rechten Kanals die Summe und die Differenz dieser Kanäle zu übertragen.

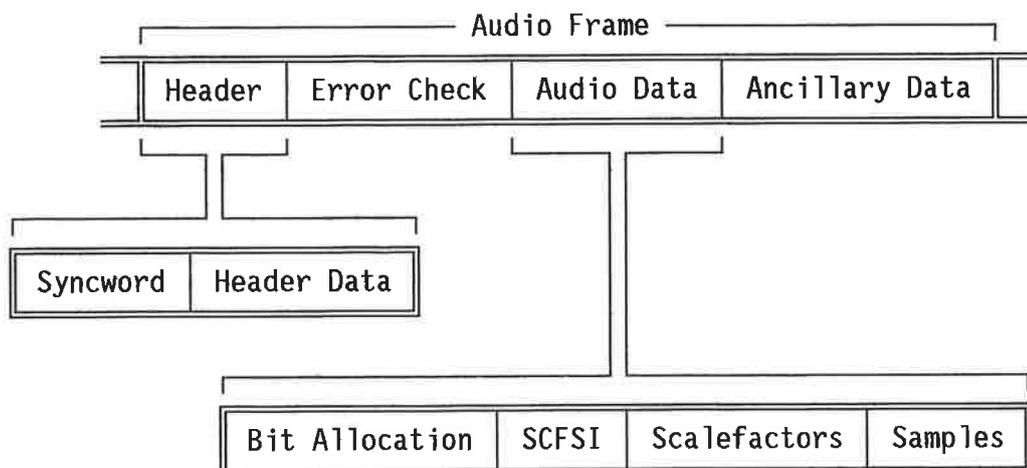
Im ISO-Standard wird nur das Format des codierten Bitstroms und der Decoder eindeutig festgelegt. Für den Coder und das verwendete psychoakustische Modell werden Vorschläge angegeben. Der Standard bietet so die Möglichkeit, durch Verbesserungen bei der Codierung die Tonqualität weiter zu optimieren.

2.2 Der ISO Audio-Codec für Layer II

Bei der Durchführung dieser Arbeit standen ein Coder und ein Decoder für die ISO-Audio Layer I und II als Software in der Programmiersprache C zur Verfügung. Da die Interkanal-Prädiktion in den Layer II integriert wurde, folgt hier eine detailliertere Beschreibung dieses Audio-Codecs. Sie ist für das Verständnis der später beschriebenen Modifikationen notwendig.

2.2.1 Struktur des Bitstroms bei Layer II

Der für Layer II festgelegte Bitstrom ist in Rahmen (Frames) gegliedert. Ein solcher Rahmen besteht aus einem Kopf (Header), der auch ein Synchronisationswort enthält, optionalen Informationen für Fehlererkennung (Error Check), den eigentlichen Audiodaten (Audio Data) sowie Raum für zusätzliche Daten (Ancillary Data). Der Aufbau eines solchen Rahmens ist Bild 2.5 zu entnehmen.



(SCFSI = Scalefactor Select Information)

Bild 2.5 Aufbau des Bitstroms bei Layer II

In einem Rahmen des Bitstroms sind jeweils 1152 Abtastwerte pro Kanal codiert. Bei Stereosignalen werden die Audiodaten für beide Kanäle gemeinsam in einem

Rahmen übertragen. Für eine Abtastfrequenz von 48 kHz ergibt sich eine Rahmenlänge von 24 ms. Bei einer Datenrate des Bitstroms von 128 kbit/s besteht ein Rahmen somit aus 3072 Bits. Bei Stereosignalen entspricht dieses einer Datenratenreduktion um den Faktor 12 - also einer Codierung mit 1,333 Bit pro Abtastwert.

Header: Der Header besteht aus 32 Bits, von denen 12 Bits für das Synchronisationswort (Syncword) benutzt werden. Die übrigen Bits geben unter anderem an, welcher Decoder zu benutzen ist (hier also: ISO Audio Layer II), welche Abtastrate zu verwenden ist und welche Bitrate der Datenstrom hat. Außerdem enthält der Kopf Informationen über die Kanalanzahl (1 oder 2) und darüber, ob und welche Teilbänder bei Stereosignalen mit Intensity Stereo gemeinsam codiert wurden. Daneben werden auch Kopierschutzinformationen übertragen. Ein weiteres Bit kennzeichnet, ob ein Teil des Bitstroms über ein Fehlerkorrekturwort (Error Check) geschützt ist. Neben einigen anderen Bits gibt es auch noch ein Bit (Private Bit), über das vom Anwender frei verfügt werden kann.

Error Check: Das optionale Fehlerkorrekturwort (Error Check) ist 16 Bit lang. Es wird mit Hilfe eines Generatorpolynoms aus den zu schützenden Bits gewonnen und ermöglicht so eine Fehlererkennung oder Fehlerkorrektur für diese Bits. Der geschützte Bereich beinhaltet den Header sowie die Bit Allocation und die Scalefactor Select Information.

Bit Allocation: Der Bereich mit den Tondaten (Audio Data) hat keine konstante Länge. Er enthält zunächst Information über die Bitzuteilung (Bit Allocation). Sie gibt an, wieviele Bits für die Quantisierung der Teilbandsample in den einzelnen Teilbändern verwendet werden. Bei Stereosignalen wird diese Information für jeden Kanal getrennt codiert. Dabei sind - je nach Teilband - 16, 8 oder 4 verschiedenen feine Quantisierer möglich, so daß für die Codierung jeweils 4, 3 oder 2 Bit benötigt werden. Die höchsten Teilbänder (z.B. > 20 kHz) sind von der Bitzuteilung ausgeschlossen. Die für die unterschiedlichen Teilbänder zugelassenen Quantisierer hängen auch von der Abtastfrequenz und der Bitrate pro Kanal ab. Eine typische Tabelle der erlaubten Bitzuteilungen ist im Anhang wiedergegeben (Anhang A.1).

Wenn aus der Bitzuteilung hervorgeht, daß einem Teilbandsignal keine Bits zugeteilt sind, wird natürlich keine weitere Information (SCFSI, Scalefactors, Samples) für dieses Teilbandsignal übertragen.

Scalefactor Select Information (SCFSI): Die SCFSI gibt für die einzelnen Teilbänder der Kanäle an, ob 1, 2 oder 3 unterschiedliche Skalierungsfaktoren übertragen werden. Sie wird mit jeweils 2 Bit codiert. Falls zwei Skalierungsfaktoren übertragen werden, gibt die SCFSI auch an, ob der erste oder der zweite der beiden übertragenen Skalierungsfaktoren zweimal angewendet werden soll. Wird nur ein Skalierungsfaktor übertragen, wird dieser dreimal verwendet.

Scalefactors: Für die einzelnen Teilbänder der Kanäle werden abhängig von der SCFSI 1 bis 3 Skalierungsfaktoren (Scalefactors) übertragen. Sie werden mit 6 Bit codiert und geben den für die Skalierung benötigten Verstärkungsfaktor mit einer Stufengröße von 2 dB an. Somit ist eine Dynamik von über 120 dB möglich.

Samples: Nun werden die quantisierten Abtastwerte (Samples) der unterabgetasteten Teilbänder der Kanäle übertragen. Pro Teilband und Kanal sind dieses 36 Abtastwerte. Sie werden je nach gewählter Bitzuteilung mit bis zu 16 Bit pro Sample codiert. Bei einigen groben Quantisierern (3, 5 oder 9 Quantisierungsstufen) werden 3 aufeinanderfolgende Sample zur Redundanzverringern in einem Codewort zusammengefaßt. Der größte Teil der in einem Frame zur Verfügung stehenden Bits wird für die so übertragenen Abtastwerte benötigt.

Werden die beiden Kanäle mit Intensity Stereo in einem Teilband gemeinsam codiert, so werden die Bit Allocation und die Samples für dieses Teilband nur einmal übertragen. Die SCFSI und die Skalierungsfaktoren werden aber weiter für beide Kanäle getrennt übertragen. Intensity Stereo wird nur für die oberen Teilbänder verwendet. Abhängig von dem im Header angegebenen Intensity Stereo Mode werden die unteren 4, 8, 12 oder 16 Teilbänder weiterhin für beide Kanäle unabhängig übertragen.

Ancillary Data: Die in einem Frame jetzt noch verbleibenden Bits stehen für weitere Daten (Ancillary Data) zur Verfügung. Diese zusätzlichen Daten sind anwenderdefinierbar und werden vom Decoder nicht ausgewertet.

2.2.2 Struktur des Audio-Coders für Layer II

Die grundlegende Struktur des Audio-Coder wurde schon in Abschnitt 2.1 erläutert. Ein Blockdiagramm des Coders ist in Bild 2.3 zu sehen. Bild 2.6 stellt den Ablauf der Codierung dar.

PCM-Audiodaten einlesen
Polyphasenfilterbank
Skalierungsfaktoren bestimmen
Psychoakustisches Modell
SCFSI bestimmen
Bitzuteilung ermitteln
Teilbandsample normieren
Teilbandsample quantisieren
Frame codieren
Bitstrom ausgeben

Bild 2.6 *Ablauf der Codierung bei Layer II*

Polyphasenfilterbank: Das Eingangssignal wird zunächst mit der Polyphasenfilterbank in 32 Teilbandsignale zerlegt, deren Abtastrate um den Faktor 32 reduziert wird.

Skalierungsfaktoren: In einem Frame sind 36 Teilbandsample pro Teilband und Kanal zu codieren. Sie werden zunächst zu 3 Blöcken mit je 12 Sample zusammengefaßt. Für jeden dieser Blöcke wird nun der kleinste mögliche Skalierungsfaktor (Scalefactor) ermittelt, der größer oder gleich dem im Block maximal auftretenden Betrag eines Teilbandsamples ist.

Psychoakustisches Modell: Für die Berechnung der Mithörschwellen kann zwischen zwei unterschiedlichen psychoakustischen Modellen gewählt werden /1/ /11/. Beide basieren hauptsächlich auf einer spektrale Analyse des zu codierenden Eingangssignals, die mit einer 1024 Punkte FFT realisiert wird. Beim ersten Modell werden außerdem die eben ermittelten Skalierungsfaktoren ausgewertet. Auf die Funktionsweise dieser Modelle soll hier nicht genauer eingegangen werden. Die psychoakustischen Modelle behandeln beide Kanäle unabhängig voneinander. Lediglich bei dem zweiten Modell wird die "Binaurale Nichtverdeckung" berücksichtigt.

Beide Modelle liefern für die einzelnen Teilbänder der Kanäle jeweils einen SMR-Wert (Signal/Mask-Ratio). Dieser Wert gibt den Abstand zwischen Signalpegel und gerade noch maskiertem Rauschpegel für das jeweilige Teilband in dB an.

SCFSI: Nun wird entschieden, ob alle 3 Skalierungsfaktoren codiert werden sollen oder ob 2 oder 1 Skalierungsfaktor ausreichen. Dazu werden die Differenzen zwischen zwei aufeinanderfolgenden Skalierungsfaktoren ausgewertet. Beträgt diese Differenz maximal zwei Stufen (also 4 dB), so wird nur der größere der beiden Skalierungsfaktoren übertragen und für beide Blöcke verwendet. Verringert sich der Skalierungsfaktor von einem Block zum nächsten um 6db oder mehr, so wird auch in diesem Fall teilweise nur der erste Skalierungsfaktor für beide Blöcke verwendet. Auf diese Weise werden Nachverdeckungseigenschaften des Gehörs ausgenutzt. Wenn diese Bedingungen auf alle 3 Skalierungsfaktoren zutreffen, wird nur der größte codiert und für alle Blöcke verwendet. Die Information, wieviele Skalierungsfaktoren codiert werden und für welche Blöcke sie gelten, wird als Scalefactor Select Information (SCFSI) bezeichnet.

Bitzuteilung: Als nächstes wird die Bitzuteilung (Bit Allocation) durchgeführt. Dabei wird versucht, die in einem Frame zur Verfügung stehenden Bits so auf die einzelnen Teilbänder der beiden Kanäle aufzuteilen, daß das bei der Quantisierung der Teilbandsample entstehende Quantisierungsgeräusch in allen Teilbändern möglichst leiser als die Maskierungsschwelle ist. Hierfür wird angenommen, daß das Quantisierungsgeräusch ein weißes Rauschen ist. Aus einer Tabelle geht hervor, welche SNR-Werte (Signal/Noise-Ratio) für die verschiedenen möglichen Quantisierer angenommen werden. Diese Tabelle ist im Anhang abgedruckt (Anhang A.2).

Für feine Quantisierer ergeben sich die SNR-Werte in dieser Tabelle zum Beispiel unter der Annahme, daß das Signal einen maximal ausgesteuerten sinusförmigen Zeitverlauf hat und daß der Quantisierungsfehler über die Breite einer Quantisiererstufe gleichverteilt ist /8/. Für grobe Quantisierer enthält die Tabelle einen bis zu $\approx 4,5$ dB kleineren SNR-Wert. Damit wird unter anderem berücksichtigt, daß bei grober Quantisierung die Annahme der Gleichverteilung des Quantisierungsfehlers nicht mehr gegeben ist. Auch kann das Quantisierungsgeräusch dann nicht mehr einfach als weißes Rauschen angesehen werden und wird daher eventuell schlechter maskiert.

Wird vom SMR-Wert, der für ein Teilband berechnet wurde, der SNR-Wert des verwendeten Quantisierers abgezogen, so ergibt sich der NMR-Wert (Noise/Mask-Ratio). Er gibt an, wie weit in diesem Teilband das Quantisierungsgeräusch über der Maskierungsschwelle liegt.

Die eigentliche Vergabe der zur Verfügung stehenden Bits erfolgt in einer Schleife. Dazu wird vor der Schleife für alle Teilbänder der beiden Kanäle der Bitzuteilungs-Index auf 0 gesetzt. Innerhalb der Schleife wird zunächst das Teilband mit dem größten NMR-Wert bestimmt. Dieses ist das Teilband, bei dem das Quantisierungsgeräusch am schlechtesten verdeckt werden würde. Falls noch genügend Bits zur Verfügung stehen und die höchste Bitzuteilung für diese Teilband noch nicht erreicht ist, wird der Bitzuteilungs-Index erhöht und so der nächstfeineren Quantisierer ausgewählt. Dabei wird die Tabelle der in den unterschiedlichen Teilbändern möglichen Quantisierer berücksichtigt (Anhang A.1). Werden einem Teilband dabei zum ersten Mal Bits zugeteilt, wird berücksichtigt, daß dann auch die SCFSI und die Skalierungsfaktoren für dieses Teilband zu übertragen sind.

Wenn bei der Codierung Intensity Stereo zugelassen wurde, wird vor der Bitzuteilung zunächst ermittelt, ob und gegebenenfalls welche Teilbänder eines Stereosignals gemeinsam codiert werden sollen. Neben der unabhängigen Codierung der beiden Kanäle stehen 4 weitere Modi zur Auswahl, bei denen für die oberen Teilbänder Intensity Stereo verwendet wird. Je nach Modus werden die untersten 4, 8, 12 oder 16 Teilbänder weiterhin unabhängig übertragen. Es wird nun der Modus gewählt, bei dem bei der Bitzuteilung in allen Teilbändern gerade ein negativer NMR-Wert - also eine Maskierung des Quantisierungsgeräusches - erreicht werden kann. Bei der eigentlichen Bitzuteilung wird der größere der SMR-Werte der beiden Kanäle verwendet. Es wird auch berücksichtigt, daß in Teilbändern mit Intensity Stereo die Information über die Bitzuteilung und die quantisierten Teilbandsample nur für das Summensignal codiert werden. Dieses Summensignal entsteht durch Addition der Teilbandsample der beiden Stereokanäle und anschließende Normierung.

Normierung: Die Teilbandsample der einzelnen Blöcke werden nun durch den später bei der Decodierung verwendeten Skalierungsfaktor geteilt. Sie sind durch diese Normierung auf den Wertebereich $[-1,1]$ beschränkt.

Quantisierung: Anschließend werden die normierten Teilbandsample quantisiert. Dazu werden gleichförmige Quantisierer mit 3 bis 65535 Stufen verwendet. Das Intervall $[-1,1]$ wird dabei in entsprechend viele Bereiche gleicher Größe aufgeteilt. Bei der Dequantisierung im Decoder wird später der jeweils in der Bereichsmittle liegende Wert als Repräsentativwert verwendet. Da alle Quantisierer eine ungrade Stufenanzahl verwenden, gibt es immer eine um 0 symmetrische Quantisierungsstufe mit dem Repräsentativwert 0.

Codierung: Abschließend werden der Kopf, gegebenenfalls ein Fehlerkorrekturwort, die gewählte Bitzuteilung, die SCFSI, die Skalierungsfaktoren und die quantisierten Teilbandsample codiert und zu einem neuen Frame des Bitstroms zusammengesetzt.

2.2.3 Struktur des Audio-Decoders für Layer II

In Abschnitt 2.1 wurde schon die grundlegende Struktur des Audio-Decoders aufgezeigt. In Bild 2.4 ist ein Blockdiagramm des Decoders zu sehen. Der Ablauf der Decodierung wird in Bild 2.7 dargestellt.

Bitstrom einlesen
Frame decodieren
Skalierungsfaktoren decodieren
Teilbandsample dequantisieren
Teilbandsample skalieren
Polyphasenfilterbank
PCM-Audiodaten ausgeben

Bild 2.7 Ablauf der Decodierung bei Layer II

Decodierung: Zunächst muß der Decoder anhand des Synchronisationswortes den Anfang des nächsten zu decodierenden Frames ermitteln. Dann werden die einzelnen Bestandteile des Bitstroms eingelesen und nacheinander decodiert. Außerdem wird gegebenenfalls das optional übertragene Fehlerkorrekturwort zur Erkennung oder Korrektur von falsch übertragenen Bits benutzt. Der so geschützte

Bereich beinhaltet den Kopf, die Bitzuteilung und die SCFSI. Die fehlerfreie Decodierung dieser Teile des Bitstroms ist notwendig, um überhaupt den Rest des Bitstroms, also die Skalierungsfaktoren und die quantisierten Teilbandsample, richtig decodieren zu können.

Rekonstruktion: Aus der SCFSI und den übertragenen Skalierungsfaktoren werden die Skalierungsfaktoren für alle 3 Blöcke eines Teilbandsignals ermittelt. Anschließend werden die decodierten Teilbandsample dequantisiert und blockweise mit den Skalenfaktoren multipliziert. Die Abtastrate der so rekonstruierten Teilbandsignale wird nun um den Faktor 32 erhöht. Abschließend werden die Teilbänder mit Hilfe des Polyphasenfilters zum Ausgangssignal zusammengefügt.

3 Interkanal-Prädiktion

Im Rahmen dieser Arbeit wurde untersucht, wie gut Interkanal-Prädiktoren zur Reduktion der Stereo-Redundanz von Tonsignalen geeignet sind. Doch bevor genauer auf die Interkanal-Prädiktion eingegangen wird, folgt zunächst eine kurze Beschreibung der prinzipiellen Funktionsweise von Prädiktoren.

3.1 Prinzipielle Funktionsweise von Prädiktoren

In ihrer üblichen Einsatzweise ermöglichen Prädiktoren eine Redundanzreduktion bei der Quellencodierung von einkanaligen zeitdiskreten Signalen, z.B. Sprachsignalen im Telefonbereich. Dabei werden die durch die Signalquelle bedingten statistischen Bindungen zwischen aufeinanderfolgenden Abtastwerten ausgenutzt. Diese Verfahren werden z.B. als DPCM (Delta Pulse Code Modulation) oder LPC (Linear Prediction Coding) bezeichnet [8/ 17].

Ein solcher Auto-Prädiktor, wie er hier zur eindeutigen Unterscheidung vom Interkanal-Prädiktor genannt werden soll, versucht, aus einer Anzahl zurückliegender Abtastwerte des Signals den aktuellen Abtastwert vorherzusagen. Da der verwendete Prädiktor sowohl dem Coder als auch dem Decoder bekannt sein muß, braucht anstelle des aktuellen Abtastwertes nur noch der Prädiktionsfehler codiert zu werden. Da dieser eine kleinere Leistung als das Originalsignal besitzt, kann er mit entsprechend weniger Stufen quantisiert werden. Auf Grund dieses Prädiktionsgewinns ist also eine Datenratenreduktion möglich. Bild 3.1 zeigt das Blockschaltbild einer DPCM-Codierung.

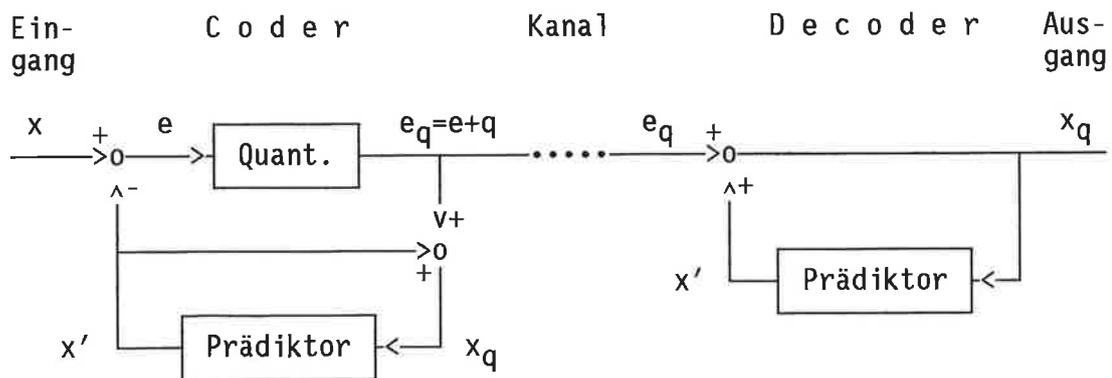


Bild 3.1 Blockschaltbild eines DPCM-Coders und -Decoders [8/

Der Vorhersagewert x' des Prädiktors wird als Linearkombination einer Anzahl vorangegangener rekonstruierte Abtastwerte x_q gebildet. Übertragen wird der quantisierte Vorhersagefehler e_q . Er ergibt sich aus dem Vorhersagefehler e und dem hinzugefügten Quantisierungsfehler q . Zur Bestimmung des optimalen Prädiktors wird die Autokorrelationsfunktion des Eingangssignals x benötigt.

Es können entweder festgelegte oder adaptive Prädiktoren verwendet werden. Im ersten Fall wird der Prädiktor einmal für die zu codierende Signalklasse (z.B. Sprachsignale) optimiert und anschließend festgelegt. Demgegenüber wird ein adaptiver Prädiktor jeweils an die momentanen statistischen Eigenschaften des zu codierenden Signals angepaßt. So ist ein höherer Prädiktionsgewinn möglich. Dabei ist jedoch zu berücksichtigen, daß der jeweils verwendete Prädiktor selbst als Seiteninformation zu übertragen ist. Codierungsverfahren mit adaptivem Prädiktor werden auch als ADPCM (Adaptive Delta Pulse Code Modulation) bezeichnet.

3.2 Grundlagen der Interkanal-Prädiktion

Im Gegensatz zur eben beschriebenen Auto-Prädiktion setzt die Interkanal-Prädiktion voraus, daß das zu codierende Gesamtsignal aus mindestens zwei Kanälen besteht. Mit Hilfe eines Interkanal-Prädiktors wird dann versucht, aus dem Signalverlauf eines Kanals den Signalverlauf eines zweiten Kanals vorherzusagen. Für den zweiten Kanal braucht dann nicht mehr das vollständige Signal sondern nur noch das Prädiktionsfehlersignal codiert zu werden. Auf diese Weise lassen sich statistische Bindungen zwischen den Signalen dieser beiden Kanäle zur Redundanzreduktion ausnutzen.

In Bild 3.2 ist die Codierung mit einem Interkanal-Prädiktor dargestellt. Die Prädiktion erfolgt dabei vom vollständig übertragenen Kanal x auf den Kanal y , für den nur der Prädiktionsfehler e übertragen wird. y' ist der Vorhersagewert des Prädiktors für den Kanal y . In diesem Abschnitt wird von unquantisierten Signalen ausgegangen. Die verschiedenen Quantisierungseffekte werden dann im nächsten Abschnitt betrachtet.

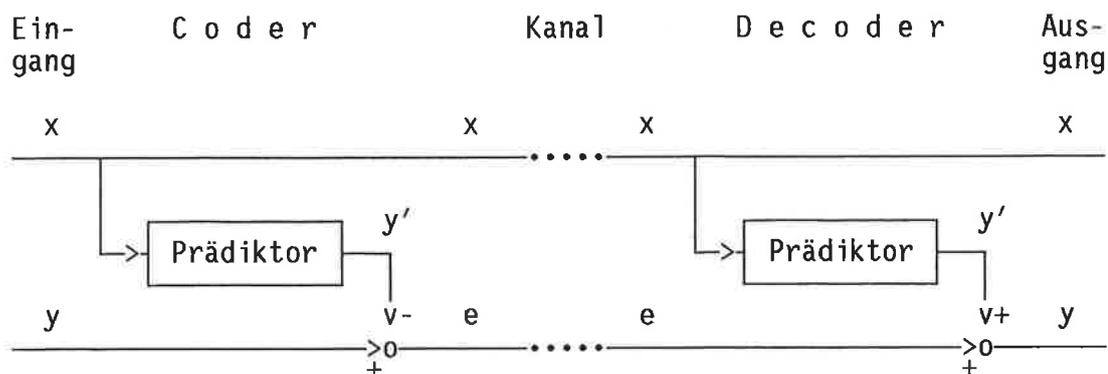


Bild 3.2 Codierung mit Interkanal-Prädiktion

Bei Stereosignalen ist die Interkanal-Prädiktion in zwei Prädiktionsrichtungen möglich: Entweder von links nach rechts (linker Kanal: x , rechter Kanal: y) oder umgekehrt von rechts nach links. Sie wird hier nicht festgelegt, um später jeweils die optimale Prädiktionsrichtung verwenden zu können.

Die Interkanal-Prädiktion wird im Rahmen dieser Arbeit in den ISO Audio-Codec Layer II integriert. Dabei wird für jedes der 32 Teilbänder ein eigener Prädiktor verwendet. Auf diese Weise lassen sich die Vorteile der Prädiktion gut mit dem psychoakustischen Codierungsverfahren von Layer II kombinieren. Wenn die Prädiktion auf die nicht teilbandgefilterten Originalsignale angewendet werden würde, wäre eine Berücksichtigung des psychoakustischen Codierungsverfahrens nur sehr schwer möglich.

Im Rahmen dieser Arbeit wird ein linearer Prädiktor verwendet. Sein Blockschaltbild ist in Bild 3.3 dargestellt.

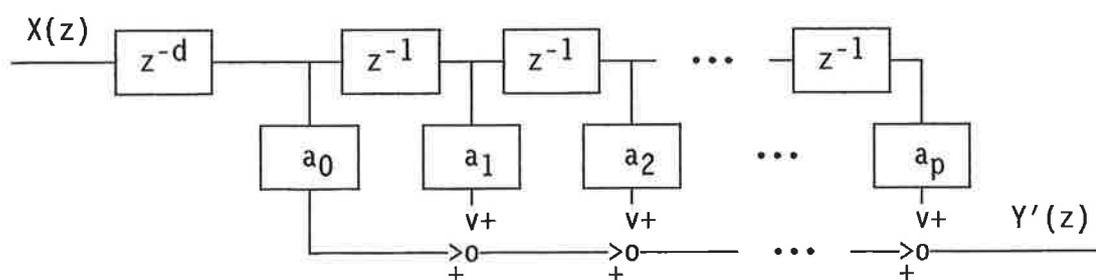


Bild 3.3 Blockschaltbild des Interkanal-Prädiktors

Im Folgenden wird nicht mit den z-Transformierten $X(z)$, $Y(z)$, ... sondern mit den Abtastwerten $x(n)$, $y(n)$, ... im Zeitbereich gerechnet. Der Vorhersagewert $y'(n)$ ergibt sich als Linearkombination von $p+1$ Werten des Signals $x(n)$. Zusätzlich ist eine Verzögerung um d Abtastschritte möglich. Der Vorhersagewert berechnet sich entsprechend dem Blockschaltbild in folgender Weise:

$$y'(n) = \sum_{k=0}^p a_k x(n-d-k) \quad (3.1)$$

Der Prädiktorgrad p ($p \geq 0$) gibt die Komplexität des Prädiktors an. a_0, a_1, \dots, a_p sind die $p+1$ Prädiktorkoeffizienten. Wird die zusätzliche Verzögerung $d = 0$ gewählt, so hat der Prädiktor genau die Struktur eines FIR-Filters.

Der Prädiktionsfehler $e(n)$ aus Bild 3.2 lautet

$$e(n) = y(n) - y'(n) . \quad (3.2)$$

Er besitzt die Leistung

$$\sigma_e^2 = E[e^2(n)] . \quad (3.3)$$

Für den Fall, daß $e(n)$ mittelwertfrei ist (also $E[e(n)] = 0$), ist die Leistung von $e(n)$ gleich der Varianz von $e(n)$, die üblicherweise mit σ^2 bezeichnet wird. Auch wenn in den folgenden Rechnungen bei keinem Signal die Mittelwertfreiheit vorausgesetzt wird, wird hier trotzdem die Bezeichnung σ^2 für die Leistung verwendet.

Die Abtastwerte $x(n)$ und $y(n)$ werden als Ereignisse von zwei zeitdiskreten, wertkontinuierlichen stochastischen Prozessen betrachtet, wobei Stationarität nicht vorausgesetzt wird. Auf mögliche Verfahren zur Berechnung der Erwartungswerte $E[\dots]$ wird am Ende dieses Abschnittes eingegangen.

Mit der Leistung

$$\sigma_y^2 = E[y^2(n)] \quad (3.4)$$

des zu prädizierenden Signals $y(n)$ läßt sich der Prädiktionsgewinn

$$G = \frac{E[y^2(n)]}{E[e^2(n)]} = \frac{\sigma_y^2}{\sigma_e^2} \quad (3.5)$$

berechnen. Ein Gewinn von $G = 4$ bedeutet, daß die Leistung des Prädiktionsfehlers $e(n)$ um $10 \log_{10}(G) = 6,02$ dB kleiner ist als die Leistung des zu prädizierenden Signals $y(n)$.

Um nun den optimalen Prädiktor zu bestimmen, muß der Prädiktionsgewinn G maximiert werden. Da der Prädiktor auf σ_y^2 keinen Einfluß hat, ist die Prädiktionsfehlerleistung σ_e^2 zu minimieren. Zur Bestimmung der Koeffizienten a_k eines optimalen Prädiktors mit vorgegebenem Grad p und vorgegebener Verzögerung d wird daher

$$\frac{d \sigma_e^2}{d a_k} \stackrel{!}{=} 0 \quad \text{für } 0 \leq k \leq p \quad (3.6)$$

gesetzt. Mit (3.1), (3.2) und (3.3) ergibt sich

$$\sigma_e^2 = E \left[\left(y(n) - \sum_{k=0}^p a_k x(n-d-k) \right)^2 \right] . \quad (3.7)$$

Mit (3.6) folgt

$$0 \stackrel{!}{=} \frac{d \sigma_e^2}{d a_i} = E \left[-2 \left(y(n) - \sum_{k=0}^p a_k x(n-d-k) \right) x(n-d-i) \right] \quad \text{für } 0 \leq i \leq p , \quad (3.8)$$

$$E \left[y(n) x(n-d-i) \right] = E \left[\sum_{k=0}^p a_k x(n-d-k) x(n-d-i) \right] \quad \text{für } 0 \leq i \leq p . \quad (3.9)$$

Daraus folgt für die a_k ein lineares Gleichungssystem mit den $p+1$ Gleichungen

$$\sum_{k=0}^p a_k E \left[x(n-d-k) x(n-d-i) \right] = E \left[y(n) x(n-d-i) \right] \quad \text{für } 0 \leq i \leq p . \quad (3.10)$$

Aus (3.7) folgt

$$\sigma_e^2 = E \left[y^2(n) \right] - 2 \sum_{k=0}^p a_k E \left[y(n) x(n-d-k) \right] + \sum_{k=0}^p a_k \sum_{i=0}^p a_i E \left[x(n-d-i) x(n-d-k) \right] . \quad (3.11)$$

Mit (3.10) ergibt sich nun die minimale Prädiktionsfehlerleistung

$$\sigma_e^2 = E[y^2(n)] - \sum_{k=0}^p a_k E[y(n) x(n-d-k)] . \quad (3.12)$$

Aus (3.4), (3.5) und (3.12) folgt der maximale Prädiktionsgewinn

$$G = \frac{1}{1 - \frac{1}{\sigma_y^2} \sum_{k=0}^p a_k E[y(n) x(n-d-k)]} . \quad (3.13)$$

Um eine übersichtlichere Schreibweise dieser Gleichungen zu ermöglichen, werden bei vorgegebener fester Verzögerung d die Leistung des Signals $x(n-d-k)$

$$\sigma_{xk}^2 = E[x^2(n-d-k)] \quad (3.14)$$

die Autokorrelationskoeffizienten

$$r_{i,k} = r_{k,i} = \frac{E[x(n-d-k) x(n-d-i)]}{\sigma_{xk} \sigma_{xi}} \quad (3.15)$$

sowie die Kreuzkorrelationskoeffizienten

$$c_i = \frac{E[y(n) x(n-d-i)]}{\sigma_y \sigma_{xi}} \quad (3.16)$$

eingeführt. Auch hier wird, wie bei den Leistungen σ^2 , keine Mittelwertfreiheit der Signale vorausgesetzt. Daher sind die obigen Definitionen der Korrelationskoeffizienten in diesem Punkt verschieden von den üblichen Definitionen.

Das lineare Gleichungssystem (3.10) lautet nun

$$\sum_{k=0}^p a_k r_{k,i} \sigma_{xk} = c_i \sigma_y \quad \text{für } 0 \leq i \leq p . \quad (3.17)$$

In Matrix-Schreibweise ergibt sich folgende Darstellung:

$$\begin{bmatrix} r_{0,0} & r_{0,1} & \dots & r_{0,p} \\ r_{1,0} & r_{1,1} & \dots & r_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p,0} & r_{p,1} & \dots & r_{p,p} \end{bmatrix} \begin{bmatrix} a_0 \sigma_{x0} \\ a_1 \sigma_{x1} \\ \vdots \\ a_p \sigma_{xp} \end{bmatrix} = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_p \end{bmatrix} \sigma_y \quad (3.18)$$

Aufgrund von $r_{k,i} = r_{i,k}$ ist die Matrix $[r_{k,i}]$ symmetrisch. Mit (3.14) und (3.15) folgt für die Diagonalelemente

$$r_{k,k} = \frac{\sigma_{xk}^2}{\sigma_{xk} \sigma_{xk}} = 1 . \quad (3.19)$$

Die Prädiktionsfehlerleistung aus (3.12) läßt sich jetzt folgendermaßen schreiben:

$$\sigma_e^2 = \sigma_y^2 - \sigma_y^2 \sum_{k=0}^p a_k c_k \sigma_{xk} \quad (3.20)$$

Und der Prädiktionsgewinn aus (3.13) lautet nun

$$G = \frac{1}{1 - \frac{1}{\sigma_y} \sum_{k=0}^p a_k c_k \sigma_{xk}} . \quad (3.21)$$

Zur Bestimmung der $p+1$ Koeffizienten a_k des optimalen Prädiktors ist also das lineares Gleichungssystem (3.18) mit $p+1$ Gleichungen zu lösen. Für Prädiktoren mit einem oder zwei Koeffizienten - dieses sind die beiden einfachsten Fälle - sollen die Lösungen hier explizit angegeben werden.

Für Prädiktoren mit $p = 0$, also mit einem Koeffizienten a_0 , ergibt sich

$$a_0 = \frac{\sigma_y}{\sigma_{x0}} c_0 , \quad (3.22)$$

$$\sigma_e^2 = \sigma_y^2 (1 - c_0^2) , \quad (3.23)$$

$$G = \frac{1}{1 - c_0^2} . \quad (3.24)$$

Für Prädiktoren mit $p = 1$, also mit zwei Koeffizienten a_0 und a_1 , ergibt sich

$$a_0 = \frac{\sigma_Y c_0 - c_1 r_{0,1}}{\sigma_{x0} (1 - r_{0,1}^2)}, \quad (3.25)$$

$$a_1 = \frac{\sigma_Y c_1 - c_0 r_{0,1}}{\sigma_{x1} (1 - r_{0,1}^2)}, \quad (3.26)$$

$$\sigma_e^2 = \sigma_Y^2 \left(1 - \frac{c_0^2 + c_1^2 - 2 c_0 c_1 r_{0,1}}{1 - r_{0,1}^2} \right), \quad (3.27)$$

$$G = \frac{1}{1 - \frac{c_0^2 + c_1^2 - 2 c_0 c_1 r_{0,1}}{1 - r_{0,1}^2}}. \quad (3.28)$$

Für komplexere Prädiktoren mit $p \geq 2$ ist es sinnvoll, das lineare Gleichungssystem (3.18) mit geeigneten Verfahren zu lösen. Da die Matrix $[r_{k,j}]$ symmetrisch und in der Praxis auch positiv definit ist, läßt sich das Verfahren von Cholesky anwenden /4/. Dabei sind deutlich weniger Rechenschritte nötig als zum Beispiel beim Gauß-Verfahren.

Die den optimalen Prädiktor bestimmenden Signaleigenschaften variieren bei den in der Praxis vorkommenden Tonsignalen sehr stark. Daher ist es sinnvoll, den Prädiktor jeweils den momentanen Signaleigenschaften anzupassen. Dieses läßt sich zum Beispiel dadurch erreichen, daß die Signale $x(n)$ und $y(n)$ in Zeitabschnitte mit jeweils N Abtastwerten gegliedert werden. Für einen solchen Block kann dann der optimale Prädiktor bestimmt und für die Codierung verwendet werden. Die in den obigen Gleichungen auftretenden Erwartungswerte sollten dabei in folgender Weise bestimmt werden:

$$E[x(n)] = \frac{1}{N} \sum_{j=0}^{N-1} x(n+j) \quad (3.29)$$

So wird nämlich die Energie des Fehlersignals

$$\sum_{j=0}^{N-1} e^2(n+j) = N \sigma_e^2 \quad (3.30)$$

in einem Block minimiert. Ein Block besteht dabei aus den N Abtastwerten n bis $n+N-1$ der Signale $x(n)$ und $y(n)$. Es ist jedoch zu beachten, daß für die Berech-

nung des optimalen Prädiktors auch noch die letzten $d+p$ Abtastwerte des Signals $x(n)$ im vorangegangenen Block benötigt werden.

Es ist auch möglich, den Prädiktor als einen Filter zu betrachten. Mit dem Ansatz

$$Y'(z) = X(z) H(z) \quad (3.31)$$

ergibt sich die Übertragungsfunktion des Prädiktors

$$H(z) = z^{-d} \sum_{k=0}^p a_k z^{-k} \quad (3.32)$$

durch z -Transformation aus der Gleichung (3.1).

Ein optimaler Prädiktor versucht also aus dem Spektrum des Signals $x(n)$ durch Filterung möglichst gut das Spektrum des Signals $y(n)$ zu formen. Dabei werden sowohl die Beträge wie auch die Phasen der Spektren berücksichtigt.

Bei der Auto-Prädiktion befinden sich die Prädiktoren in einer rückgekoppelten Schleife (siehe Bild 3.1). Somit ergibt sich die Struktur eines FIR-Filters. Es ist daher notwendig, bei der Auto-Prädiktion auch Betrachtungen zur Stabilität des rückgekoppelten Prädiktors anzustellen. Da bei der Interkanal-Prädiktion solche Rückkopplungsschleifen nicht auftreten (siehe Bild 3.2), sind dort auch keine Stabilitätsbetrachtungen notwendig.

3.3 Quantisierungseffekte bei der Interkanal-Prädiktion

Wird die Interkanal-Prädiktion in einem Codec eingesetzt, so sind die Effekte der für die Codierung notwendigen Quantisierung zu berücksichtigen. Ausgehend vom bisher verwendeten Blockschaltbild (Bild 3.2) ist eine Quantisierung an zwei verschiedenen Stellen notwendig. Zum einen müssen die übertragenen Signale x und e quantisiert werden. Außerdem muß auch der für einen Block verwendete Prädiktor übertragen werden. Dieses macht eine Quantisierung der Prädiktorkoeffizienten a_k notwendig.

Bild 3.4 zeigt ein Blockschaltbild für die Codierung mit Interkanal-Prädiktion, in dem diese beiden Quantisierungseffekte berücksichtigt sind. "Q" steht dabei für einen Quantisierer. "QPräd" stellt den Prädiktor mit quantisierten Koeffizienten dar. Damit sich bei der Decodierung des Signals y keine Quantisierungsfehler un-

günstig überlagern, ist es wichtig, daß auch bei der Codierung die Prädiktion von dem schon quantisierten Signal x_q ausgeht.

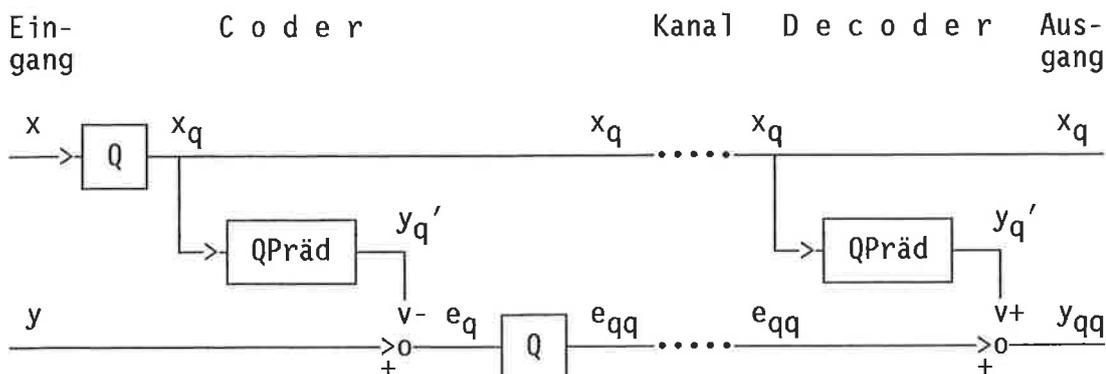


Bild 3.4 Codierung mit Interkanal-Prädiktion und Quantisierung

Auf die bei der Quantisierung der Prädiktorkoeffizienten und Signale auftretenden Effekte wird in den beiden nächsten Abschnitten eingegangen.

3.3.1 Quantisierung der Prädiktorkoeffizienten

Der für einen Block verwendete Prädiktor muß codiert und zum Decoder übertragen werden, um so eine Rekonstruktion des Signals y zu ermöglichen. Dazu müssen auch die $p+1$ Koeffizienten a_k , die in der Prädiktorgleichung (3.1) auftreten, quantisiert werden. Die quantisierten Prädiktorkoeffizienten a_{qk} lassen sich folgendermaßen darstellen:

$$a_{qk} = a_k + q_{ak} \quad \text{für } 0 \leq k \leq p \quad (3.33)$$

q_{ak} ist dabei der Quantisierungsfehler des Koeffizienten a_k .

In den folgenden Rechnungen wird das Eingangssignal des Prädiktors mit $x(n)$ bezeichnet. Daß später, wie in Bild 3.4 dargestellt, als Eingangssignal das quantisierte Signal $x_q(n)$ verwendet wird, hat auf die Rechnungen in diesem Abschnitt keinen Einfluß. Der bei der Prädiktion mit quantisierten Prädiktorkoeffizienten erzeugte Vorhersagewert wird mit $y_q'(n)$ bezeichnet. Das Prädiktionsfehlersignal heißt $e_q(n)$. Aus den Gleichungen (3.1) und (3.2) folgt für quantisierte Koeffizienten

$$y_q'(n) = \sum_{k=0}^p a_{qk} x(n-d-k) , \quad (3.34)$$

$$e_q(n) = y(n) - y_q'(n) . \quad (3.35)$$

Für die Leistung des Prädiktionsfehlers $e_q(n)$ ergibt sich damit anstelle von (3.7) nun

$$\sigma_{eq}^2 = E[e_q^2(n)] = E[(y(n) - \sum_{k=0}^p a_{qk} x(n-d-k))^2] . \quad (3.36)$$

Daraus folgt

$$\sigma_{eq}^2 = E[y^2(n)] - 2 \sum_{k=0}^p a_{qk} E[y(n) x(n-d-k)] + \sum_{k=0}^p a_{qk} \sum_{i=0}^p a_{qi} E[x(n-d-k) x(n-d-i)] , \quad (3.37)$$

$$\sigma_{eq}^2 = \sigma_y^2 - 2 \sum_{k=0}^p (a_k + q_{ak}) c_k \sigma_y \sigma_{xk} + \sum_{k=0}^p (a_k + q_{ak}) \sum_{i=0}^p (a_i + q_{ai}) r_{k,i} \sigma_{xk} \sigma_{xi} . \quad (3.38)$$

Für die folgenden Rechnungen müssen einige Annahmen über die Quantisierungsfehler q_{ak} der Prädiktorkoeffizienten gemacht werden. So werden die q_{ak} als mittelwertfrei angenommen. Dieses bedeutet

$$E[q_{ak}] = 0 \quad \text{für } 0 \leq k \leq p . \quad (3.39)$$

Außerdem wird davon ausgegangen, daß die q_{ak} statistisch unabhängig von den Signalen $x(n)$ und $y(n)$ und den Prädiktorkoeffizienten a_k sind. Auch die Quantisierungsfehler q_{ak} und q_{ai} zweier verschiedener Koeffizienten a_k und a_i sollen statistisch unabhängig voneinander sein. Daraus folgt mit (3.39)

$$E[q_{ak} q_{ai}] = 0 \quad \text{für } k \neq i . \quad (3.40)$$

Bei einer hinreichend feinen Quantisierung der Prädiktorkoeffizienten sind diese Annahmen berechtigt.

Aus (3.38) folgt mit diesen Annahmen für den Erwartungswert von σ_{eq}^2 bezüglich der Quantisierungsfehler der Prädiktorkoeffizienten

$$E[\sigma_{eq}^2] = \sigma_y^2 - 2 \sum_{k=0}^p (a_k + q_{ak}) c_k \sigma_y \sigma_{xk} + \sum_{k=0}^p (a_k + q_{ak}) \sum_{i=0}^p (a_i + q_{ai}) r_{k,i} \sigma_{xk} \sigma_{xi} \quad (3.41)$$

und

$$E[\sigma_{eq}^2] = \sigma_y^2 - 2 \sum_{k=0}^p (a_k + E[q_{ak}]) c_k \sigma_y \sigma_{xk} + \sum_{k=0}^p \sum_{i=0}^p E[a_k a_i + a_k q_{ai} + q_{ak} a_i + q_{ak} q_{ai}] r_{k,i} \sigma_{xk} \sigma_{xi} \quad (3.42)$$

Mit (3.39), (3.40) und (3.19) läßt sich (3.42) vereinfachen:

$$E[\sigma_{eq}^2] = \sigma_y^2 - 2 \sum_{k=0}^p a_k c_k \sigma_y \sigma_{xk} + \sum_{k=0}^p a_k \sum_{i=0}^p a_i r_{k,i} \sigma_{xk} \sigma_{xi} + \sum_{k=0}^p E[q_{ak}^2] \sigma_{xk}^2 \quad (3.43)$$

Mit (3.17) folgt nun

$$E[\sigma_{eq}^2] = \sigma_y^2 - \sigma_y \sum_{k=0}^p a_k c_k \sigma_{xk} + \sum_{k=0}^p E[q_{ak}^2] \sigma_{xk}^2 \quad (3.44)$$

Mit Hilfe der Gleichung (3.20) für die Prädiktionsfehlerleistung σ_e^2 bei unquantisierten Koeffizienten ergibt sich aus (3.44) dann folgende Gleichung:

$$E[\sigma_{eq}^2] = \sigma_e^2 + \sum_{k=0}^p E[q_{ak}^2] \sigma_{xk}^2 \quad (3.45)$$

Mit zwei weiteren Annahmen läßt sich diese Gleichung weiter vereinfachen. Ist der stochastische Prozeß für das Signal $x(n)$ stationär, so gilt

$$\sigma_{xk}^2 = \sigma_x^2 \quad \text{für } 0 \leq k \leq p \quad (3.46)$$

Bei hinreichend feiner Quantisierung der Prädiktorkoeffizienten ist außerdem folgende Annahme sinnvoll:

$$E[q_{ak}^2] = E[q_a^2] \quad \text{für } 0 \leq k \leq p \quad (3.47)$$

Mit diesen beiden Annahmen folgt aus (3.45)

$$E[\sigma_{eq}^2] = \sigma_e^2 + (p+1) E[q_a^2] \sigma_x^2 \quad (3.48)$$

Diese Gleichung zeigt deutlich, daß sich die Prädiktionsfehlerleistung σ_{eq}^2 bei quantisierten Koeffizienten aus der Prädiktionsfehlerleistung σ_e^2 bei unquantisierten Koeffizienten sowie einer zusätzlichen, durch die Quantisierung der Koeffi-

zienten bedingten Leistung zusammensetzt. Diese zusätzliche Leistung hängt von dem Grad p des Prädiktors, dem Erwartungswert $E[q_a^2]$ des Koeffizientenquantisierungsfehlerquadrates sowie der Leistung σ_x^2 des Signals $x(n)$ ab.

Unter der Annahme, daß mit dem optimalen Prädiktor eine exakte Vorhersage des Signals $y(n)$ möglich ist, also $\sigma_e^2 = 0$ gilt, ergibt sich für den Erwartungswert der Leistung des Prädiktionsfehlersignals bei quantisierten Koeffizienten folgende Gleichung:

$$E[\sigma_{eq}^2] = (p+1) E[q_a^2] \sigma_x^2 \quad (3.49)$$

Besitzen nun die Signale $x(n)$ und $y(n)$ die gleiche Leistung $\sigma_x^2 = \sigma_y^2$, so errechnet sich für diesen Erwartungswert $E[\sigma_{eq}^2]$ folgender mittlerer Prädiktionsgewinn

$$G_{qa,mittel} = \frac{\sigma_y^2}{E[\sigma_{eq}^2]} = \frac{1}{(p+1) E[q_a^2]} \quad (3.50)$$

Wird für die Prädiktorkoeffizienten eine hinreichend feine gleichförmige Quantisierung mit der Stufenbreite δ verwendet, so darf angenommen werden, daß der Quantisierungsfehler q_a im Intervall $[-\delta/2, \delta/2]$ gleichverteilt ist. Für $E[q_a^2]$ folgt daraus (siehe /8/)

$$E[q_a^2] = \delta^2/12 \quad (3.51)$$

Der mittleren Prädiktionsgewinn aus (3.50) lautet nun

$$G_{qa,mittel} = \frac{12}{(p+1) \delta^2} \quad (3.52)$$

Bild 3.0 zeigt eine Tabelle, in der diese Prädiktionsgewinne in dB angegeben sind. Die "Anzahl Bits" ist für die Codierung eines Koeffizienten notwendig, wenn er im Intervall $[-2,2]$ mit der Stufenbreite δ quantisiert wird. Bei einer Halbierung der Stufenbreite δ erhöht sich der mittlere Prädiktionsgewinn um 6,02 dB. Wird die Anzahl der Prädiktorkoeffizienten $p+1$ vervierfacht, so verringert er sich um 6,02 dB.

δ	Anz. Bits	Prädiktorgrad p							
		0	1	2	3	5	7	9	15
1/2	3	16,8	13,8	12,0	10,8	9,0	7,8	6,8	4,8
1/4	4	22,8	19,8	18,1	16,8	15,1	13,8	12,8	10,8
1/8	5	28,9	25,8	24,1	22,8	21,1	19,8	18,9	16,8
1/16	6	34,9	31,9	30,1	28,9	27,1	25,8	24,9	22,8
1/32	7	40,9	37,9	36,1	34,9	33,1	31,9	30,9	28,9
1/64	8	46,9	43,9	42,1	40,9	39,1	37,9	36,9	34,9

Bild 3.5 Mittlerer Prädiktionsgewinn in dB bei $\sigma_e^2 = 0$, $\sigma_x^2 = \sigma_y^2$ und quantisierten Prädiktorkoeffizienten (Quantisierungsstufenbreite δ)

Aus dieser Tabelle läßt sich ablesen, welche Größe der mittlere Prädiktionsgewinn bei quantisierten Koeffizienten maximal erreicht. Ist $\sigma_e^2 > 0$, so folgt aus (3.48), daß der mittlere Prädiktionsgewinn kleiner als er hier tabellierte Wert ist.

Die hier berechneten Obergrenzen für die mittleren Prädiktionsgewinne bei quantisierten Koeffizienten stimmen mit den bei Untersuchungen an realen Tonsignalen ermittelten Prädiktionsgewinnen gut überein. Die in diesem Abschnitt ermittelten Auswirkungen der Quantisierung der Prädiktorkoeffizienten werden in dem später beschriebenen Verfahren für die Codierung der Prädiktoren berücksichtigt.

3.3.2 Prädiktion bei quantisierten Signalen

Um die bei der Interkanal-Prädiktion zu übertragenen Signale $x(n)$ und $e(n)$ codieren zu können, ist es notwendig, sie zu quantisieren. Im Blockschaltbild in Bild 3.4 sind die beiden benötigten Quantisierer mit "Q" bezeichnet.

Dem Signal $x(n)$ wird der Quantisierungsfehler $q_x(n)$ überlagert. Für das quantisierte Signal $x_q(n)$ gilt

$$x_q(n) = x(n) + q_x(n) . \quad (3.53)$$

Die Prädiktion geht nun sowohl im Coder wie im Decoder vom quantisierten Signal $x_q(n)$ aus. Das Ausgangssignal $y_q'(n)$ des Prädiktors enthält daher neben dem eigentlichen Vorhersagewert $y'(n)$ auch den durch den Prädiktor gefilterten Quantisierungsfehler $q_x(n)$. Dieser gefilterte Quantisierungsfehler ist auch im Prädiktionsfehlersignal $e_q(n)$ enthalten. Es lautet

$$e_q(n) = y(n) - y_q'(n) . \quad (3.54)$$

Auch dieses Signal $e_q(n)$ muß für die Codierung quantisiert werden. Dabei wird der Quantisierungsfehler $q_e(n)$ hinzugefügt. Es ergibt sich folgendes quantisiertes Prädiktionsfehlersignal $e_{qq}(n)$:

$$e_{qq}(n) = e_q(n) + q_e(n) \quad (3.55)$$

Bei der Rekonstruktion im Decoder entsteht das Signal $y_{qq}(n)$. Dabei gilt

$$y_{qq}(n) = e_{qq}(n) + y_q'(n) \quad (3.56)$$

Mit (3.54) und (3.55) folgt daraus

$$y_{qq}(n) = y(n) - y_q'(n) + q_e(n) + y_q'(n) = y(n) + q_e(n) \quad (3.57)$$

Dem Signal $y(n)$ wird also bei der Rekonstruktion nur der Quantisierungsfehler $q_e(n)$ überlagert. Wäre die Prädiktion beim Coder nicht von $x_q(n)$ sondern von $x(n)$ ausgegangen, so würde das im Decoder rekonstruierte Signal $y_{qq}(n)$ zusätzlich noch den durch den Prädiktor gefilterten Quantisierungsfehler $q_x(n)$ enthalten.

Der Prädiktor selbst sowie die Quantisierung der Prädiktorkoeffizienten hat keinen Einfluß auf den in Gleichung (3.57) dargestellten Zusammenhang. Jedoch hat die Tatsache, daß die Prädiktion vom quantisierten Signal $x_q(n)$ ausgeht, Auswirkungen auf den erzielbaren Prädiktionsgewinn. Diese sollen nun genauer betrachtet werden.

Wird in die Gleichung (3.1) für den Vorhersagewert anstelle von $x(n)$ das quantisierte Signal $x_q(n)$ eingesetzt, so ergibt sich mit (3.53)

$$y_q'(n) = \sum_{k=0}^p a_k (x(n-d-k) + q_x(n-d-k)) \quad (3.58)$$

Aus dem Quantisierungsfehler $q_x(n)$ entsteht bei der Filterung durch den Prädiktor das Fehlersignal $q_{xy}(n)$, für das gilt

$$q_{xy}(n) = \sum_{k=0}^p a_k q_x(n-d-k) \quad (3.59)$$

Mit (3.1) und (3.59) läßt sich (3.58) nun in folgender Weise schreiben:

$$y_q'(n) = y'(n) + q_{xy}(n) \quad (3.60)$$

Mit (3.54), (3.60) und (3.2) folgt für das Prädiktionsfehlersignal $e_q(n)$

$$e_q(n) = y(n) - y'(n) - q_{xy}(n) = e(n) - q_{xy}(n) . \quad (3.61)$$

Dabei wird hier davon ausgegangen, daß die Koeffizienten a_k des Prädiktors gemäß Gleichung (3.17) aus den unquantisierten Signalen $x(n)$ und $y(n)$ berechnet wurden.

Für die folgenden Rechnungen sind einige Annahmen über den Quantisierungsfehler $q_x(n)$ notwendig, die bei hinreichend feiner Quantisierung von $x(n)$ auch berechtigt sind. So wird davon ausgegangen, daß $q_x(n)$ mittelwertfrei ist. Es gilt also

$$E[q_x(n)] = 0 . \quad (3.62)$$

Außerdem soll $q_x(n)$ von den Signalen $x(n)$ und $y(n)$ statistisch unabhängig sein. Daraus folgt, daß $q_x(n)$ auch von $e(n)$ statistisch unabhängig ist. Mit (3.62) folgt daraus

$$E[q_x(n-d-k) e(n)] = 0 \quad \text{für } 0 \leq k \leq p . \quad (3.63)$$

Es wird auch angenommen, daß das Quantisierungsfehlersignal $q_x(n)$ ein konstantes (also weißes) Leistungsdichtespektrum hat und die Leistung $\sigma_{q_x}^2$ besitzt. Daraus folgt

$$E[q_x(n-d-k) q_x(n-d-i)] = 0 \quad \text{für } k \neq i , \quad (3.64)$$

$$E[q_x^2(n-d-k)] = \sigma_{q_x}^2 \quad \text{für } 0 \leq k \leq p . \quad (3.65)$$

Da $q_{xy}(n)$ gemäß Gleichung (3.59) eine Linearkombination von $p+1$ Werten von $q_x(n)$ ist, ist auch $q_{xy}(n)$ mittelwertfrei und von $e(n)$ statistisch unabhängig. Es gilt also

$$E[q_{xy}(n)] = 0 , \quad (3.66)$$

$$E[q_{xy}(n) e(n)] = 0 . \quad (3.67)$$

Für die Leistung des Prädiktionsfehlersignals $e_q(n)$ aus Gleichung (3.61) gilt

$$\sigma_{e_q}^2 = E[e_q^2(n)] = E[e^2(n) - 2 e(n) q_{xy}(n) + q_{xy}^2(n)] . \quad (3.68)$$

Mit (3.67) vereinfacht sich diese Gleichung zu

$$\sigma_{e_q}^2 = E[e^2(n)] + E[q_{xy}^2(n)] = \sigma_e^2 + \sigma_{q_{xy}}^2 . \quad (3.69)$$

Für die Leistung des prädiktorgefilterten Quantisierungsfehlersignals $q_{xy}(n)$ gilt mit (3.59)

$$\sigma_{q_{xy}}^2 = E[q_{xy}^2(n)] = E\left[\sum_{k=0}^p \sum_{i=0}^p a_k a_i q_x(n-d-k) q_x(n-d-i)\right]. \quad (3.70)$$

Mit (3.64) und (3.65) vereinfacht sich diese Gleichung zu

$$\sigma_{q_{xy}}^2 = \sum_{k=0}^p a_k^2 E[q_x^2(n-d-k)] = \sigma_{q_x}^2 \sum_{k=0}^p a_k^2. \quad (3.71)$$

Es wird nun der Rauschleistungsübertragungsfaktor K des Prädiktors eingeführt. Er gibt das Verhältnis der Rauschleistung $\sigma_{q_{xy}}^2$ am Prädiktorausgang zur Rauschleistung $\sigma_{q_x}^2$ des weißen Rauschens am Prädiktoreingang an. Der Rauschleistungsübertragungsfaktor läßt sich auch aus der Übertragungsfunktion $H(z)$ des Prädiktors nach Gleichung (3.32) bestimmen /10/. Es gelten folgende Zusammenhänge:

$$K = \frac{\sigma_{q_{xy}}^2}{\sigma_{q_x}^2} = \sum_{k=0}^p a_k^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |H(e^{j\omega})|^2 d\omega \quad (3.72)$$

Die Gleichung (3.71) lautet nun

$$\sigma_{q_{xy}}^2 = K \sigma_{q_x}^2. \quad (3.73)$$

Wird dieser Ausdruck für $\sigma_{q_{xy}}^2$ in (3.69) eingesetzt, so ergibt sich für die Leistung von $e_q(n)$

$$\sigma_{e_q}^2 = \sigma_e^2 + K \sigma_{q_x}^2. \quad (3.74)$$

Aus dieser Gleichung geht hervor, daß die Prädiktionsfehlerleistung $\sigma_{e_q}^2$ bei vom quantisierten Signal $x_q(n)$ ausgehender Prädiktion neben σ_e^2 eine zusätzliche Leistung enthält, die von der Quantisierungsfehlerleistung $\sigma_{q_x}^2$ und vom Rauschleistungsübertragungsfaktor K des Prädiktors abhängt.

Wird nun angenommen, daß der optimale Prädiktor von $x(n)$ ausgehend eine exakte Vorhersage von $y(n)$ ermöglicht, so gilt $\sigma_e^2 = 0$. Außerdem sollen $x(n)$ und $y(n)$ die gleiche Leistung $\sigma_x^2 = \sigma_y^2$ besitzen. Trotz dieser Annahmen ist es

schwierig, die Summe der a_k^2 abzuschätzen, da die Prädiktorkoeffizienten stark von der Art des Zusammenhanges zwischen $x(n)$ und $y(n)$ abhängen. An einem einfachen Beispiel (Beispiel 3) wird diese in Abschnitt 3.4.1 deutlich.

Um trotz dieses Problem es weiterrechnen zu können, wird jetzt angenommen, daß

$$\sum_{k=0}^p a_k^2 = 1 \quad (3.75)$$

gilt. Eine Betrachtung der in der Praxis vorkommenden Prädiktorkoeffizienten zeigt, daß diese Annahme relativ realistisch ist.

Aus den hier gemachten Annahmen und (3.74) folgt

$$\sigma_{eq}^2 = \sigma_{qx}^2 . \quad (3.76)$$

Für den Prädiktionsgewinn gilt damit

$$G_{qx} = \frac{\sigma_y^2}{\sigma_{qx}^2} = \frac{\sigma_x^2}{\sigma_{qx}^2} . \quad (3.77)$$

Der Prädiktionsgewinn G_{qx} ist also unter den getroffenen Annahmen so groß wie der Signal/Rausch-Abstand des quantisierten Signals $x_q(n)$.

Bisher wurde davon ausgegangen, daß der optimale Prädiktor für die Signale $x(n)$ und $y(n)$ berechnet wurde, die Prädiktion dann aber vom quantisierten Signal $x_q(n)$ ausgeht. Wird jedoch der optimale Prädiktor für das schon quantisierte Signal $x_q(n)$ und $y(n)$ bestimmt, so kann der erzielte Prädiktionsgewinn dadurch nur größer werden. Dieser Fall soll nun etwas genauer betrachtet werden.

Für die folgenden Rechnungen werden die oben über den Quantisierungsfehler $q_x(n)$ gemachten Annahmen vorausgesetzt. Die Signalleistungen σ_x^2 und σ_y^2 dürfen unterschiedlich sein. Außerdem sei $x(n)$ durch einen stationären Zufallsprozeß charakterisierbar. Mit (3.14) lautet die Leistung von $x_q(n)$ dann

$$\sigma_{xq}^2 = \sigma_x^2 + \sigma_{qx}^2 . \quad (3.78)$$

Für die Korrelationskoeffizienten ergibt sich mit (3.15) und (3.16)

$$r_{qi,k} = \frac{r_{i,k} \sigma_x^2 + \sigma_{qx}^2 \delta(i-k)}{\sigma_x^2 + \sigma_{qx}^2} \quad (\text{mit } \delta(0)=1 \text{ und } \delta(i)=0 \text{ für } i \neq 0), \quad (3.79)$$

$$c_{qi} = \frac{c_i \sigma_x}{(\sigma_x^2 + \sigma_{qx}^2)^{1/2}}. \quad (3.80)$$

Für einen Prädiktor mit einem Koeffizienten (also $p = 0$) sei eine exakte Vorhersage von $x(n)$ auf $y(n)$ möglich. Daraus folgt $\sigma_e^2 = 0$ und $c_0^2 = 1$. Für den Prädiktorkoeffizienten a_0 gilt Gleichung (3.22). Wird der für das unquantisierte Signal $x(n)$ berechnete Prädiktor nun auf das quantisierte Signal angewendet, so ergibt sich für diesen Fall I die Prädiktionsfehlerleistung

$$\sigma_{eq,I}^2 = \sigma_y^2 \frac{\sigma_{qx}^2}{\sigma_x^2} \quad (3.81)$$

und der Prädiktionsgewinn

$$G_{qx,I} = \frac{\sigma_y^2}{\sigma_{eq,I}^2} = \frac{\sigma_x^2}{\sigma_{qx}^2}. \quad (3.82)$$

Wird der Prädiktor jedoch mit (3.78), (3.79) und (3.80) für das schon quantisierte Signal $x_q(n)$ berechnet, so ergibt sich für den neuen Prädiktorkoeffizienten a_{q0}

$$a_{q0} = a_0 \frac{\sigma_x^2}{\sigma_x^2 + \sigma_{qx}^2}. \quad (3.83)$$

Für diesen Fall II lautet die Prädiktionsfehlerleistung

$$\sigma_{eq,II}^2 = \sigma_y^2 \frac{\sigma_{qx}^2}{\sigma_x^2 + \sigma_{qx}^2} \quad (3.84)$$

und der Prädiktionsgewinn

$$G_{qx,II} = \frac{\sigma_y^2}{\sigma_{eq,II}^2} = 1 + \frac{\sigma_x^2}{\sigma_{qx}^2}. \quad (3.85)$$

Wenn der optimale Prädiktor für das schon quantisierte Signal $x_q(n)$ berechnet wird, ergibt sich also ein betragsmäßig etwas kleinerer Prädiktorkoeffizient und ein etwas größerer Prädiktionsgewinn. Dieser Unterschied zwischen Fall I und Fall II

wird jedoch um so kleiner, je größer der Signal/Rausch-Abstand σ_x^2/σ_{qx}^2 des quantisierten Signals $x_q(n)$ ist.

Um die Auswirkung der Quantisierung des Signals $x(n)$ auf den erzielbaren Prädiktionsgewinn zu berücksichtigen, wird später der optimale Prädiktor für das schon quantisierte Signal $x_q(n)$ berechnet. Dieses macht aber eine Schätzung des für $x(n)$ gewählten Quantisiereres notwendig.

3.4 Ermittlung des erzielbaren Prädiktionsgewinns

Um zu untersuchen, welche Prädiktionsgewinne in der Praxis möglich sind, wurden verschiedene teilbandgefilterte Musiksignale entsprechend analysiert. Dabei sind die Randbedingungen zu berücksichtigen, die sich bei der Erweiterung des Audio-Codecs für Layer II um die Interkanal-Prädiktion ergeben.

So wird eine Blocklänge von $N = 36$ Teilbandsample gewählt, da bei Layer II in einem Frame pro Teilband und Kanal jeweils 36 Teilbandsample codiert werden. Die Interkanal-Prädiktion soll die systembedingte Verzögerung in Coder und Decoder nicht erhöhen. Daher darf bei der Prädiktion nur auf den aktuellen Block oder auf schon übertragene Blöcke zurückgegriffen werden.

Damit möglicherweise in einem Block auftretende Übertragungsfehler nicht aufgrund der Prädiktion auch mehrere Blöcke später noch Störungen des decodierten Signals zur Folge haben, soll bei der Decodierung nur auf die im aktuellen und im vorhergehenden Block übertragenen Teilbandsample zurückgegriffen werden. Daher dürfen bei der Rekonstruktion eines Teilbandsamples des Kanals y mit Hilfe des Prädiktors nur Teilbandsample aus dem Kanal x benutzt werden, die 0 bis $N/2 = 18$ Abtastschritte zurückliegen. Würden noch weiter zurückliegende Teilbandsample benutzt werden, so könnte bei von Block zu Block wechselnder Prädiktionsrichtung das oben beschriebene Problem auftreten. Somit sind die für Prädiktorgrad p und Verzögerung d möglichen Werte durch die folgenden Ungleichungen begrenzt:

$$p + d \leq 18 \quad \text{und} \quad p \geq 0 \quad \text{und} \quad d \geq 0 \quad (3.86)$$

Es ist auch zu berücksichtigen, daß der verwendete Prädiktor selbst als Seiteninformation zu übertragen ist und sich daher der bei der Codierung erzielte Prädiktionsgewinn entsprechend verringert. Außerdem muß ein geeignetes Verfahren zur

Quantisierung der Prädiktorkoeffizienten und zur anschließenden Codierung des gesamten Prädiktors benutzt werden.

Zunächst werden jedoch die bei theoretischen Signalemodellen möglichen Prädiktionsgewinne untersucht. Dabei werden Zusammenhänge aufgezeigt, die für das genauere Verständnis der Interkanal-Prädiktion interessant sind.

3.4.1 Prädiktionsgewinn bei theoretischen Signalmodellen

In diesem Abschnitt werden die bei verschiedenen theoretischen Signalmodellen möglichen Prädiktionsgewinne anhand von vier Beispielen untersucht. Dabei werden keine Quantisierungseffekte berücksichtigt.

Beispiel 1: Im einfachsten Fall wird angenommen, das sich die Signale $x(n)$ und $y(n)$ aus einem Ausgangssignal $s(n)$ durch Multiplikation mit den Faktoren A_x und A_y ergeben. Es gilt also

$$x(n) = A_x s(n) , \quad (3.87)$$

$$y(n) = A_y s(n) . \quad (3.88)$$

Es soll nun ein Prädiktor mit dem Grad $p = 0$ und der Verzögerung $d = 0$ verwendet werden. Mit der Signalleistung

$$\sigma_s^2 = E[s^2(n)] \quad (3.89)$$

gilt für die Leistungen und Korrelationskoeffizienten nach (3.4), (3.14), (3.15) und (3.16)

$$\sigma_{x0}^2 = \sigma_x^2 = A_x^2 \sigma_s^2 , \quad (3.90)$$

$$\sigma_y^2 = A_y^2 \sigma_s^2 , \quad (3.91)$$

$$r_{0,0} = 1 , \quad (3.92)$$

$$c_0 = \frac{A_x A_y \sigma_s^2}{|A_x| |A_y| \sigma_s^2} = \text{sgn}(A_x A_y) = \pm 1 . \quad (3.93)$$

Mit (3.22), (3.23) und (3.24) folgt daraus

$$a_0 = \frac{|A_Y| \sigma_S}{|A_X| \sigma_S} \operatorname{sgn}(A_X A_Y) = \frac{A_Y}{A_X} , \quad (3.94)$$

$$\sigma_e^2 = A_Y^2 \sigma_S^2 (1 - (\pm 1)^2) = 0 , \quad (3.95)$$

$$G = \frac{1}{0} = \infty . \quad (3.96)$$

Es ist also, wie zu erwarten war, eine exakte Prädiktion mit $\sigma_e^2 = 0$ möglich. Der Prädiktorkoeffizient a_0 läßt sich mit der Annahme $y'(n) = y(n)$ auch direkt aus der Prädiktorgleichung (3.1) bestimmen. Aus

$$y(n) = y'(n) = a_0 x(n) \quad (3.97)$$

folgt der gleiche Wert für den Koeffizienten a_0 wie in

$$a_0 = \frac{y(n)}{x(n)} = \frac{A_Y s(n)}{A_X s(n)} = \frac{A_Y}{A_X} . \quad (3.98)$$

Dieses recht triviale Beispiel zeigt, daß eine exakte Prädiktion erreicht wird, wenn $x(n)$ und $x(n)$ linear voneinander abhängen. Dieses ist für Mono-Signale der Fall, die an beliebiger Stelle im Amplitudenstereo-Panorama angeordnet sein können. Wenn $|A_X| > 0$ und $|A_Y| > 0$ gilt, so sind beide Prädiktionsrichtungen gleichwertig.

Beispiel 2: Werden nun den beiden Kanälen zwei statistisch unabhängige weiße Rauschsignale $r_X(n)$ und $r_Y(n)$ mit den Leistungen σ_{rX}^2 und σ_{rY}^2 überlagert, so ergibt sich

$$x(n) = A_X s(n) + r_X(n) , \quad (3.99)$$

$$y(n) = A_Y s(n) + r_Y(n) . \quad (3.100)$$

Die Verhältnisse von Signal- zu Rauschleistung bei $x(n)$ und $y(n)$ erhalten die Bezeichnung SR_X und SR_Y :

$$SR_X = \frac{A_X^2 \sigma_S^2}{\sigma_{rX}^2} \quad (3.101)$$

$$SR_Y = \frac{A_Y^2 \sigma_S^2}{\sigma_{rY}^2} \quad (3.102)$$

Für den Kreuzkorrelationskoeffizienten gilt damit

$$c_0 = \frac{1}{((1 + SR_x^{-1})(1 + SR_y^{-1}))^{1/2}} \operatorname{sgn}(A_x A_y) . \quad (3.103)$$

Der Prädiktorkoeffizient lautet jetzt

$$a_0 = \frac{A_y}{A_x} \frac{1}{1 + SR_x^{-1}} . \quad (3.104)$$

Er ergibt sich aus dem a_0 in Gleichung (3.98) durch Multiplikation mit einem Faktor, der vom Verhältnis SR_x der Signal- und Rauschleistung von $x(n)$ abhängt und Werte von 0 bis 1 annehmen kann. Der Betrag von a_0 wird hier also kleiner als in Gleichung (3.98).

Für den Prädiktionsgewinn gilt

$$G = \frac{1}{1 - \frac{1}{(1 + SR_x^{-1})(1 + SR_y^{-1})}} . \quad (3.105)$$

Da die Signal-/Rauschleistungs-Verhältnisse SR_x und SR_y von $x(n)$ und $y(n)$ in gleicher Weise in den Prädiktionsgewinn eingehen, ist auch in diesem Fall der Prädiktionsgewinn von der Prädiktionsrichtung unabhängig.

Beispiel 3: Es wird jetzt angenommen, daß die Signale $x(n)$ und $y(n)$ Sinusschwingungen der gleichen Frequenz w aber unterschiedlicher Phase und Amplitude enthalten. Es gilt

$$x(n) = A_x \cos(w n) , \quad (3.106)$$

$$y(n) = A_y \cos(w n + \alpha) . \quad (3.107)$$

Es soll ein Prädiktor mit zwei Koeffizienten (also $p = 1$) und einer Verzögerung von $d = 0$ verwendet werden. Für die verschiedenen Leistungen folgt

$$\sigma_{x0}^2 = \sigma_{x1}^2 = 1/2 A_x^2 , \quad (3.108)$$

$$\sigma_y^2 = 1/2 A_y^2 . \quad (3.109)$$

Für die Korrelationskoeffizienten ergibt sich

$$r_{0,1} = \cos(w) , \quad (3.110)$$

$$c_0 = \cos(\alpha) , \quad (3.111)$$

$$c_1 = \cos(\alpha + w) . \quad (3.112)$$

Damit lassen sich die Prädiktorkoeffizienten berechnen:

$$a_0 = \frac{A_Y}{A_X} (\cos(\alpha) + \sin(\alpha) \cotan(w)) \quad (3.113)$$

$$a_1 = - \frac{A_Y \sin(\alpha)}{A_X \sin(w)} \quad (3.114)$$

Für die Prädiktionsfehlerleistung gilt $\sigma_e^2 = 0$. Für den Prädiktionsgewinn folgt daraus $G = \infty$. Es ist also eine exakte Prädiktion mit $y(n) = y'(n)$ möglich. Daher lassen sich die Prädiktorkoeffizienten auch durch Einsetzen von (3.106) und (3.107) in die Prädiktorgleichung (3.1) berechnen. Für $|A_X| > 0$ und $|A_Y| > 0$ sind hier auch wieder beide Prädiktionsrichtungen gleichwertig.

Für $w = 0$ (konstanter Signalwert) und $w = \pi$ (Shannon-Grenze) bekommt die Determinante der Matrix $[r_{k,i}]$ in (3.18) den Wert 0:

$$\det[r_{k,i}] = 1 - \cos^2(w) = 0 \quad \text{für } w = 0 \text{ und } w = \pi \quad (3.115)$$

Das lineare Gleichungssystem (3.18) ist nun nicht eindeutig lösbar, da in diesen beiden Fällen ein Prädiktor mit einem Koeffizienten (also $p = 0$) ausreichen würde.

In der Praxis treten solche einfachen Sinusschwingungen zum Beispiel auf, wenn in ein Teilband nur eine Spektrallinie eines Tones fällt. Setzen sich die Signale $x(n)$ und $y(n)$ aus mehrere Sinusschwingungen zusammen, werden entsprechend komplexere Prädiktoren mit mehr Koeffizienten benötigt. Die Prädiktion ermöglicht es jedoch nicht, aus einer Sinusschwingung mit einer Frequenz w_x eine Sinusschwingung mit einer anderen Frequenz w_y vorherzusagen. Dieses läßt sich gut mit der Darstellung des Prädiktors als Filter erläutern. Mit einem linearen zeitinvarianten Filter läßt sich nur die Phase und Amplitude einer Spektrallinie verändern, nicht jedoch ihre Frequenz.

Beispiel 4: Bei ersten Überlegungen im Rahmen dieser Arbeit wurde angenommen, das die beiden Kanäle eines Stereosignals zum Beispiel durch Zeitverschiebung und Verstärkung voneinander abgeleitet werden können. Davon ausgehend wurde für die Untersuchungen ein Prädiktor mit vorgeschalteter Verzögerung um d Abtastschritte verwendet. Wenn zwischen den Signalen $x(n)$ und $y(n)$ der Zusammenhang

$$y(n) = a_0 x(n-d) \quad (3.116)$$

besteht, ist sofort sichtbar, daß für eine exakte Prädiktion ein Prädiktor des Grades $p = 0$ mit dem Koeffizienten a_0 und der Verzögerung d ausreicht. Der Wert der Verzögerung d darf jedoch aufgrund von (3.86) nicht negativ sein. Anderenfalls ist die umgekehrte Prädiktionsrichtung zu verwenden. Hieraus wird deutlich, warum in dieser Arbeit bei der Interkanal-Prädiktion keine feste Prädiktionsrichtung benutzt wird.

Ist $y(n)$ aus $x(n)$ durch eine nicht ganzzahlige Verzögerung d hervorgegangen, so ermöglicht ein Prädiktor mit $p > 0$ die Interpolation mehrerer Abtastwerte des Signals $x(n)$ zu einem Vorhersagewert $y'(n)$. Auch in diesem Fall bietet die variable Verzögerung d Vorteile. Zwischen den Signalen $x(n)$ und $y(n)$ bestehe zum Beispiel folgender Zusammenhang:

$$y(n) = 0,7 x(n-3) + 0,7 x(n-4) \quad (3.117)$$

Bei einer Verzögerung von $d = 0$ ist für die exakte Prädiktion ein Prädiktorgrad von $p \geq 4$ notwendig. Für die Prädiktorkoeffizienten gilt

$$a_0 = 0, \quad a_1 = 0, \quad a_2 = 0, \quad a_3 = 0,7, \quad a_4 = 0,7, \quad a_5 = 0, \quad \dots \quad (3.118)$$

Es sind also mindestens 5 Prädiktorkoeffizienten zu codieren, von denen mindestens 3 den Wert 0 haben. Ist eine zusätzliche Verzögerung d möglich, so ist ein Prädiktor des Grades $p = 1$ mit der Verzögerung $d = 3$ und den Prädiktorkoeffizienten

$$a_0 = 0,7, \quad a_1 = 0,7 \quad (3.119)$$

ausreichend. Jetzt brauchen nur 2 Prädiktorkoeffizienten codiert zu werden. Zusätzlich muß aber auch die gewählte Verzögerung d übertragen werden. Dieses Beispiel zeigt, daß die variable Verzögerung d auch bei der notwendigen Codierung des gewählten Prädiktors Vorteile bieten kann.

3.4.2 Prädiktionsgewinn bei realen Tonsignalen

Um die mit der Interkanal-Prädiktion im praktischen Einsatz erzielbaren Prädiktionsgewinne abschätzen zu können, wurden verschiedene Tonsignale untersucht. Diese Tonsignale, mit denen später auch der modifizierte Coder getestet wurde, lagen in digitaler Form als Datei vor. Bei einer Abtastrate von 48 kHz sind die Sample für den linken und rechten Kanal mit jeweils 16 Bit gleichförmig quantisiert. Die verwendeten Tonsignale sind im Anhang verzeichnet (Anhang A.4).

Für die Berechnung der in einem Teilband erzielbaren Prädiktionsgewinne wurde ein Programm in der Programmiersprache C geschrieben. Die zu untersuchenden Tonsignale standen diesem Programm in einer teilbandgefilterten und abtastratenreduzierten Fassung zur Verfügung. Für diese Teilbandzerlegung war die Polyphasenfilterbank von Layer II verwendet worden.

Dem Programm zur Berechnung der Prädiktionsgewinne sind eine Reihe verschiedener Parameter vorzugeben. So sind unter anderem das zu untersuchende Teilband SB , die Blockgröße N , der Prädiktorgrad p und der für die Verzögerung zugelassene Wertebereich $d_{\min} \leq d \leq d_{\max}$ zu wählen. Außerdem muß angegeben werden, welche Prädiktionsrichtungen p_{dir} erlaubt sein sollen. Die berechneten Gewinne können dann in tabellarischer und graphischer Form dargestellt werden. Außerdem können die jeweils ermittelten optimalen Prädiktoren ausgegeben werden. Aus Gründen der Übersichtlichkeit und des besseren Verständnis werden die Prädiktionsgewinne jeweils in dB angegeben. Dabei gilt folgender Zusammenhang:

$$G_{dB} = 10 \log_{10}(G) \quad (3.120)$$

Mit diesem Programm wurden jeweils mehrere Teilbänder verschiedener Tonsignale untersucht. Die graphischen Darstellungen von wichtigen dabei ermittelten Ergebnissen sind im Anhang wiedergegeben (Anhang A.5).

Um in einem Teilband den optimalen Prädiktor für einen Block von N Teilbandsample pro Kanal zu bestimmen, werden zunächst in der in Gleichung (3.29) beschriebenen Weise die benötigten Leistungen gemäß (3.4) und (3.14) sowie die benötigten Korrelationskoeffizienten gemäß (3.15) und (3.16) berechnet. Anschließend werden die Prädiktorkoeffizienten a_k bestimmt, in dem das lineare Gleichungssystem (3.18) mit Hilfe des Cholesky-Verfahrens /4/ gelöst wird. Falls dieses nicht möglich ist, weil die Matrix $[r_{k,i}]$ nicht positiv definit ist oder weil numerische Probleme auftreten, so werden diese Fehler abgefangen und gegebenenfalls

gemeldet. Da solche Fehler nur bei extrem leisen Signalen oder bei nicht eindeutig lösba- ren Gleichungssystemen (Prädiktorgrad p zu groß) auftreten, ergeben sich daraus keine weiteren Probleme. Mit Gleichung (3.21) wird nun der erzielte Prä- diktionsgewinn berechnet.

Wenn keine feste Prädiktionsrichtung p_{dir} vorgegeben wurde oder wenn mehrere Werte für die Verzögerung d zugelassen wurden (also $d_{min} < d_{max}$), so wird für alle damit möglichen Kombinationen von d und p_{dir} jeweils der optimale Prädiktor und der erzielte Prädiktionsgewinn bestimmt. Anschließend wird der größte dieser Prädiktionsgewinne gesucht und so die optimale Kombination von Verzögerung d und Prädiktionsrichtung p_{dir} ermittelt. Dieses Verfahren ist zwar relativ rechen- aufwendig, stellt aber sicher, daß wirklich der unter den gegebenen Randbedingun- gen beste Prädiktor mit dem optimalen Prädiktionsgewinn G_{opt} ermittelt wird.

Einige der so berechneten Prädiktionsgewinne G_{opt} sind im Anhang dargestellt (Anhang A.5.1). Dabei wurde, wie auch bei allen weiteren Diagrammen, jeweils die optimale der beiden möglichen Prädiktionsrichtungen gewählt. Die Verzöge- rung wurde entweder fest auf $d = 0$ gesetzt oder es wurde die optimale Verzöge- rung d mit $d_{min}=0 \leq d \leq d_{max}=16$ verwendet. Es wurde immer eine Block- länge von $N = 36$ Teilbandsample (24 ms) verwendet. Die Diagramme zeigen, daß das Zulassen einer variablen Verzögerung d teilweise deutlich höhere Gewinne ermöglicht. Außerdem ist zu sehen, daß die erzielbaren Prädiktionsgewinne mit wachsendem Prädiktorgrad zunehmen, die variable Verzögerung dann aber nur noch geringere Vorteile bietet. Dieses läßt sich mit den im vorhergehenden Ab- schnitt an "Beispiel 4" aufgezeigten Zusammenhängen erklären.

Die erzielbaren Prädiktionsgewinne hängen stark vom Signalcharakter ab. So sind beim Tonsignal "Stravinsky" nur selten mehr als 5 dB Gewinn möglich. Beim Ton- signal "Klarinette" sind teilweise Prädiktionsgewinne von über 40 dB erreichbar. Die manchmal direkt am Anfang oder Ende der eigentlichen Tonsignale auftreten- den "Ausreißer" hängen mit der Quantisierung der Tonsignale auf 16 Bit sowie mit dem "Digital Null" enthaltenden Vor- und Nachspann zusammen.

Bei den Tonsignalen "Cembalo" und "Klarinette" ist ein interessantes Phänomen zu erkennen: Der von Prädiktoren mit einem niedrigen Prädiktorgrad p erzielbare Prädiktionsgewinn nimmt zum Ende des Signals hin deutlich zu! Dieses ist damit zu erklären, daß in beiden Fällen auf den Musikinstrumenten eine aufsteigende

Tonfolge gespielt wurde. Beim "Cembalo", einem extrem obertonreichen Instrument, liegen die Grundfrequenzen zwischen etwa 100 Hz (erster Ton) und etwa 700 Hz (letzter Ton). Mit zunehmender Tonhöhe steigt auch der Frequenzabstand der harmonischen Obertöne entsprechend. Daher fallen mit zunehmender Tonhöhe auch immer weniger dieser Spektrallinien in ein 750 Hz breites Teilband. Und wie im vorhergehenden Abschnitt in "Beispiel 3" gezeigt wurde, reicht zum Beispiel bei genau einer Spektrallinie in einem Teilband ein Prädiktorgrad $p = 1$ für die exakte Prädiktion aus.

Um später ein geeignetes Verfahren zur Codierung der verwendeten Prädiktoren entwerfen zu können, sind mit dem Programm verschiedene statistische Untersuchungen der ermittelten optimalen Prädiktoren und Prädiktionsgewinne möglich. So wurde untersucht, in welcher Weise der über die Dauer eines Tonsignals gemittelte Prädiktionsgewinn vom für die Verzögerung zugelassenen Wertebereich $d_{\min} \leq d \leq d_{\max}$ und vom Prädiktorgrad p abhängt. Im Anhang sind graphische Darstellungen dieser Zusammenhänge wiedergegeben (Anhang A.5.2, A.5.3, A.5.4). Sie geben die oben erläuterten Abhängigkeiten des Prädiktionsgewinns von d , d_{\max} und p gut wieder. Die Diagramme, in denen der mittlere Prädiktionsgewinn als Funktion des Prädiktorgrades p dargestellt ist, zeigen, daß es nicht sehr sinnvoll ist, komplexere Prädiktoren als solche mit maximal etwa 7 bis 10 Koeffizienten ($p = 6 \dots p = 9$) einzusetzen. Bei einer Quantisierung der einzelnen Koeffizienten mit jeweils 6 Bit ist pro Koeffizient nämlich etwa 1 dB Seiteninformation vom erzielten Prädiktionsgewinn abzuziehen. Dabei wird davon ausgegangen, daß ein Prädiktionsgewinn von 6,02 dB dazu führt, daß bei der Quantisierung der Teilbandsample 1 Bit pro Sample eingespart werden kann. Bei einer Blocklänge von $N = 36$ entspricht ein Bit pro Block an Seiteninformation daher $6,02 \text{ dB} / 36 = 0,167 \text{ dB}$.

Beim Tonsignal "Suzanne Vega", das starken Mono-Charakter hat, ist zu erkennen, daß statistische Abhängigkeiten fast nur zwischen zeitgleichen Teilbandsamples bestehen (siehe Anhang A.5.2, A.5.3). Eine variable Verzögerung d bietet hier also fast keine Vorteile.

Der bei den drei anderen Tonsignalen bei niedrigen Prädiktorgraden p auftretende Zusammenhang zwischen dem mittleren Prädiktionsgewinn und der maximal zugelassenen Verzögerung d_{\max} läßt sich auch begründen: Dazu wird davon ausgegangen, daß die erzielbaren Prädiktionsgewinne unabhängig von der Verzögerung

d seien und zwischen 0 und einem maximalen Gewinn G_{\max} gleichverteilt seien. Je mehr Werte der Verzögerung d nun bei der Ermittlung des maximalen Gewinns (also des optimalen Prädiktors) zugelassen sind, um so höher wird der Erwartungswert für G_{opt} . Daraus ergibt sich folgender Zusammenhang:

$$E[G_{\text{opt}}] = G_{\max} \frac{d_{\max} + 1}{d_{\max} + 2} \quad (3.121)$$

Er ist in Bild 3.6 graphisch dargestellt.

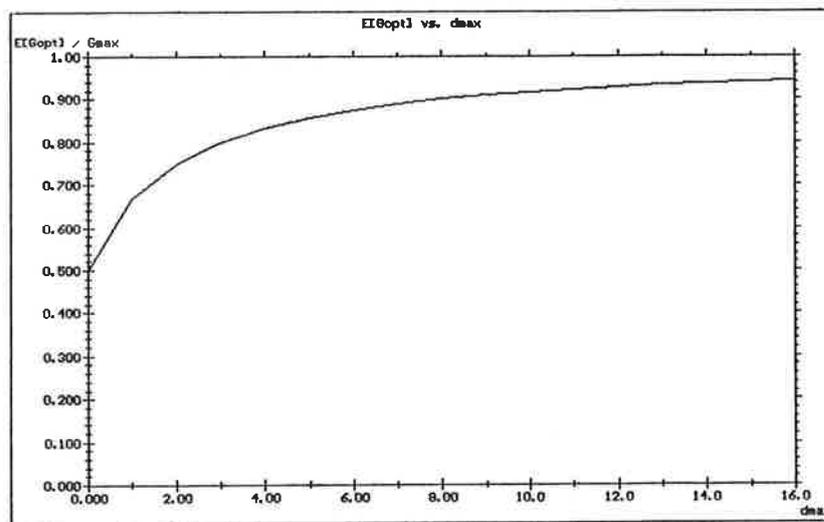


Bild 3.6 Abhängigkeit des Prädiktionsgewinns $E[G_{\text{opt}}]$ von der maximal zugelassenen Verzögerung d_{\max}

Für die Codierung der verwendeten Prädiktoren ist eine Quantisierung ihrer Koeffizienten notwendig. Um ein geeignetes Quantisierungsverfahren zu entwickeln, wurden mit dem hier beschriebenen Programm Histogramme über die Verteilung der Werte für die Prädiktorkoeffizienten a_k ermittelt. Einige der so berechneten Wahrscheinlichkeitsdichten sind exemplarisch im Anhang wiedergegeben (Anhang A.5.5).

Beim Tonsignal "Klarinette" ist zu erkennen, daß der Betrag der Prädiktorkoeffizienten a_k selten deutlich größer als 1 wird. Obwohl die Verzögerung $d = 0$ benutzt wurde, weichen die Wahrscheinlichkeitsdichten der unterschiedlichen Koeffizienten a_k eines Prädiktors mit $p > 0$ nicht sehr stark voneinander ab. Beim Tonsignal "Suzanne Vega" ist zu erkennen, daß der Koeffizient a_0 sehr häufig einen

Wert um 1 annimmt. Dieses liegt daran, daß zeitgleiche Teilbandsample bei diesem Tonsignal sehr ähnliche Werte besitzen, das Tonsignal also Mono-Charakter hat. Wenn der Prädiktorgrad $p = 1$ gewählt wird, so nimmt der zweite Koeffizient a_1 nur Werte in der Nähe von 0 an. Aufgrund eines Pegelunterschiedes zwischen den beiden Kanäle und der nicht festgelegten Prädiktionsrichtung ergeben sich hier bei der Wahrscheinlichkeitsdichte für a_0 zwei dicht nebeneinander liegende Maxima.

Da später bei der Quantisierung der Teilbandsample im Coder nicht die Leistungen der Signale sondern ihre Spitzenwerte berücksichtigt werden müssen, wird für den nach dem oben beschriebenen Verfahren ermittelten Prädiktor zusätzlich der Gewinn G_{peak} als Verhältnis der Spitzenwertquadrate der Signale $y(n)$ und $e(n)$ berechnet:

$$G_{\text{peak}} = \frac{\max(|y(n)|^2)}{\max(|e(n)|^2)} \quad (3.122)$$

$\max(f(n))$ gibt dabei den größten Wert von $f(n)$ im aktuellen Block von N Abtastwerten an. Um G_{peak} zu bestimmen, müssen für den gewählten Prädiktor alle N Werte von $e(n)$ im aktuellen Block berechnet werden. G_{opt} und G_{peak} hängen über die Crest-Faktoren der Signale $y(n)$ und $e(n)$ zusammen. Der Crest-Faktor gibt das Verhältnis von Spitzenwert zu Effektivwert eines Signals an. Sind die Crest-Faktoren von $y(n)$ und $e(n)$ gleich, so gilt $G_{\text{peak}} = G_{\text{opt}}$.

Mit dem hier beschriebenen Programm wurde so auch die erzielbaren Prädiktionsgewinne G_{peak} ermittelt. Einige der so ermittelten Ergebnisse sind im Anhang graphisch dargestellt (Anhang A.5.6). Es zeigt sich, daß die Gewinne G_{peak} im Mittel etwas kleiner sind als die Gewinne G_{opt} . Außerdem variieren die Gewinne G_{peak} von Block zu Block stärker als die Gewinne G_{opt} .

Da die Crest-Faktoren stark von den jeweiligen Kurvenformen und internen Phasenbeziehungen der Signale abhängen, konnten sie im Rahmen dieser Arbeit nicht theoretisch untersucht werden. Auch eine Optimierung der Prädiktoren zur Maximierung des Gewinns G_{peak} aus (3.122) konnte im Rahmen dieser Arbeit nicht behandelt werden.

Das hier beschriebene Programm ermöglicht auch, die Auswirkungen der Quantisierung der Prädiktorkoeffizienten a_k auf den erzielbaren Prädiktionsgewinn G_{peak}

zu untersuchen. Das bei quantisierten Prädiktorkoeffizienten a_{qk} entstehende Fehlersignal wird mit $e_q(n)$ bezeichnet. Für den Prädiktionsgewinn G_{quant} gilt

$$G_{\text{quant}} = \frac{\max(|y(n)|^2)}{\max(|e_q(n)|^2)}. \quad (3.123)$$

Dabei wird weiterhin von den unquantisierten Signalen $x(n)$ und $y(n)$ ausgegangen. Die Berücksichtigung der bei der Quantisierung der Teilbandsample auftretenden Effekte erfolgte erst im Rahmen der Modifikation des Audio-Codecs für Layer II.

Zur Quantisierung der Prädiktorkoeffizienten wird ein gleichförmiger Quantisierer verwendet. Dazu wird das vorzugebene Intervall $[-a_{\text{max}}, a_{\text{max}}]$ in eine vorzugebene Anzahl von gleichgroßen Bereichen eingeteilt. Die Repräsentativwerte liegen jeweils in der Mitte dieser Bereiche. Bei den durchgeführten Untersuchungen wurde zum Beispiel häufig das Intervall $[-2,2]$ in 64 Quantisierungsstufen unterteilt.

Einige der so ermittelten Prädiktionsgewinne G_{quant} sind im Anhang graphisch dargestellt (Anhang A.5.6). Die Prädiktorkoeffizienten a_k wurden dabei im Intervall $[-2,2]$ mit 64 möglichen Stufen quantisiert. Dieses entspricht einer Seiteninformation von 6 Bit = 1 dB pro Koeffizient. Zusätzlich ist ein Bit Seiteninformation für die gewählte Prädiktionsrichtung und gegebenenfalls etwa 4 Bit für die variable Verzögerung zu berücksichtigen. Für den Prädiktograd $p = 3$ und $d_{\text{max}} = 16$ sind somit etwa 29 Bit = 4,8 dB Seiteninformation zu berücksichtigen. Diese Werte sind auch in den Diagrammen eingezeichnet. Außerdem ist zu erkennen, daß die Quantisierung der Prädiktorkoeffizienten zur Folge hat, daß die Prädiktionsgewinne G_{quant} einen gewissen Maximalwert nur selten übersteigen. Für die hier verwendete Quantisierungsstufenbreite $\delta = 1/16$ und den Prädiktograd $p = 7$ ergibt sich aus der Tabelle in Bild 3.5 ein mittlerer maximaler Prädiktionsgewinn von 25,8 dB. Wie im Anhang zu erkennen ist, stimmt die Größenordnung dieses theoretischen Wertes relativ gut mit den maximal ermittelten Prädiktionsgewinnen G_{quant} überein.

3.4.3 Verfahren zur Codierung der Prädiktorparameter

Von diesen Untersuchungsergebnissen ausgehend wurden ein Verfahren zu Codierung der Prädiktorparameter entworfen. Damit ein für die jeweiligen Signaleigenschaften optimaler Prädiktor verwendet werden kann, müssen Prädiktoren mit un-

terschiedlichem Prädiktorgrad p und variabler Verzögerung d codiert werden können. Außerdem müssen die Prädiktorkoeffizienten in geeigneter Weise quantisiert werden. Das im Rahmen dieser Arbeit entwickelte Codierungsverfahren ist im Anhang dargestellt (Anhang A.3).

Die im vorhergehenden Abschnitt beschriebenen Untersuchungen haben gezeigt, daß es nur bei Prädiktoren mit einem niedrigen Prädiktorgrad p deutliche Vorteile bietet, eine variable Verzögerung d zuzulassen. Daher ist sie im Codierungsverfahren nur für Prädiktoren mit $p \leq 3$ berücksichtigt. Es besteht aber auch bei diesen einfachen Prädiktoren die Möglichkeit, direkt die feste Verzögerung $d = 0$ zu benutzen. Außerdem hat sich gezeigt, daß es sich nicht lohnt, sehr komplexe Prädiktoren zu verwenden, da dabei die Prädiktorkoeffizienten zu einer entsprechend großen Menge an Seiteninformation führen. So ermöglicht das entwickelte Verfahren nur die Codierung von Prädiktoren mit maximal 8 Koeffizienten ($p \leq 7$). Für die Verzögerung d sind maximal die nach Ungleichung (3.86) erlaubten Werte zugelassen.

Die zunächst untersuchte gleichförmige Quantisierung der Prädiktorkoeffizienten im Intervall $[-2,2]$ mit 6 Bit, also 64 Stufen, hat sich als nicht optimal herausgestellt. Zum einen ist sie so grob, daß der bei komplexen Prädiktoren erzielbare Gewinn deutlich begrenzt wird. Außerdem treten auch häufiger Prädiktorkoeffizienten mit außerhalb des Intervalles $[-2,2]$ liegenden Werten auf, so daß diese Prädiktoren dann nicht mehr sinnvoll codiert werden können. Eine feinere Quantisierung in einem größeren Intervall würde die Anzahl der für die Seiteninformation benötigten Bits stark erhöhen.

Da die einzelnen Koeffizienten eines Prädiktors in der Regel in der gleichen Größenordnung liegen, diese Größenordnung jedoch vom Verhältnis der Pegel der beiden Signale $x(n)$ und $y(n)$ abhängt, wurde ein Skalierungsfaktor für die Koeffizienten eines Prädiktors eingeführt. Dieser Skalierungsfaktor, der mit 3 Bit codiert wird, gibt das Intervall an, in dem die Prädiktorkoeffizienten weiterhin gleichförmig mit 64 Stufen quantisiert werden. Das kleinste dabei mögliche Intervall ist $[-1/4, 1/4]$, das größte ist $[-8,8]$. Dieses Verfahren ist speziell bei Prädiktoren mit vielen Koeffizienten sinnvoller als eine zunächst in Erwägung gezogene ungleichförmige Quantisierung der Prädiktorkoeffizienten.

Für Prädiktoren mit einem Koeffizienten a_0 und keiner zusätzlichen Verzögerung, also $p = 0$ und $d = 0$, wurde zusätzlich ein sehr grobes ungleichförmiges Quantisierungsverfahren für a_0 festgelegt, das nur 3 Bit für die Codierung benötigt. Im Falle eines Mono-Signals mit $x(n) = y(n)$, also $a_0 = 1$, ist so nur sehr wenig Seiteninformation für die Übertragung des Prädiktors notwendig.

Zusammen mit dem Fall, daß keine Prädiktion verwendet wird, ergeben sich insgesamt 14 verschiedene Modi für die Codierung der Prädiktoren. Diese Prädiktormodi p_{mode} werden mit einem Präfixcode /8/ codiert. Die Länge der Codewörter liegt zwischen 2 Bit (bei den einfachsten Prädiktoren) und 7 Bit (bei den komplexen Prädiktoren). Anschließend an das Codewort für den Prädiktormodus p_{mode} werden die jeweils notwendigen weiteren Prädiktorkenngrößen übertragen. Diese sind die gewählte Prädiktionsrichtung p_{dir} (1 Bit), die gewählte Verzögerung d (4 Bit), der Skalierungsfaktor sf für die Prädiktorkoeffizienten (3 Bit) sowie die Prädiktorkoeffizienten a_k selbst (3 Bit oder 6 Bit pro Koeffizient).

3.4.4 Prädiktionsgewinn bei Berücksichtigung der Seiteninformation

Um den resultierenden Prädiktionsgewinn bei der Interkanal-Prädiktion zu bestimmen, ist vom Prädiktionsgewinn G_{quant} , der bei der Teilbandsamplequantisierung ausgenutzt werden kann, die für die Codierung des verwendeten Prädiktors benötigte Seiteninformation SI (in dB dargestellt) abzuziehen. Für den so ermittelten verbleibenden Prädiktionsgewinn G_{remain} gilt:

$$G_{remain,dB} = G_{quant,dB} - SI_{dB} \quad (3.124)$$

Bei der Umrechnung der Seiteninformation in dB wird angenommen, daß bei einem Prädiktionsgewinn von $G_{quant,dB} = 6,02$ dB bei der Quantisierung des Signals $e(n)$ ein Bit pro Teilbandsample weniger benötigt wird als bei der Quantisierung von $y(n)$ notwendig wäre. Bei einer Blocklänge von $N = 36$ entspricht 1 Bit Seiteninformation also $6,02 \text{ dB} / 36 = 0,167$ dB. Die bei den verschiedenen Modi p_{mode} für die Codierung der Prädiktoren notwendige Seiteninformation SI ist im Anhang angegeben (Anhang A.3).

Um den für einen Block eines Teilbands optimalen Prädiktor zu bestimmen, der unter Berücksichtigung der Seiteninformation den größten Prädiktionsgewinn G_{remain} ermöglicht, wird in folgender Weise vorgegangen: Zunächst wird für alle Kombinationen aus den beiden Prädiktionsrichtung p_{dir} und 13 Prädiktormodi

p_{mode} der optimale Prädiktor und gegebenenfalls auch die optimale Verzögerung d ermittelt. Dazu wird - wie auch schon weiter oben beschrieben - die Verzögerung d gewählt, die bei gegebener Prädiktionsrichtung p_{dir} und Prädiktorgrad p den größten Prädiktionsgewinn G_{opt} ermöglicht. Für die insgesamt $13 * 2 = 26$ so ermittelten Prädiktoren wird nun die Quantisierung der Prädiktorkoeffizienten durchgeführt. Anschließend werden mit den quantisierten Koeffizienten die Prädiktionsgewinne G_{quant} ermittelt. Mit Gleichung (3.124) können jetzt die verbleibenden Prädiktionsgewinne G_{remain} berechnet werden. Es wird dann der größte Gewinn G_{remain} gesucht und so der unter Berücksichtigung der Seiteninformation optimale Prädiktor ermittelt. Ist mit keinem der Prädiktoren ein verbleibender Gewinn $G_{remain,dB} > 0$ dB möglich, so kommt kein Prädiktor zum Einsatz ($p_{mode} = 0$).

Das oben beschriebene Programm zur Ermittlung der erzielbaren Prädiktionsgewinne wurde nun um dieses Verfahren zur Bestimmung der optimalen Prädiktoren unter Berücksichtigung der Seiteninformationen erweitert. Einige der damit gewonnenen Ergebnisse sind im Anhang graphisch dargestellt (Anhang A.6).

Die Diagramme, in denen der verbleibende Prädiktionsgewinn G_{remain} über den Verlauf eines Tonsignals dargestellt ist (Anhang A.6.1), zeigen, daß in einigen Teilbändern der Tonsignale "Klarinette" und "Suzanne Vega" resultierende Prädiktionsgewinne im Bereich von 30 dB bis 40 dB möglich sind. Der ebenfalls dargestellte Umfang der Seiteninformation SI zeigt, daß bei den größeren Prädiktionsgewinnen in der Regel auch komplexere Prädiktoren verwendet wurden. Die Tatsache, daß abhängig von den Signaleigenschaften alle insgesamt 14 möglichen Prädiktormodi p_{mode} auch zum Einsatz kommen, läßt den Schluß zu, daß es sinnvoll ist, diese sehr verschiedenen Modi zuzulassen.

Im Anhang sind auch Diagramme wiedergegeben, die den über den Signalverlauf gemittelten verbleibenden Prädiktionsgewinn G_{remain} in dB in Abhängigkeit vom Teilband SB darstellen (Anhang A.6.2). Sie zeigen, daß die Interkanal-Prädiktion hauptsächlich in den ersten Teilbändern deutliche Gewinne ermöglicht. Dieses ist nicht ungünstig, da auch ein großer Teil der Signalleistung in den unteren Teilbändern liegt und beim Audio-Codec für Layer II ein großer Teil zur Verfügung stehenden Bits für die Codierung der Sample dieser Teilbänder eingesetzt wird. Nur beim Tonsignal "Stravinsky" nimmt der mittlere Prädiktionsgewinn zu den höheren Teilbändern hin zu. Dieses kann damit erklärt werden, daß nur wenige Instrumente

bei dieser Orchesteraufnahme Obertöne erzeugen, die in die oberen Teilbänder fallen.

4 Modifizierter Audio-Codec mit Interkanal-Prädiktion

Um die Eigenschaften und den erzielbaren Prädiktionsgewinn der Interkanal-Prädiktion in der praktischen Anwendung untersuchen zu können, wurde der ISO Audio-Codec für Layer II entsprechend modifiziert. Dabei konnte von einer Implementation dieses Codecs als C-Programm ausgegangen werden.

Zunächst wurde die Definition des Bitstroms zwischen Coder und Decoder um einen Abschnitt ergänzt, der die Codierung der verwendeten Prädiktoren ermöglicht. Anschließend wurden der Coder und der Decoder um die Interkanal-Prädiktion erweitert. Bei der Modifikation des Coders haben sich verschiedene, teilweise prinzipielle Probleme ergeben, die gelöst werden mußten. Die im Laufe der Entwicklung des erweiterten Codecs erzielte Tonqualität der codierten Signale wurde nach objektiven und subjektiven Kriterien beurteilt.

4.1 Struktur des modifizierten Audio-Codecs

Durch die Erweiterung um die Interkanal-Prädiktion hat sich die grundlegende Struktur des Codecs nur an einigen Stellen verändert. So ist im Blockschaltbild für den Coder (Bild 2.3) der Block "Normierung, Quantisierung, Codierung" um die Aufgaben "Prädiktoren ermitteln" und "Prädiktion" zu ergänzen. Entsprechend ist im Blockschaltbild für den Decoder (Bild 2.4) der Block "Decodierung, Dequantisierung, Skalierung" um die "Prädiktion" zu erweitern.

Die Polyphasenfilterbank und das psychoakustische Modell wurden nicht geändert. Auch die Normierung, Quantisierung, und Codierung der Teilbandsample (siehe Abschnitt 2.2 sowie Anhang A.1 und A.2) wurde beibehalten. Wenn in einem Teilband ein Prädiktor zum Einsatz kommt, werden jedoch nicht mehr die Originalsignale $x(n)$ und $y(n)$ sondern nur $x(n)$ und das Prädiktionsfehlersignal $e_q(n)$ quantisiert und codiert. Dieses ist in Bild 3.4 dargestellt.

Die modifizierte Bitzuteilung muß berücksichtigen, daß für das Signal $e_q(n)$ nur ein um den Prädiktionsgewinn $G_{\text{quant,dB}}$ kleineres SMR (Signal/Mask-Ratio) benötigt wird, als beim Signal $y(n)$ notwendig wäre. Außerdem sind die für die Codierung des Prädiktors benötigten Bits zu berücksichtigen. Um bei der Bestimmung des optimalen Prädiktors und des Prädiktionsgewinns $G_{\text{quant,dB}}$ für ein Teilband die Auswirkungen der Quantisierung des Originalsignals $x(n)$

berücksichtigen zu können, muß jedoch der verwendete Quantisierer und somit auch die gewählte Bitzuteilung schon bekannt sein. Für eine möglichst optimale Prädiktion ist es nämlich sinnvoll, den Prädiktor für das schon quantisierte Signal $x_q(n)$ zu berechnen.

Dieser ein wenig an den "Hauptmann von Köpenick" erinnernde Widerspruch stellt das größte Problem bei der Modifikation des Audio-Coders dar. Er kann mit zwei unterschiedlichen Ansätzen angegangen werden:

Iterative Bitzuteilung: Für die Bitzuteilung kann ein iteratives Verfahren verwendet werden. Dabei werden in einer Schleife zunächst für die aktuelle Bitzuteilung die optimalen Prädiktoren und Prädiktionsgewinne berechnet. Im zweiten Teil der Schleife wird dann mit den neu berechneten Prädiktionsgewinnen eine aktualisierte Bitzuteilung bestimmt. Als Startwert kann für die erste Ermittlung der Prädiktoren zum Beispiel von unquantisierten Signalen ausgegangen werden. Die Iterative Bitzuteilung hat jedoch den Nachteil, daß sie relativ viel Rechenaufwand erfordert, da die optimalen Prädiktoren mehrfach berechnet werden müssen. Außerdem müssen für diese "rückgekoppelte" Schleife Stabilitätsbetrachtungen angestellt werden und es muß ein geeignetes Schleifenabbruch-Kriterium verwendet werden. Der Ablauf der iterativen Bitzuteilung ist in Bild 4.1 dargestellt.

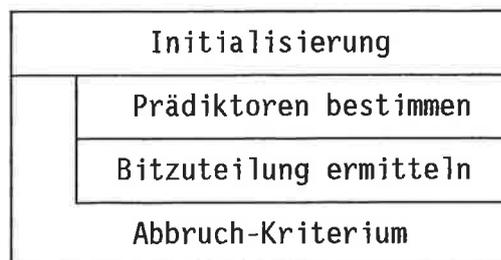


Bild 4.1 *Ablauf der iterativen Bitzuteilung bei Layer II mit Interkanal-Prädiktion*

Approximierte Bitzuteilung: Anstelle der iterativen Bitzuteilung wurde im Rahmen dieser Arbeit ein anderes einfacheres Verfahren für die Bitzuteilung entwickelt, das eine gute Annäherung an die optimale Bitzuteilung ermöglicht. Es läßt sich als Abfolge eines halben, eines ganzen und eines halben Schleifendurchlaufes der iterativen Bitzuteilung beschreiben. Der Ablauf dieser approximierten Bitzuteilung ist in Bild 4.2 dargestellt.

Bitzuteilung schätzen (vorl.)
Prädiktoren bestimmen (vorl.)
Bitzuteilung ermitteln
ggf. Prädiktorkoeffizienten neu bestimmen

Bild 4.2 *Ablauf der approximierten Bitzuteilung bei Layer II mit Interkanal-Prädiktion*

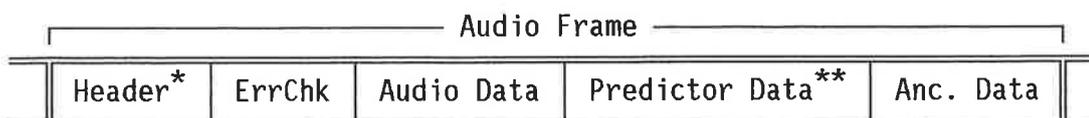
Bei der Schätzung der Bitzuteilung wird davon ausgegangen, daß die Summe der verbleibenden Prädiktionsgewinne $G_{\text{remain,dB}}$ im aktuellen Block den gleichen Wert wie im vergangenen Block haben wird. Anschließend werden von der geschätzten Bitzuteilung ausgehend die optimalen Prädiktoren und Prädiktionsgewinne ermittelt. Dabei werden dann die Prädiktormodi und die Prädiktionsrichtungen festgelegt. Mit diesen Prädiktionsgewinnen wird nun die endgültige Bitzuteilung durchgeführt. In den Fällen, wo sie von der geschätzten Bitzuteilung abweicht, werden für den zugehörigen Prädiktor die Prädiktorkoeffizienten neu berechnet. Die Auswahl eines anderen Prädiktormodus oder einer anderen Prädiktionsrichtung ist jedoch nicht mehr möglich, da die Bitzuteilung ja schon festgelegt ist.

Aufgrund der bei der iterativen Bitzuteilung zu erwartenden Probleme (Stabilität, Rechenaufwand) konnte im Rahmen dieser Arbeit nur die approximierte Bitzuteilung entwickelt und implementiert werden.

In den folgenden Abschnitten wird der modifizierte Audio-Codec für Layer II mit Interkanal-Prädiktion genauer beschrieben.

4.1.1 Struktur des erweiterten Bitstroms

Um auch die verwendeten Prädiktoren im Bitstrom codieren und übertragen zu können, wurde der in Bild 2.5 dargestellte Bitstrom von Layer II um einen Abschnitt (Predictor Data) erweitert. Dieser Abschnitt wurde zwischen die Tondaten (Audio Data) und die möglichen Zusatzdaten (Ancillary Data) eingefügt. Der Aufbau des so modifizierten Bitstroms ist in Bild 4.3 dargestellt.



(* = modifiziert / ** = neu)
(ErrChk = Error Check / Anc. Data = Ancillary Data)

Bild 4.3 Aufbau des erweiterten Bitstroms bei Layer II mit Interkanal-Prädiktion

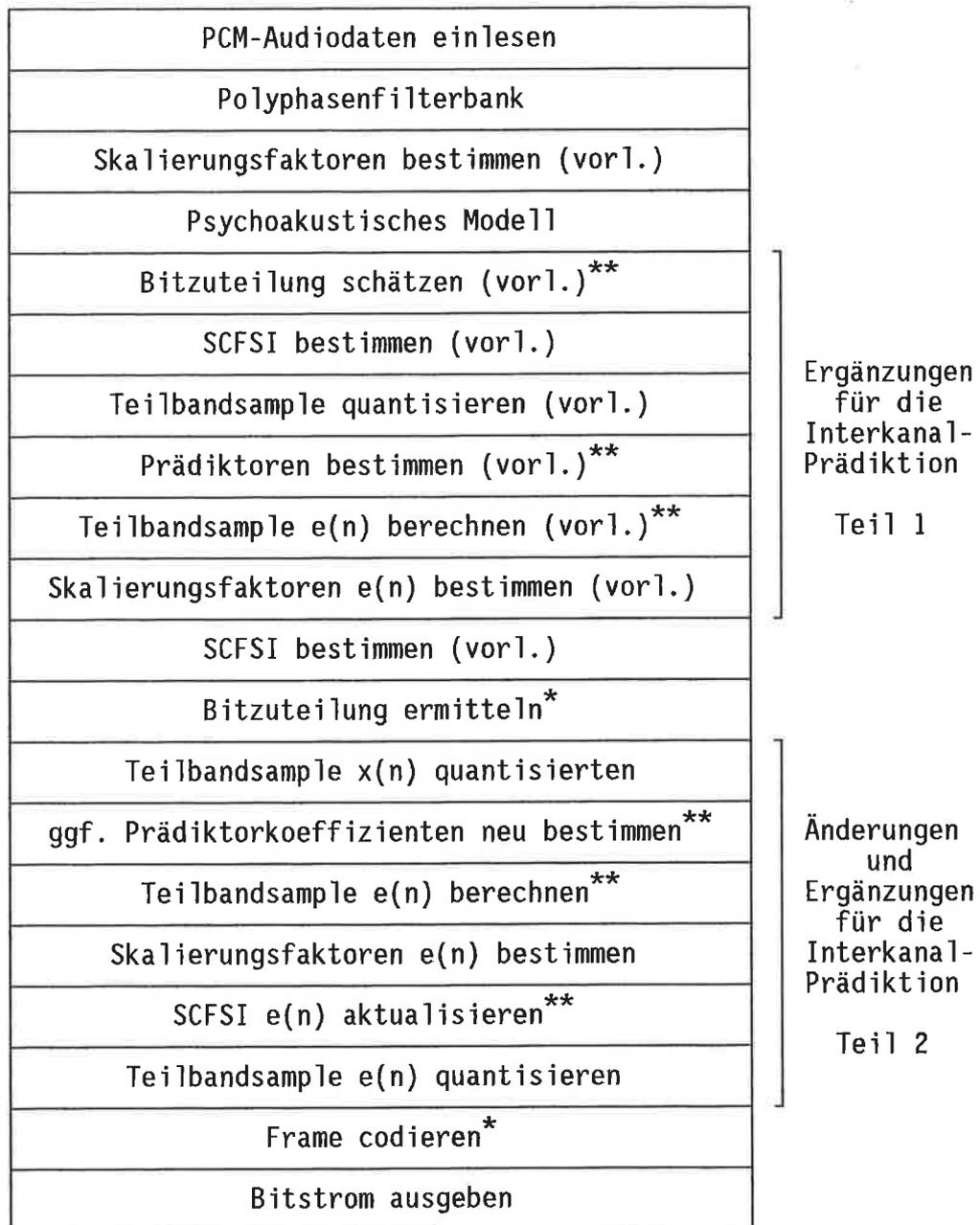
Die einzelnen Prädiktoren für die Teilbänder werden jeweils mit dem schon weiter oben beschriebenen Verfahren codiert (Abschnitt 3.4.3 und Anhang A.3). Sie werden aber nur für die Teilbänder im Bitstrom übertragen, in denen mindestens einer der beiden Stereo-Kanäle überhaupt eine Bitzuteilung erhalten hat. Denn in einem Teilband, in dem noch nicht einmal das Signal $x(n)$ übertragen wird, bringt in der Praxis auch eine Prädiktion mit einer Verzögerung $d > 0$ keinen Gewinn.

Das im Header für anwenderdefinierbare Aufgaben zur Verfügung stehende "Private Bit" wird dazu verwendet, den Einsatz der Interkanal-Prädiktion zu kennzeichnen. Ist es nicht gesetzt, so wurde keine Interkanal-Prädiktion verwendet. Wird ein mit Interkanal-Prädiktion erzeugter Bitstrom von einem dem ISO-Standard entsprechenden Decoder entschlüsselt, so wertet dieser den Zustand des "Private Bit" nicht aus. Er rekonstruiert daher jeweils statt der Teilbandsignale $x(n)$ und $y(n)$ die Signale $x(n)$ und $e(n)$. Der Abschnitt "Predictor Data" wird dabei als Teil des Blocks "Ancillary Data" betrachtet und nicht ausgewertet.

Können bei der Übertragung des Bitstroms Fehler auftreten, so ergibt sich beim hier beschriebenen erweiterten Bitstrom das Problem, das der Abschnitt "Predictor Data" nicht vom Fehlerkorrekturwort "Error Check" geschützt werden kann. Tritt in einem den Prädiktormodus $pmode$ angehenden Codewort ein Bitfehler auf, so kann auch der Rest des codierten Prädiktors nicht mehr richtig decodiert werden, da die Anzahl der für die Codierung eines Prädiktors verwendeten Bits vom Modus $pmode$ abhängt. Auch die fehlerfreie Decodierung der noch folgenden Prädiktoren ist nicht mehr möglich. Um dieses Problem zu lösen, müßte zumindest für die Bits, die für die Codierung der Modi $pmode$ verwendet werden, eine Fehlererkennung oder -korrektur möglich sein.

4.1.2 Struktur des modifizierten Audio-Coders

Der Ablauf der modifizierten Codierung ist in Bild 4.4 dargestellt.



(* = modifiziert / ** = neu)
 ("Teilbandsample quantisieren" bedeutet hier:
 Teilbandsample normieren, quantisieren, dequantisieren & skalieren)

**Bild 4.4 Ablauf der modifizierten Codierung bei Layer II
 mit Interkanal-Prädiktion**

Bei der Modifikation des Audio-Coders für Layer II wurde das Verfahren der approximierten Bitzuteilung verwendet. Dazu wurde der in Bild 2.6 dargestellte Ablauf um zwei große Blöcke erweitert. Diese Änderungen und Ergänzungen sind in Bild 4.4 mit "Teil 1" und "Teil 2" bezeichnet. Der "Teil 2" beinhaltet dabei die endgültige Normierung und Quantisierung der Teilbandsample. Außerdem wurden natürlich die Ermittlung der Bitzuteilung und die Codierung des Frames modifiziert.

Der Ablauf der Codierung wird nun im einzelnen beschrieben. Die Teile des Coders, deren Funktionsweise sich gegenüber der in Abschnitt 2.2.2 beschriebenen nicht geändert hat, werden hier nicht noch einmal detailliert erklärt.

Polyphasenfilterbank: Die Polyphasenfilterung mit anschließender Abtastratenreduktion wird ohne Modifikationen für die beiden Stereokanäle durchgeführt. Es werden weiterhin 36 Teilbandsample je Teilband und Kanal in einem Frame codiert.

Skalierungsfaktoren: Für die in 12er-Blöcken zusammengefaßten Teilbandsample der beiden Kanäle werden wie bisher nun die Skalierungsfaktoren bestimmt. Für die Prädiktionsfehlersignale $e(n)$ müssen jedoch später erneut die Skalierungsfaktoren ermittelt werden.

Psychoakustisches Modell: Das unmodifizierte psychoakustische Modell berechnet jetzt die SMR-Werte (Signal/Mask-Ratio) für die Teilbänder der beiden Kanäle.

Schätzung der Bitzuteilung: Um die Prädiktoren für die quantisierten Signale $x(n)$ berechnen zu können, ist eine Schätzung der Bitzuteilung notwendig. Dazu werden die im letzten Block erzielten SNR-Werte (Signal/Noise-Ratio) aller Teilbänder der beiden Kanäle aufsummiert. Der SNR-Wert für ein Teilbandsignal $y(n)$ berechnet sich dabei als Summe des SNR-Wertes des für das Prädiktionsfehlersignal $e(n)$ verwendeten Quantisierers und des Prädiktionsgewinns $G_{\text{quant,dB}}$. Grundlegende Annahme bei der folgenden Schätzung der Bitzuteilung ist nun, daß im aktuellen Block die gleiche Summe der SNR-Werte erreicht wird wie im letzten Block. Dieses bedeutet, daß davon ausgegangen wird, daß sich die Summe der verbleibenden Prädiktionsgewinne $G_{\text{remain,dB}}$ über die Teilbänder nicht ändert, da

bei konstanter Bitrate mit den Teilbandquantisierern etwa eine konstante SNR-Summe erzielt wird.

Die eigentliche Schätzung der Bitzuteilung erfolgt nun in einer der "richtigen" Bitzuteilung sehr ähnlichen Weise. Zunächst werden allen Teilbändern der beiden Kanäle keine Bits zugeteilt. Dieses bedeutet einen SNR-Wert von 0 dB. Anschließend wird mit Hilfe der vom psychoakustischen Modell schon berechneten SMR-Werte das Teilband eines Kanals herausgesucht, das den größten NMR-Wert (Noise/Mask-Ratio) hat. Für dieses Teilband wird dann gemäß der entsprechenden Tabelle (Anhang A.1) die nächstgrößere mögliche Bitzuteilung gewählt. Dieser Vorgang wird so lange wiederholt, bis die Summe der SNR-Werte der geschätzten Bitzuteilung gerade nicht mehr kleiner ist als die SNR-Summe des letzten Blocks. Die so geschätzte Bitzuteilung gibt nun an, welche Quantisierer für die Teilbänder der Kanäle voraussichtlich verwendet werden, wenn sie als vollständiges Signal $x(n)$ und nicht als Prädiktionsfehlersignal $e(n)$ übertragen werden.

Es hat sich gezeigt, daß bei diesem Schätzverfahren in etwa 80% bis 90% aller Fälle die geschätzte Bitzuteilung mit der später verwendeten Bitzuteilung übereinstimmt. Abweichungen um mehr als eine Bitzuteilungsstufe sind sehr selten (<1%).

SCFSI: Aus den schon berechneten Skalierungsfaktoren wird nun wie bisher die Scalefactor Select Information (SCFSI) ermittelt. Für die Prädiktionsfehlersignale $e(n)$ muß die SCFSI später jedoch erneut berechnet werden.

Quantisierung: Mit Hilfe der SCFSI und der Skalierungsfaktoren werden nun alle Teilbandsignale normiert. Anschließend werden sie gemäß der geschätzten Bitzuteilung quantisiert. Um die Prädiktoren für die schon quantisierten Signale $x_q(n)$ ermitteln zu können, werden die Signale zusätzlich wieder dequantisiert und skaliert, so daß sie in der gleichen Form wie später im Decoder vorliegen. Die verwendeten Normierungs- und Quantisierungsverfahren wurden nicht modifiziert.

Ermittlung der Prädiktoren: Das Verfahren für die Ermittlung der optimalen Prädiktoren wurde schon in Abschnitt 3.4.4 erläutert. Da alle Teilbänder der beiden Kanäle nun sowohl in quantisierter als auch in unquantisierte Form zur Verfügung stehen, können die optimalen Prädiktoren jetzt aber jeweils für die Prädiktion vom quantisierten Signal $x_q(n)$ auf das unquantisierte Signal $y(n)$ ermittelt werden.

Erst anschließend werden der Prädiktormodus p_{mode} und die Prädiktionsrichtung p_{dir} gewählt, die den höchsten verbleibenden Prädiktionsgewinn $G_{\text{remain,dB}}$ in einem Teilband ermöglichen. Auf diese Weise werden automatisch die in Abschnitt 3.3.2 erläuterten Auswirkungen der Quantisierung von $x(n)$ auf den Prädiktionsgewinn berücksichtigt. Dieses ist speziell dann wichtig, wenn zum Beispiel eine grobe Quantisierung des Signals $x(n)$ den ohne Signalquantisierung erzielten hohen Prädiktionsgewinn stark reduziert. Die Auswirkungen der Quantisierung der Prädiktorkoeffizienten werden in der in Abschnitt 3.4.4 beschriebenen Weise berücksichtigt.

Um nicht unnötig viele Bits für die Codierung der Prädiktorparameter zu benutzen, ist es sinnvoll, keinen Prädiktor mit einem so hohen Gewinn $G_{\text{quant,dB}}$ zu verwenden, daß selbst dann, wenn das Prädiktionsfehlersignal $e_q(n)$ nicht quantisiert und codiert wird, beim Teilbandsignal $y(n)$ ein deutlich besserer NMR-Wert erzielt wird als bei den anderen codierten Teilbandsignalen. Als Grenze für den NMR-Wert wird der Mittelwert der NMR-Werte der im letzten Block codierten Teilbandsignale benutzt. Treten bei der Ermittlung der optimalen Prädiktoren Gewinne $G_{\text{quant,dB}}$ auf, die größer sind, als die Differenz zwischen dem SMR-Wert des Teilbandsignals $y(n)$ und obigem NMR-Mittelwert, so wird der Prädiktor mit dem kleinsten Modus p_{mode} gewählt, der diese Eigenschaft hat.

Prädiktionsrichtungen, bei denen die Prädiktion von einem Signal $x(n)$ ausgehen würde, dem bei der Schätzung der Bitzuteilung keine Bits zugewiesen wurden (also $x_q(n) = 0$), werden bei der Ermittlung der optimalen Prädiktoren explizit ausgeschlossen. Ist in einem Teilband der Einsatz eines Prädiktors nicht sinnvoll, da kein verbleibender Prädiktionsgewinn $G_{\text{remain,dB}} > 0$ dB erzielt werden kann oder da beide Prädiktionsrichtungen explizit ausgeschlossen wurden, so wird der Prädiktormodus in diesem Teilband auf $p_{\text{mode}} = 0$ (keine Prädiktion) gesetzt.

Berechnung von $e(n)$: Mit den so ermittelten Prädiktoren werden nun ausgehend von den quantisierten Signalen $x_q(n)$ zunächst die Prädiktionssignale $y_q'(n)$ berechnet. Zusammen mit den unquantisierten Signalen $y(n)$ können nun die Prädiktionsfehlersignale $e_q(n)$ bestimmt werden. Falls die geschätzte Bitzuteilung von der später verwendeten abweicht, ist dann eine erneute Berechnung von $e_q(n)$ erforderlich. Wenn in einem Teilband keine Prädiktion zum Einsatz kommt ($p_{\text{mode}} = 0$), so wird weiter das Signal $y(n)$ verwendet.

Skalierungsfaktoren für $e(n)$: Für die eben berechneten $e_q(n)$ werden nun die Skalierungsfaktoren ermittelt. Das verwendete Verfahren wurde nicht modifiziert.

SCFSI für $e(n)$: Es wird nun die Scalefactor Select Information für die neu berechneten Skalierungsfaktoren von $e_q(n)$ bestimmt. Die SCFSI für die Signale $x(n)$ ändert sich dabei nicht.

Bitzuteilung: Es folgt nun die endgültige Festlegung der Bitzuteilung für das aktuelle Frame. Dazu muß der bisher verwendete Algorithmus (siehe Abschnitt 2.2.2) an mehreren Stellen modifiziert werden.

Zunächst werden die vom psychoakustischen Modell berechneten SMR-Werte für die Teilbänder, in denen an Stelle des Signals $y(n)$ das Prädiktionsfehlersignal $e_q(n)$ übertragen werden soll, um den jeweiligen Prädiktionsgewinn $G_{\text{quant,dB}}$ reduziert.

Die Erweiterung des Bitstroms um die codierten Prädiktoren muß berücksichtigt werden. Dazu wird als Startwert davon ausgegangen, daß keine Prädiktoren übertragen werden. Werden nun in einem Teilband irgendeinem der beiden Kanäle das erste Mal Bits zugeteilt, so muß entsprechend auch der für dieses Teilband verwendete Prädiktor codiert werden. Es daher auch die Anzahl der für die Codierung des Prädiktors notwendigen Bits (siehe Anhang A.3) zu berücksichtigen.

Bei der Bitzuteilung wird zunächst sichergestellt, daß die Teilbandsignale $x(n)$, die später einem Prädiktor als Ausgangssignal dienen, auf jeden Fall übertragen werden. Dazu wird für diese Teilbandsignale zunächst der jeweils einfachste größte Quantisierer ausgewählt. Anschließend wird - solange noch Bits zur Verfügung stehen - in der Schleife für die Bitzuteilung weiterhin bei dem Teilband eines Kanals mit dem größten NMR-Wert der jeweils nächstfeinere Quantisierer ausgewählt.

Quantisierung von $x(n)$: Mit den bei der Bitzuteilung endgültig festgelegten Quantisierern werden die Teilbandsignale $x(n)$ nun normiert, quantisiert und anschließend wieder dequantisiert und deskaliert. So stehen dem Coder jetzt genau die quantisierten Signale $x_q(n)$ zur Verfügung, die auch im Decoder rekonstruiert werden. Falls die geschätzte Bitzuteilung mit der inzwischen endgültig festgelegten

Bitzuteilung übereinstimmt, waren genau diese $x_q(n)$ ja auch schon weiter oben berechnet worden.

Prädiktorkoeffizienten: Für den Fall, daß die für ein Teilbandsignal $x(n)$ geschätzte Bitzuteilung nicht mit der nun endgültig festgelegten Bitzuteilung übereinstimmt, ist es sinnvoll, den in diesem Teilband verwendeten Prädiktor für das jetzt anders quantisierte Signal $x_q(n)$ neu zu berechnen. Aufgrund der schon festgelegten Bitzuteilung müssen dabei jedoch der Prädiktormodus p_{mode} und die Prädiktionsrichtung p_{dir} beibehalten werden. So können nur noch die Prädiktorkoeffizienten und gegebenenfalls die variable Verschiebung d neu bestimmt werden. Abgesehen von diesen Einschränkungen wird dabei in der gleichen Weise wie bei der "Ermittlung der Prädiktoren" vorgegangen.

Berechnung von $e(n)$: Für die nun endgültig quantisierten Signale $x_q(n)$ und die endgültig festgelegten Prädiktoren werden jetzt die Vorhersagesignale $y_q'(n)$ berechnet, die später auch im Dekoder erzeugt werden. Mit den $y_q'(n)$ und den Signalen $y(n)$ werden die Prädiktionsfehlersignale $e_q(n)$ bestimmt. Falls in einem Teilband die geschätzte Bitzuteilung mit der endgültig festgelegten Bitzuteilung übereinstimmt, waren die $e_q(n)$ dieses Teilbands auch schon weiter oben richtig berechnet worden.

Skalierungsfaktoren für $e(n)$: Für das nun ebenfalls endgültig festliegende Prädiktionsfehlersignal $e_q(n)$ werden jetzt die Skalierungsfaktoren bestimmt.

SCFSI für $e(n)$: Bei der Bestimmung der Scalefactor Select Information (SCFSI) für die Signale $e_q(n)$ ergibt sich das Problem, daß die Bitzuteilung schon festgelegt ist. Damit steht auch schon fest, ob in einem Teilband 1, 2 oder 3 Skalierungsfaktoren für das Signal $e_q(n)$ übertragen werden. Wenn 3 Skalierungsfaktoren codiert werden, sind keine weiteren Überlegungen notwendig. Wird nur ein Skalierungsfaktor codiert, so muß dieses natürlich der größte der insgesamt 3 Faktoren sein. Wenn 2 Skalierungsfaktoren übertragen werden, so kann jetzt noch festgelegt werden, ob der erste oder der zweite der beiden Skalierungsfaktoren später für zwei 12er-Blöcke verwendet wird. Dazu werden, ähnlich wie bei dem sonst verwendeten Verfahren zur Bestimmung der SCFSI, die Differenzen zweier aufeinanderfolgender Skalierungsfaktoren ausgewertet.

Stimmen in einem Teilband die geschätzte und die richtige Bitzuteilung für das Signal $x(n)$ überein, so sind die Skalierungsfaktoren und die SCFSI für das Prädiktionsfehlersignal $e_q(n)$ ja auch schon weiter oben richtig berechnet worden. Bei der SCFSI ergibt sich jedoch das prinzipielle Problem, daß die Hüllkurven der Signale $y(n)$ und $e(n)$ nicht unbedingt ähnlich sein müssen. Auf die mit der psychoakustischen Nachverdeckung begründeten Teile des für die Bestimmung der SCFSI benutzten Verfahrens dürfte also eventuell nicht zurückgegriffen werden.

Quantisierung von $e(n)$: Mit den jetzt endgültig festgelegten SCFSI und den Skalierungsfaktoren können nun auch die Prädiktionsfehlersignale $e_q(n)$ normiert und quantisiert werden.

Codierung: Die dann folgende Codierung eines Frames des Bitstroms wurde nur an einer Stelle modifiziert. So werden entsprechend des im vorhergehenden Abschnitt beschriebenen erweiterten Bitstroms nach den Tondaten (Audio Data) die benutzten Prädiktorparameter codiert. Dabei wird das im Anhang beschriebene Verfahren benutzt (Anhang A.3).

4.1.3 Struktur des modifizierten Audio-Decoders

Der Ablauf der modifizierten Decodierung ist in Bild 4.5 dargestellt.

Bitstrom einlesen
Frame decodieren*
Skalierungsfaktoren decodieren
Teilbandsample dequantisieren
Teilbandsample skalieren
Prädiktoren decodieren**
Teilbandsample $y(n)$ rekonstruieren**
Polyphasenfilterbank
PCM-Audiodaten ausgeben

(* = modifiziert / ** = neu)

Bild 4.5 Ablauf der modifizierten Decodierung bei Layer II mit Interkanal-Prädiktion

Der Ablauf der Decodierung wird nun im einzelnen beschrieben. Die in Abschnitt 2.2.3 vorgestellte Funktionsweise des Decoders ändert sich dabei durch das Hinzufügen der Interkanal-Prädiktion nur an wenigen Stellen.

Decodierung: Die Decodierung eines Frames wird zunächst wie bisher durchgeführt. Nachdem die Tondaten (Audio Data) vollständig decodiert wurden, können nun auch die verwendeten Prädiktorparameter entschlüsselt werden. Dabei geht aus der schon decodierten Bitzuteilung hervor, für welche Teilbänder überhaupt ein Prädiktor übertragen wurde.

Dequantisierung: Zunächst werden die Teilbandsample dequantisiert. Anschließend werden die SCFSI und die Skalierungsfaktoren für die Skalierung der Teilbandsample verwendet. Diese Dequantisierung und Skalierung wurde nicht modifiziert.

Prädiktion: Mit Hilfe der decodierten Prädiktoren können nun in den Teilbändern, in denen bei der Codierung die Interkanal-Prädiktion zum Einsatz kam, aus den decodierten Signalen $x_q(n)$ die Prädiktionssignale $y_q'(n)$ berechnet werden. Zusammen mit den ebenfalls decodierten Prädiktionsfehlersignalen $e_{qq}(n)$ können daraus die Signale $y_{qq}(n)$ berechnet werden.

Rekonstruktion: Bei den so rekonstruierten Teilbandsignalen $x_q(n)$ und $y_{qq}(n)$ wird die Abtastrate erhöht. Mit Hilfe des Polyphasenfilters werden die Teilbandsignale abschließend zu den Ausgangssignalen für beiden Stereokanäle zusammengefügt.

4.2 Untersuchung der codierten Tonsignale

Es wurden objektive und subjektive Verfahren verwendet, um die durch den Einsatz der Interkanal-Prädiktion erzielten Änderungen der Qualität der codierten Tonsignale zu untersuchen. Dabei wurden die im Anhang A.4 angegebenen Tonsignale benutzt. Es wurde in der Regel mit einer Bitrate von 128 kbit/s für ein Stereo-Signal gearbeitet. Dieses entspricht einer Verringerung der Datenrate um dem Faktor 12. Pro Abtastwert, der im Originalsignal ja mit 16 Bit quantisiert ist, stehen bei der Übertragung also 1,33 Bit zur Verfügung. Zur Berechnung der SMR-Werte (Signal/Mask-Ratio) wurde in der Regel das zweite psychoakustische Modell verwendet, da es speziell bei der verwendeten niedrigeren Bitrate eine höhere subjektive Tonqualität bietet als das erste Modell /11/.

Die Optimierung des um die Interkanal-Prädiktion erweiterten Audio-Codecs für Layer II wurde hauptsächlich mit den Signalen "Klarinette" und "Suzanne Vega" durchgeführt, da bei diesen beiden Signalen die Änderungen in der Tonqualität auch subjektiv deutlich wahrnehmbar sind.

4.2.1 Objektive Qualität der codierten Tonsignale

Die hier angewendeten Verfahren zur Bestimmung der objektiven Qualität der codierten Tonsignale basieren auf den vom verwendeten psychoakustischen Modell berechneten SMR-Werten (Signal/Mask-Ratio). Es wird dabei angenommen, daß das psychoakustische Modell fehlerfrei funktioniert. Diese Annahme wird in Abschnitt 5.1 noch genauer betrachtet.

Die Codierung erfolgt in Blöcken mit 36 Teilbandsamples je Kanal und Teilband. Bei der hier verwendeten Abtastrate haben diese Blöcke daher eine Länge von 24 ms. Aus der in einem Block verwendeten Bitzuteilung läßt sich auf die SNR-Werte (Signal/Noise-Ratio) schließen, die bei der Quantisierung der einzelnen Teilbandsignale erreicht wurden. Dazu wird die auch bei der Bitzuteilung verwendete Tabelle benutzt (Anhang A.2). Die folgenden Betrachtungen beziehen sich immer auf einen solchen 24 ms langen Block.

Bei den Teilbandsignalen, die unter Verwendung eines Interkanal-Prädiktors codiert wurden, ist zu dem SNR-Wert des für das Prädiktionsfehlersignal $e_q(n)$ verwendeten Quantisierers noch der Prädiktionsgewinn $G_{\text{quant, dB}}$ zu addieren, um den resultierenden SNR-Wert für das rekonstruierte Signal $y_{qq}(n)$ zu erhalten. Dabei ist der Prädiktionsgewinn des nach der endgültigen Festlegung der Bitzuteilung gegebenenfalls noch aktualisierten Prädiktors zu verwenden. Wenn der Schätzwert für die Bitzuteilung des Signals $x(n)$ nicht richtig war, weicht dieser Gewinn in der Regel von dem bei der Bitzuteilung angenommenen Gewinn ab.

Wird nun vom SMR-Wert der SNR-Wert abgezogen, so ergibt sich der NMR-Wert (Noise/Mask-Ratio). Er gibt an, wie weit bei diesem Teilbandsignal das Quantisierungsgeräusch über der Maskierungsschwelle liegt. Ist der NMR-Wert negativ, so wird das Quantisierungsgeräusch in diesem Teilbandsignal vollständig maskiert und ist daher (zumindest in der Regel) nicht hörbar.

Der verwendete Bitzuteilungsalgorithmus führt in der Regel dazu, daß sich die NMR-Werte zweier Teilbandsignale nur selten um mehr als 5 dB bis 9 dB unterscheiden - dem Abstand der SNR-Werte zweier "benachbarter" Bitzuteilungsstufen (siehe Anhang A.1 und A.2). Um eine übersichtliche Darstellung zu ermöglichen, ist es daher sinnvoll, die SMR-, SNR- und NMR-Werte aller Teilbänder der beiden Kanäle zu addieren. Bei dieser Summation werden negative Werte nicht berücksichtigt, da ein Teilbandsignal mit negativem SMR-Wert nicht codiert zu werden braucht, ein negativer NMR-Wert die Maskierung des Quantisierungsgeräusches anzeigt und ein negativer SNR-Wert nicht auftritt.

Die SMR-Summe gibt nun an, welche Signalqualität bei der Codierung erzielt werden muß, damit die durch den Codierungsprozeß hinzugefügten Störgeräusche gerade nicht mehr hörbar sind und somit eine transparente Codierung möglich ist. Die SNR-Summe gibt die bei der Codierung erzielte Signalqualität an. Und die

NMR-Summe gibt an, wie stark die hinzugefügten Störgeräusche hörbar sind. Hat die NMR-Summe den Wert 0, so sind die Quantisierungsgeräusche in allen Teilbändern beider Kanäle maskiert. Es wurde also eine transparente Codierung erreicht. Diese Aussagen gehen natürlich von einer fehlerfreien Funktion des psychoakustischen Modells aus.

Die Summe der SMR-Werte kann relativ einfach in die für eine transparente Codierung notwendige Bitrate umgerechnet werden, wenn die Seiteninformation mit Bitzuteilung und Skalierungsfaktoren vernachlässigt wird und angenommen wird, daß pro Bit, das für die Teilbandsamplequantisierung verwendet wird, der SNR-Wert eines Teilbandsignals um etwa 6 dB größer wird. Eine SNR-Summe von 6 dB entspricht so einer Bitrate von 1,5 kbit/s, da bei 48 kHz Abtastrate die Teilbandsignale mit $48 \text{ kHz} / 32 = 1,5 \text{ kHz}$ abgetastet werden. Eine Bitrate von 1 kbit/s entspricht einer SNR-Summe von 4 dB.

Bei der hier verwendeten Bitrate von 128 kbit/s ergibt sich also eine SNR-Summe von 512 dB. In der Praxis werden aufgrund der Seiteninformation mit Bitzuteilung und Skalierungsfaktoren jedoch nur SNR-Summen von etwa 400 bis 450 dB erreicht. Hat das codierte Signal eine kleinere SMR-Summe, so ist davon auszugehen, daß die Bitzuteilung eine transparente Codierung ermöglicht.

Die durch den Einsatz der Interkanal-Prädiktion erzielten Prädiktionsgewinne lassen sich auch in dieser Form darstellen. Die Summe der in den einzelnen Teilbändern erzielten Prädiktionsgewinne $G_{\text{quant,dB}}$ wird hier als PG-Summe (Prediction Gain) bezeichnet. Die Summe der verbleibenden Prädiktionsgewinne $G_{\text{remain,dB}}$ wird entsprechend RG-Summe (Remaining Gain) genannt. Mit Hilfe der RG-Summe läßt sich der durch die Interkanal-Prädiktion erzielt Codierungsgewinn in Relation zu der benötigten Signalqualität (SMR-Summe) und zur bisher erzielten Signalqualität (SNR-Summe) setzen. Mit der Entsprechung $4 \text{ dB} = 1 \text{ kbit/s}$ läßt sich der durch die RG-Summe beschriebene Prädiktionsgewinn auch in eine Bitrate umrechnen. Außerdem gibt die RG-Summe an, um welchen Wert sich die SNR-Summe beim Einsatz der Interkanal-Prädiktion etwa erhöhen wird.

Die jeweils bei der Codierung der verschiedenen Tonsignale mit und ohne Interkanal-Prädiktion erzielte objektive Signalqualität ist im Anhang graphisch dargestellt (Anhang A.7). Dabei wurde immer eine Bitrate von 128 kbit/s sowie das zweite psychoakustische Modell verwendet.

Im Anhang A.7.1 sind für alle fünf untersuchten Tonsignale (siehe Anhang A.4) jeweils zwei Diagramme dargestellt. Das erste zeigt die benötigte Signalqualität (SMR-Summe) sowie die Stärke der hörbaren Störgeräusche (NMR-Summe) bei Codierung ohne und mit Interkanal-Prädiktion. Das zweite Diagramm zeigt die erzielte Signalqualität (SNR-Summe) mit und ohne Interkanalprädiktion. Außerdem ist der Prädiktionsgewinn in Form der PG-Summe und der RG-Summe eingezeichnet.

Klarinette: Beim Tonsignal "Klarinette" werden für die RG-Summe Werte von zum Teil mehr als 150 dB erreicht. Dieses entspricht einer Bitrate von fast 40 kbit/s, die durch die Interkanal-Prädiktion gewonnen wird. Der maximale und der mittlere Wert der NMR-Summe werden durch den Einsatz der Interkanal-Prädiktion um etwa ein Drittel kleiner. Besonders in den Signalabschnitten, in denen die Störsignale am stärksten sind, ermöglicht die Interkanal-Prädiktion eine deutliche Verbesserung.

Cembalo: Bei Tonsignal "Cembalo" werden verbleibende Prädiktionsgewinne von insgesamt bis zu etwa 100 dB erzielt. Diese RG-Summe entspricht rund 25 kbit/s. Aufgrund der höheren SMR-Werte und eines ungünstigeren zeitlichen Verlaufes des Prädiktionsgewinns werden durch die Interkanal-Prädiktion jedoch nur relativ geringe Verbesserungen bei den hörbaren Störgeräuschen (NMR-Summe) erzielt.

Stravinsky & Kastagnetten: Bei den Tonsignalen "Stravinsky" und "Kastagnetten" werden auch ohne Interkanal-Prädiktion bei einer Bitrate von 128 kbit/s schon sehr kleine NMR-Werte erzielt. Daher bietet die Interkanal-Prädiktion hier - trotz der teilweise recht hohen Spitzenwerte für die RG-Summe - nur relativ geringe Vorteile.

Suzanne Vega: Beim Tonsignal "Suzanne Vega" ist der durch die Interkanal-Prädiktion erzielte Gewinn sehr deutlich zu erkennen. Die RG-Summe liegt in der Regel zwischen 150 dB und 200 dB, was einer eingesparten Bitrate von rund 45 kbit/s entspricht. Werden ohne Interkanal-Prädiktion NMR-Summen von bis zu 50 dB erreicht, so ist bei Einsatz der Prädiktion die NMR-Summe (von extrem seltenen Ausnahmen abgesehen) immer 0. Die Interkanal-Prädiktion ermöglicht hier also eine theoretisch fast transparente Codierung mit 128 kbit/s.

Bei der Bestimmung der SNR-Werte werden im Coder einige Näherungen angenommen. Diese Annahmen werden im Folgenden genauer untersucht:

Da die tabellierten SNR-Werte der Quantisierer nur unter speziellen Annahmen für die Signale exakt gelten, ist auch der wirklich erzielte Abstand zwischen dem Teilbandsignal und dem bei der Codierung hinzugefügten Störgeräusch von Interesse. Er wird hier mit SQR (Signal/Quantization_Noise-Ratio) bezeichnet. Der SQR-Wert gibt das Verhältnis der Signalenergie zur Störgeräuschenergie in einem Block der Länge N an. Er berechnet sich zum Beispiel für das Signal $y(n)$ in folgender Weise:

$$SQR_{dB} = 10 \log_{10} \frac{\sum_{n=0}^{N-1} y^2(n)}{\sum_{n=0}^{N-1} (y_{qq}(n) - y(n))^2} \quad (4.1)$$

Damit die SQR-Werte schon bei der Codierung ermittelt werden können, müssen die Teilbandsignale $x_q(n)$ und $y_{qq}(n)$ auch schon im Coder rekonstruiert werden.

In Anhang A.7.2 sind die SQR-Summen im Vergleich zu den SNR-Summen für die Tonsignale "Klarinette" und "Cembalo" graphisch dargestellt. Die SNR-Werte eines Teilbandsignals sind in der Regel bis zu 6 dB größer als die SQR-Werte. Durch den Einsatz der Interkanal-Prädiktion verringert sich diese Differenz bei den prädizierten Teilbandsignalen $y(n)$ etwas. Die Ursachen für die Differenz zwischen den SNR- und den SQR-Werten liegen in den bei der Bestimmung der SNR-Werte gemachten Annahmen begründen. Dieses Problem wird in Abschnitt 5.1 genauer untersucht.

In Anhang A.7.3 wird die Codierung eines Blocks des Tonsignals "Klarinette" genauer analysiert. Dieses geschieht sowohl für den nicht modifizierten als auch für den um die Interkanal-Prädiktion erweiterten Codec. Es werden jeweils die ermittelten SMR-, SNR-, SQR-, PG- und RG-Werte für die einzelnen Teilbänder der beiden Kanäle aufgeführt. Außerdem sind dort auch die geschätzte und die dann endgültig verwendete Bitzuteilung sowie eine Reihe weitere Daten angegeben. Mit Hilfe von in dieser Form ausgegebenen Datensätzen wurde die Funktion des um die Interkanal-Prädiktion erweiterten Codecs analysiert und optimiert.

Wird bei Layer II "Joint Stereo Coding" verwendet, so wird dabei hauptsächlich die Stereo-Irrelevanz und nicht die Stereo-Redundanz ausgenutzt. Daher kann die objektive Qualität der so codierten Tonsignale nicht in einer Form dargestellt werden, die einen Vergleich mit den oben beschriebenen Ergebnissen möglich macht.

Abschließend kann festgestellt werden, daß durch den Einsatz der Interkanal-Prädiktion bei zwei der fünf untersuchten Tonsignale eine deutliche Verbesserung der objektiven Signalqualität erzielt werden kann. Bei den drei anderen Signalen ist jeweils eine leichte Verbesserung möglich.

4.2.2 Subjektive Qualität der codierten Tonsignale

Bei den im vorhergehenden Abschnitt durchgeführten objektiven Untersuchungen der Tonqualität der codierten Signale mußte von einer fehlerfreien Funktion des psychoakustischen Modells ausgegangen werden. Da ein solches Modell jedoch nicht die volle Komplexität des Wahrnehmungsvorgangs im menschlichen Gehör wiedergeben kann, wurde die bei der Codierung erzielte Tonqualität auch subjektiv untersucht. Dazu wurden die codierten Tonsignale über Kopfhörer oder Lautsprecher abgehört. Da die bei der Codierung entstehenden Störsignale bei Wiedergabe über Kopfhörer in der Regel deutlicher und differenzierbarer wahrgenommen werden können als bei Wiedergabe über Lautsprecher, wurde im Rahmen dieser Arbeit hauptsächlich die Kopfhörerwiedergabe verwendet. Die subjektiv empfundenen Unterschiede in der Tonqualität zwischen den verschiedenen Versionen der codierten Tonsignalen hängen auch etwas von den Eigenschaften des verwendeten Wiedergabesystems - wie zum Beispiel dem Frequenzgang - ab.

Vorwiegend wurden die durch den Einsatz der Interkanal-Prädiktion erzielten Veränderungen der subjektiven Tonqualität im Vergleich zum unmodifizierten Codec untersucht. Dabei wurde in den meisten Fällen eine Bitrate von 128 kbit/s und das zweite psychoakustische Modell verwendet. Außerdem wurde auch das uncodierte Originalsignal und das mit "Joint Stereo Coding" vom Layer II codierte Signal zum Vergleich herangezogen.

Die von verschiedenen Menschen empfundenen Unterschiede der Tonqualität können durchaus etwas von einander abweichen, da sie auf einer subjektiven Wahrnehmung beruhen. Leider konnten im Rahmen dieser Arbeit keine Hörtests mit einer größeren Anzahl von Testhörerinnen und Testhörern durchgeführt werden. Die subjektive Tonqualität der codierten Tonsignale mit und ohne Interkanal-Prädiktion

kann hier also nicht in Form von gemittelten Qualitätskennzahlen dargestellt werden, wie dieses bei den Hörtests im Rahmen der ISO-Standardisierung möglich war /5/. Statt dessen wird versucht, die subjektiven Qualitätsunterschiede zwischen den verschiedenen codierten Versionen der fünf verwendeten Tonsignale (siehe Anhang A.4) verbal zu beschreiben.

Klarinette: Beim Tonsignal "Klarinette" ist ein Unterschied zwischen den mit und ohne Interkanal-Prädiktion codierten Versionen deutlich hörbar. Der nicht modifizierte Codec erzeugt bei diesem Signal relativ starke rauschartige Störungen im Bereich niedriger Frequenzen. Beim modifizierten Codec sind diese Störungen nur noch sehr schwach und das Signal klingt sauberer. Außerdem hört es sich etwas heller an, da mehr hochfrequente Signalanteile übertragen werden können. Zusätzlich tritt zum Ende des Signals - also dort, wo die größten Prädiktionsgewinne erzielt werden - teilweise ein im mittleren Frequenzbereich liegendes leises Störgeräusch auf, das am ehesten mit dem Begriff "Zwitschern" umschrieben werden kann. Von allen Personen, die diese beiden Signale im Vergleich gehört haben, wurde die mit dem modifizierten Codec erzeugte Version jeweils als die mit der höheren Tonqualität bezeichnet. Für das Tonsignal "Klarinette" kann also unter den gegebenen Randbedingungen gesagt werden, daß durch den Einsatz der Interkanal-Prädiktion eine Verbesserung der subjektiven Tonqualität erzielt werden kann. Mit beiden Versionen des Codecs ist jedoch bei 128 kbit/s keine transparente Codierung dieses Tonsignals möglich. Diese Aussagen stimmen mit den im vorhergehenden Abschnitt erläuterten Ergebnissen der objektiven Untersuchungen überein.

Die Tonsignale "Klarinette" und "Cembalo" haben sich bei den ISO-Hörtests von Layer II mit 128 kbit/s jeweils als sehr kritisch herausgestellt und wurden durch die Codierung deutlich störend verfälscht /5/. Speziell aus diesem Grund ist die beim Tonsignal "Klarinette" erzielte Qualitätsverbesserung durch Interkanal-Prädiktion positiv zu bewerten.

Cembalo: Beim Tonsignal "Cembalo" sind nur geringe Unterschiede in der Tonqualität zwischen den beiden Codec-Versionen zu hören. Der modifizierte Coder erzeugt zum Ende des Signals hin etwas weniger Störungen als der Coder ohne Interkanal-Prädiktion. Daß nur geringe Tonqualitätsunterschiede hörbar sind, liegt zum Teil sicher daran, daß beide codierten Versionen dieses Tonsignals sehr deutlich gestört sind. Von allen hier untersuchten Tonsignalen wird das "Cembalo" eindeutig am schlechtesten codiert. Eine transparente Codierung mit 128 kbit/s ist

mit beiden Versionen des Codecs nicht möglich. Diese Aussagen stimmen mit den Ergebnissen der objektiven Untersuchungen überein.

Stravinsky: Das Tonsignal "Stravinsky" wird schon vom unmodifizierten Codec relativ gut codiert. Zwischen dem Original und den beiden codierten Versionen dieses Tonsignals sind leichte Unterschiede in den Höhen und bei der Räumlichkeit speziell bei Lautsprecherwiedergabe hörbar. Der Einsatz der Interkanal-Prädiktion hat bei diesem Tonsignal also ebenfalls eine leichte subjektive Qualitätsverbesserung zur Folge. Auch diese Ergebnisse weichen nicht von den objektiv erzielten ab.

Suzanne Vega: Beim Tonsignal "Suzanne Vega" ist, ähnlich wie bei der "Klarinette", eine deutliche Verbesserung der subjektiven Tonqualität zu hören, wenn der modifizierte Codec verwendet wird. Beim unmodifizierten Codec sind deutliche zwitscherartige Störgeräusche zu hören. Außerdem klingt das codierte Tonsignal etwas dumpf, da hochfrequente Anteile fehlen. Wird die Interkanal-Prädiktion bei der Codierung verwendet, so werden die Störgeräusche leiser und treten bei höheren Frequenzen als vorher auf. Auch klingt das Signal dann heller, da nun mehr höherfrequente Signalanteile übertragen werden. Daneben ergibt sich eine etwas bessere Ortung im Stereo-Panorama. Eine transparente Codierung dieses Signals ist jedoch auch bei Einsatz der Interkanal-Prädiktion unter den hier angenommenen Voraussetzungen noch nicht möglich. Die deutliche Verbesserung der subjektiven Tonqualität, die mit dem modifizierten Codec erzielt wurde, war nach den objektiven Untersuchungen zu erwarten. Danach hätte jedoch bei Einsatz der Interkanal-Prädiktion auch eine transparente Codierung möglich sein müssen.

Kastagnetten: Aufgrund seiner sehr harten Anschläge ist das Tonsignal "Kastagnetten" speziell für die Untersuchung von sogenannten Vor-Echo-Störungen geeignet. Diese Störungen können infolge der blockweisen Codierung entstehen, wenn am Ende eines Blocks ein lautes impulsartiges Signal auftritt. Sie werden in Abschnitt 5.1 noch etwas näher erläutert. Wird das Tonsignal "Kastagnetten" mit Hilfe des unmodifizierten Codecs übertragen, so treten leichte Vor-Echo-Störungen auf. Bei Einsatz der Interkanal-Prädiktion ändert sich der Charakter dieser Störungen etwas. Mögliche Ursachen hierfür werden in Abschnitt 5.2 beschrieben. Die Vor-Echo-Störungen sind jedoch nur bei der Kopfhörerwiedergabe eindeutig wahrzunehmen. Bei Wiedergabe über Lautsprecher sind diese Störungen bei beiden Codec-Versionen nicht mehr eindeutig wahrzunehmen. Von einer eindeutigen Ver-

besserung oder Verschlechterung der Tonqualität kann bei beiden Wiedergabeverfahren nicht gesprochen werden. Mit beiden Versionen des Codecs ist jedoch auch hier keine transparente Codierung möglich. Die Aussagen über die Vor-Echo-Störungen können mit den Ergebnissen der objektiven Untersuchungen nicht begründet werden. Die Ursache dafür ist, daß die Vor-Echo-Störungen von ihrer zeitlichen Größenordnung unterhalb der Blocklänge von hier 24 ms liegen. Sie können daher mit den auf einen gesamten Block bezogenen objektiven Qualitätskenngrößen, die in vorhergehenden Abschnitt verwendet wurden, nicht erfaßt werden.

Es wurde auch die bei höheren Bitraten als 128 kbit/s erzielbare subjektive Tonqualität mit und ohne Interkanal-Prädiktion untersucht. Dabei ergeben sich jedoch nur geringfügig höhere Prädiktionsgewinne, die ja mit Hilfe der RG-Summe angegeben werden können. Da gleichzeitig die erzielte Signalqualität, angegeben durch die SNR-Summe, etwa proportional zur Bitrate wächst, wird die relative Qualitätsverbesserung durch den Einsatz der Interkanal-Prädiktion geringer. Der Effekt, daß die Vorteile der Interkanal-Prädiktion bei steigender Bitrate - relativ betrachtet - geringer werden, war auch bei subjektiven Vergleichen festzustellen.

Wenn an Stelle des zweiten das erste psychoakustische Modell verwendet wurde, so ergaben sich bei gleicher Bitrate teilweise eine andere subjektive und objektive Tonqualität. Diese Effekte stehen jedoch nicht mit der Interkanal-Prädiktion in Zusammenhang. Mögliche Ursachen werden in Abschnitt 5.1 genauer betrachtet.

Wird Layer II mit "Joint Stereo Coding" benutzt, so ergibt sich bei den Tonsignalen "Klarinette" und "Cembalo" sogar eine schlechtere subjektive Qualität als beim "Independent Stereo Coding" des unmodifizierten Codecs. Beim Tonsignal "Suzanne Vega" ermöglicht "Joint Stereo Coding" eine eindeutige Verbesserung gegenüber "Independent Stereo Coding". Beim Tonsignal "Stravinsky" wird eine leichte Verbesserung erzielt. Bei diesen beiden Tonsignalen ermöglicht der Einsatz der Interkanal-Prädiktion eine ähnliche oder geringfügig bessere subjektive Tonqualität. Beim Tonsignal "Kastagnette" ergeben sich zwar auch subjektive Änderungen, wenn "Joint Stereo Coding" verwendet wird. Von eindeutigen Verbesserungen oder Verschlechterungen gegenüber "Independent Stereo Coding" oder dem Codec mit Interkanal-Prädiktion kann aber nicht gesprochen werden.

Abschließend läßt sich feststellen, daß bei zwei der fünf untersuchten Tonsignale durch den Einsatz der Interkanal-Prädiktion eine deutliche Verbesserung der subjektiven Tonqualität erzielt werden konnte. Bei diesen beiden Signalen war auch schon eine deutliche Verbesserung der objektiven Signalqualität zu erkennen. Bei zwei weiteren Tonsignalen wurde zumindest eine leichte Verbesserung der subjektiven Tonqualität erzielt. Bei einem fünften Tonsignal blieb die subjektive Tonqualität etwa gleich.

5 Ausblick auf mögliche Weiterentwicklungen

Während der Entwicklung der Interkanal-Prädiktion und besonders bei der Modifikation, Analyse und Optimierung des Audio-Codecs für Layer II haben sich eine Reihe von Überlegungen zu möglichen Änderungen und Weiterentwicklungen dieses Toncodierungsverfahrens ergeben. Sie konnten zum großen Teil im Rahmen dieser Arbeit nicht weiter untersucht werden. In den drei folgenden Abschnitten werden diese Ansätze jeweils kurz erläutert.

5.1 Optimierungsmöglichkeiten beim ISO Audio-Codec

In diesem Abschnitt werden verschiedene Probleme, die bei der Arbeit mit dem ISO Audio-Codec für Layer II aufgefallen sind, erläutert. Dabei werden jeweils Ansätze für eventuell mögliche Optimierungen aufgeführt.

Bitzuteilung: Der momentan verwendete Bitzuteilungsalgorithmus führt dazu, daß Abstand zwischen Quantisierungsgeräusch und Maskierungsschwelle in allen Teilbändern etwa gleich groß ist. Stehen aufgrund der verwendeten Bitrate nicht ausreichend Bits zur Verfügung, so sind alle überhaupt codierten Teilbandsignale subjektiv etwa gleich stark gestört. Da nicht übertragene Teilbandsignale häufig die subjektive Tonqualität weniger beeinträchtigen als zu grob quantisierte, kann es daher eventuell sinnvoll sein, die Bitzuteilung so zu modifizieren, daß zwar weniger Teilbandsignale codiert werden, in diesen aber jeweils das Quantisierungsgeräusch unterhalb der Maskierungsschwelle liegt. Diese Bitzuteilung sollte bei den Teilbändern für die niedrigen Frequenzen beginnen und würde dazu führen, daß der Codec einen Tiefpaß-Charakter mit sich ändernder Eckfrequenz hätte. Dadurch könnten die bisher teilweise auftretenden zwitscherartigen Störgeräusche eventuell verhindert werden.

Quantisierung: Sowohl im Coder wie auch im Decoder sind Quantisierer notwendig. So müssen im Coder die Teilbandsignale quantisiert werden. Im Decoder muß das rekonstruierte Ausgangssignal wieder quantisiert werden, damit es als 16 Bit PCM-Signal ausgegeben werden kann. Es ist zu überlegen, ob der Einsatz eines Noise-Shapers - also eine Rückkopplung des Quantisierungsfehlers - möglicherweise Vorteile bieten kann /14/. Auch der Einsatz eines Dithering-Verfahrens sollte untersucht werden, da dabei mit Hilfe eines überlagerten definierten Rauschsignals dafür gesorgt wird, daß Mittelwert und Leistung des Quantisierungsfehler-

signals unabhängig vom quantisierten Signal sind /14/. Dithering kann jedoch zu 4 dB bis 5 dB schlechteren SNR-Werten führen.

Wird bei einem tonalen Teilbandsignal, das aus einer oder wenigen Spektrallinien besteht, ein grober Quantisierer verwendet, so kann das Quantisierungsgeräusch teilweise auch nur aus wenigen Spektrallinien bestehen. Dieses tritt besonders dann auf, wenn die Teilbandfrequenz in einem einfachen Verhältnis zur Teilbandabtastrate steht. Da das Quantisierungsgeräusch somit nicht unbedingt Rausch-Charakter hat, wird es eventuell schlechter maskiert. Außerdem können solche Störgeräusche eventuell eher als "unsauberer Klang" (ähnlich Klirrstörungen und Intermodulationstönen) und weniger als "verrauschter Klang" wahrgenommen werden.

Psychoakustisches Modell: Bei der Untersuchung codierter Tonsignale hat sich herausgestellt, daß eine NMR-Summe von 0 keine subjektiv transparente Codierung garantiert (Beispiel: "Suzanne Vega"). Dieses deutet auf Probleme mit den psychoakustischen Modellen hin. Außerdem hat sich zum Beispiel herausgestellt, daß das Tonsignal "Klarinette" bei Codierung mit dem unmodifizierten Codec und einer Bitrate von 256 kbit/s deutlich besser klingt, wenn statt des zweiten das erste psychoakustische Modell verwendet wird. Zur subjektiven Untersuchung dieser Eigenschaften der psychoakustischen Modelle können die zu codierenden Tonsignale jeweils an der Maskierungsschwelle quantisiert werden. Das Quantisierungsgeräusch muß dabei immer gerade unter der Maskierungsschwelle liegen.

Die Anforderungen an das psychoakustische Modell können jedoch davon abhängen, ob entweder bei hoher Bitrate eine transparente Codierung oder bei niedriger Bitrate eine möglichst gute Codierung erzielt werden soll.

Crest-Faktoren der Teilbandsignale: Bei der Codierung werden mit den SMR-Werten und den SNR-Werten zunächst Leistungsverhältnisse beschrieben. Da ein Teilbandsignal vor der Quantisierung aber nicht auf seine maximale Leistung sondern auf seinen Spitzenwert normiert werden muß, geht nun auch sein Crest-Faktor (Verhältnis von Spitzenwert zu Effektivwert) in den wirklich erzielten Abstand zwischen Signal und Quantisierungsgeräusch ein. Untersuchungen dieses SQR-Werte (Signal/Quantization_Noise-Ratio), die schon in Abschnitt 4.2.1 kurz erläutert wurden, zeigen, daß die SQR-Werte in den codierten Teilbandsignalen im Mittel um bis zu 6 dB kleiner sind als die SNR-Werte. Die Differenz zwischen

SNR- und SQR-Wert variiert aber von Teilband zu Teilband und von Block zu Block. Die SNR-Summen und SQR-Summen einiger codierter Tonsignale sind im Anhang dargestellt (Anhang A.7.2).

Der SQR-Wert für ein Teilband eines Kanals kann aus verschiedenen Gründen kleiner als der entsprechende SNR-Wert sein: So kann das quantisierte Signal zum Beispiel einen höheren Crest-Faktor als das für die Berechnungen der SNR-Werte angenommene Sinussignal haben. Auch das Quantisierungsfehlergeräusch kann andere als die angenommenen Eigenschaften haben. Außerdem ermöglichen die Skalierungsfaktoren nur eine Normierung in Schritten von 2 dB, so daß eventuell der Quantisierer nicht voll ausgesteuert ist. Die durch die SCFSI gesteuerte mehrfache Verwendung von Skalierungsfaktoren kann diesen Effekt verstärken. Es ist nun zu überlegen, bis in wie weit das aktuelle Codierungsverfahren in diesem Punkt psychoakustisch gerechtfertigt ist oder gegebenenfalls verbessert werden kann.

Aliasing-Korrektur: Die Übertragungsfunktion der Teilbandfilter, die in der Polyphasenfilterbank für die Teilbandzerlegung verwendet werden, ist an den Teilbandgrenzen nicht beliebig steilflankig. Ein Signal, dessen Frequenz $1/4$ Teilbandbreite von der Grenze eines Teilbands entfernt ist, wird mit rund 20 dB gedämpft. Liegt die Frequenz mehr als $1/2$ Teilbandbreite außerhalb des Teilbands, so beträgt die Dämpfung über 100 dB. Bei der nach der Teilbandfilterung durchgeführten Abtastratenreduktion der Teilbandsignale entstehen somit durch Rückfaltung Alias-Komponenten. Die Synthesefilterbank im Decoder besitzt nun die Eigenschaft, diese Alias-Komponenten wieder auszulöschen (Aliasing Cancellation). Dafür muß aber jeweils auch das Teilbandsignal, in das die Frequenzen der Alias-Komponenten fallen, übertragen werden.

Das Problem ist nun, daß die Notwendigkeit der Übertragung dieser Alias-Komponenten von den psychoakustischen Modellen teilweise nicht berücksichtigt wird. Von beiden Modellen werden zunächst die für die Teilbandsignale geltenden Maskierungsschwellen berechnet. Dazu wird das zu codierende Signal mit Hilfe einer FFT spektral analysiert. Um aus den Maskierungsschwellen die bei der Bitzuteilung benutzten SMR-Werte zu berechnen, müssen auch die Signalpegel der Teilbandsignale bekannt sein. Beim ersten psychoakustischen Modell wird der Signalpegel aus betragsgrößten FFT-Linie, die in dem betrachteten Teilband liegt, und dem maximalen Skalierungsfaktor bestimmt. Da der Skalierungsfaktor auch von

gegebenenfalls entstandenen Alias-Komponenten abhängt, wird auf diese Weise die für die Alias-Korrektur notwendige Übertragung dieser Komponenten berücksichtigt.

Beim zweiten psychoakustischen Modell wird zur Berechnung des Signalpegels die Summe der Betragsquadrate der in ein Teilband fallenden FFT-Linien bestimmt. Da die FFT auf das nicht teilbandgefilterte Gesamtsignal angewendet wurde, werden daher auch keine Alias-Komponenten berücksichtigt. Dieses kann im Extremfall dazu führen, daß ein Teilband, in dem nur Alias-Komponenten liegen, gar nicht übertragen wird. Bei der Decodierung bleiben diese Alias-Störungen dann vollständig erhalten. In der Praxis tritt dieses Problem hauptsächlich bei sehr tonalen Signalen auf. Der oben erwähnte Extremfall trat zum Beispiel auf, als ein Rechtecksignal mit einer Frequenz von 960 Hz mit dem zweiten psychoakustischen Modell bei einer niedrigen Bitrate codiert wurde. Ein möglicher Lösungsansatz ist, daß bei der Berechnung der Signalpegels auch die etwas außerhalb des Teilbands liegenden FFT-Linien, mit dem Betragsfrequenzgang des Teilbandfilters gewichtet, in die Summation einbezogen werden.

Dieses Problem könnte voraussichtlich aber auch dadurch gelöst werden, daß vom psychoakustischen Modell nur die Maskierungsschwellen und nicht die SMR-Werte berechnet werden würden. Bei der Bestimmung des Quantisierungsgeräuschpegels könnte dann zum Beispiel auch der wirklich für die Normierung verwendete Skalierungsfaktor berücksichtigt werden. Damit im Zusammenhang stehen dann auch die oben erwähnten Überlegungen zu den Crest-Faktoren der Teilbandsignale.

Vor-Echo-Störungen: Vor-Echo-Störungen können infolge der blockweisen Codierung entstehen, wenn am Ende eines Blocks ein lautes impulsartiges Signal - wie zum Beispiel ein harter Anschlag - auftritt. Die vom psychoakustischen Modell für einen Block berechneten Maskierungsschwellen sind nun hauptsächlich durch das laute impulsartige Signal und nicht die leiseren Signalanteile vor dem Impuls bestimmt. Da der aus der Psychoakustik bekannte Vorverdeckungseffekt jedoch nur für kürzere Zeiten als die Blocklänge von hier 24 ms deutlich wirksam ist, hat das am Blockende liegende impulsartige Signal zu Beginn des Blocks noch keine starke Maskierungswirkung. Trotzdem wird der für einen gesamten Block eines Teilbandsignals verwendete Quantisierer von der oben berechneten Maskierungsschwelle bestimmt und kann daher am Anfang des Blocks zu noch hörbaren Quantisierungsgeräuschen führen. Zwar werden in einem Block

entsprechend der Scalefactor Select Information (SCFSI) maximal 3 verschiedene Skalierungsfaktoren für die Teilblöcke von 8 ms Länge verwendet, so daß damit der Verlauf des Quantisierungsgeräuschpegels an den Verlauf des Teilbandsignalpegels etwas angepaßt werden kann. Dennoch können aber Vor-Echo-Störungen hörbar sein. Als möglicher Ansatz, auch zu Blockbeginn gültige Maskierungsschwellen zu erhalten, kann zum Beispiel jeweils das Minimum der Maskierungsschwellen für den aktuellen und den vorangegangenen Block verwendet werden. Auch ist es möglich, daß eine Modifikation des Verfahrens zur Bestimmung der SCFSI Vor-Echo-Störungen verringern kann.

5.2 Optimierung der Codierung von Stereosignalen

In diesem Abschnitt werden zunächst verschiedene Optimierungsmöglichkeiten bei der Interkanal-Prädiktion erläutert. Anschließend werden verschiedene Ansätze vorgestellt, mit denen unabhängig von der Interkanal-Prädiktion eventuell die Codierung von Stereosignalen verbessert werden kann.

Iterative Bitzuteilung: In Abschnitt 4.1 wurden zwei Verfahren zur Bitzuteilung erläutert, die den Einsatz der Interkanal-Prädiktion möglich machen. Im Rahmen dieser Arbeit konnte nur das Verfahren der approximierten Bitzuteilung implementiert werden. Da das dabei benötigte Verfahren zur Schätzung der Bitzuteilung bei etwa 10 % bis 20 % aller codierten Teilbandsignale nicht genau die richtige Bitzuteilung schätzt, wird in diesen Fällen bei der endgültigen Bitzuteilung von einem anderen als dem später erzielten Prädiktionsgewinn ausgegangen. Dieses Problem könnte durch den Einsatz der iterativen Bitzuteilung gelöst werden.

Crest-Faktoren von $y(n)$ und $e(n)$: In Abschnitt 3.4.2 wird erläutert, daß das Teilbandsignal $y(n)$ und das Prädiktionsfehlersignal $e(n)$ unterschiedliche Crest-Faktoren aufweisen können. Daher wird zusätzlich zum Gewinn G_{opt} der dieses berücksichtigende Gewinn G_{peak} eingeführt. Könnte G_{opt} anstelle von G_{peak} verwendet werden, so würde dieses den notwendigen Rechenaufwand bei der Codierung verringern. Es wäre daher interessant, die Auswirkungen dieser Änderung auf die subjektive Signalqualität zu untersuchen. Dieses Problem hängt eng mit den in Abschnitt 5.1 erwähnten Überlegungen zu den Crest-Faktoren zusammen.

Prädiktionsfehlergeräusch: In Abschnitt 4.2.1 wurden die objektiven Untersuchungen der codierten Tonsignale erläutert. Bei diesen Untersuchungen wurde

auch die jeweils für einen Block gewählte Bitzuteilung analysiert. Wenn bei einer Bitrate von 128 kbit/s der Coder mit Interkanal-Prädiktion verwendet wurde, so ist es häufiger vorgekommen, daß das Prädiktionsfehlersignal $e_q(n)$ gar nicht quantisiert und codiert wurde. Daher ist das Teilbandsignal $y(n)$ nach der Decodierung nicht durch ein Quantisierungsfehlergeräusch sondern durch ein Prädiktionsfehlergeräusch gestört. Es könnte nun interessant sein, die objektiven und subjektiven Charakteristika des Prädiktionsfehlergeräusches im Vergleich zum Quantisierungsfehlergeräusch zu untersuchen.

SCFSI für $e(n)$: In Abschnitt 4.1.2 wird erläutert, daß sich bei der Ermittlung der SCFSI für das Prädiktionsfehlersignal $e(n)$ das prinzipielle Problem ergibt, daß die Hüllkurven der Signale $y(n)$ und $e(n)$ nicht unbedingt ähnlich sein müssen. Daher könnte es sinnvoll sein, die mit dem psychoakustischen Effekt der Nachverdeckung begründeten Teile des Verfahrens für die Bestimmung der SCFSI hier nicht zu benutzen.

Vor-Echo-Störungen: Zur Bestimmung eines optimalen Prädiktors wird - wie in Abschnitt 3.2 erläutert - die Prädiktionsfehlerleistung minimiert. Wenn nun zum Beispiel die Teilbandsignale $x(n)$ und $y(n)$ am Anfang eines Blocks relativ leise und am Ende des Blocks deutlich lauter sind, so werden dennoch die Prädiktionsfehlerleistungen am Blockanfang und -ende in gleicher Weise und mit gleicher Gewichtung berücksichtigt. Ergibt sich dadurch nun eine über die gesamte Länge des Blocks etwa konstante Prädiktionsfehlerleistung, so kann dieses als Vor-Echo-Störung wahrgenommen werden. Es wäre von Interesse, dieses Problem genauer zu untersuchen.

SCFSI bei Intensity Stereo: Da bei Intensity Stereo die Skalierungsfaktoren zur Nachbildung der Hüllkurven der beiden Stereosignale verwendet werden, kann es Vorteile bieten, das Verfahren zur Bestimmung der SCFSI in diesem Fall so zu ändern, daß nur dann nicht alle drei Skalierungsfaktoren übertragen werden, wenn zwei aufeinanderfolgende Skalierungsfaktoren genau gleich sind. Im Decoder würden dann genau die Skalierungsfaktoren verwendet werden, die auch zunächst im Coder bestimmt wurden.

Psychoakustisches Modell: Bei beiden psychoakustischen Modellen wird nicht berücksichtigt, daß ein Geräusch auf dem einen Stereo-Kanal auch Geräusche auf dem anderen Stereo-Kanal verdecken kann /6/. Dieser Effekt (Interchannel Masking) hängt jedoch stark von der Art der Wiedergabe ab - also zum Beispiel

davon, ob Lautsprecher oder Kopfhörer benutzt werden. Er würde wahrscheinlich eine Reduzierung der Stereo-Irrelevanz ermöglichen. Jedoch existieren für die Interkanal-Maskierung bisher keine entsprechenden psychoakustischen Modelle.

Interkanal-Prädiktion bei Layer III: Die Interkanal-Prädiktion wurde im Rahmen dieser Arbeit hauptsächlich für den ISO Audio-Codec für Layer II entwickelt, da er eine höhere Datenratenreduktion als Layer I ermöglicht. Ein als C-Programm realisierter Codec für Layer III stand während der Durchführung dieser Arbeit noch nicht zur Verfügung. Auch ist - im Vergleich zu Layer II - die Ergänzung von Layer III um die Interkanal-Prädiktion aufgrund der deutlich höheren Komplexität voraussichtlich wesentlich schwieriger.

In Abschnitt 4.2.1 ist zu erkennen, daß große Prädiktionsgewinne hauptsächlich bei Mono-ähnlichen Tonsignalen ("Suzanne Vega") und bei stark tonalen Tonsignalen ("Klarinette") erzielt werden. Da bei Layer III bei Joint Stereo Coding ein Summen- und Differenz-Signal-Verfahren eingesetzt wird, ermöglicht dieses auch eine Verbesserung der Codierung von Mono-ähnlichen Signalen /1/ /5/. Da anstelle von Teilbandsignalen bei Layer III deutlich feiner aufgelöste Spektralkomponenten codiert werden, ist es daher durchaus möglich, daß die Interkanal-Prädiktion in der hier vorgestellten Form bei Layer III geringere Vorteile bietet als bei Layer II. Es ist aber zu überlegen, ob bei Layer III jeweils die dem gleichen Frequenzbereich zugeordneten Spektralkomponenten der beiden Stereosignale gemeinsam codiert werden können.

5.3 Interkanal-Prädiktion bei Mehrkanalsignalen

Es bestehen Überlegungen, den ISO-Standard für die Audio-Codierung so zu erweitern, daß auch eine Übertragung von Mehrkanalsignalen möglich ist /12/. Mehrkanalsysteme ermöglichen eine "räumlichere" Wiedergabe von Tonsignalen als zum Beispiel Stereo-Signale und kommen gegenwärtig im Zusammenhang mit dem Medium Film schon häufig zum Einsatz. Eine Übertragung von Mehrkanalsignalen kann aber auch ohne begleitendes Bild sinnvoll eingesetzt werden. Dabei werden zum Beispiel Systeme mit 3 Hauptkanälen (Links, Mitte, Rechts) und 1 oder 2 Surround-Kanälen eingesetzt /2/. In der Praxis auftretende Mehrkanalsignalen enthalten voraussichtlich eine relativ hohe Redundanz und Irrelevanz, die bei der Codierung ausgenutzt werden sollte.

Erste einfache Untersuchungen an teilbandgefilterten Mehrkanalsignalen haben gezeigt, daß auch dort der Einsatz der Interkanal-Prädiktion deutliche Vorteile bei der Codierung bieten kann. Dabei wurden hauptsächlich statistische Abhängigkeiten zwischen zeitgleichen Teilbandsamples festgestellt. Es kann daher eventuell ausreichen, nur Prädiktoren mit einem Koeffizienten ($p = 0$) und keiner zusätzlichen Verzögerung ($d = 0$) einzusetzen. Außerdem könnte es Vorteile bieten, bei der Prädiktion von mehr als einem Kanal auszugehen. So könnte zum Beispiel ein Prädiktor einen Vorhersagewert für den Mitten-Kanal als Linearkombination der Abtastwerte des linken und des rechten Kanals bilden. Die Interkanal-Prädiktion bei Mehrkanalsignalen konnte im Rahmen dieser Arbeit nicht eingehender untersucht werden.

6 Zusammenfassung

Der ISO Audio-Codec stellt ein Verfahren zur Quellencodierung von digitalen Tonsignalen dar, das eine Reduzierung der bei der Übertragung benötigten Datenrate bei möglichst hoher subjektiver Tonqualität erlaubt. Dabei wird hauptsächlich eine Irrelevanz-Reduktion durchgeführt, bei der Verdeckungseffekte des menschlichen Gehörs ausgenutzt werden. Abhängig von der Komplexität des Codecs ist eine Reduzierung der Datenrate etwa um den Faktor 4 bis 12 möglich, so daß Stereosignale mit Bitraten in der Größenordnung von 384 kbit/s bis 128 kbit/s übertragen werden können. Bei den niedrigen Bitraten werden jedoch teilweise Störungen hörbar.

Die beiden Kanäle eines Stereosignals werden vom ISO Audio-Codec im Modus "Independent Stereo Coding" praktisch unabhängig voneinander codiert. Im Modus "Joint Stereo Coding" werden beide Kanäle gemeinsam codiert. Dabei wird mit relativ einfachen Verfahren die Stereo-Redundanz und die Stereo-Irrelevanz ausgenutzt. Da dabei nur Abhängigkeiten zwischen zeitgleichen Abtastwerte der beiden Kanäle eines Stereosignals ausgewertet werden, wird in der Regel nur eine geringe Reduktion der vorhandenen Stereo-Redundanz erzielt.

Ziel dieser Arbeit war es daher, die Stereo-Redundanz von Tonsignalen genauer zu untersuchen und anschließend bei der Codierung auszunutzen. Dabei wurde das Verfahren der Interkanal-Prädiktion angewendet, das die zwischen den beiden Kanälen eines Stereosignals bestehenden Kreuzkorrelationen sowie eventuelle Laufzeitunterschiede auswertet.

Im Rahmen dieser Arbeit wurde zunächst ein Verfahren zur Bestimmung der optimalen Parameter von Interkanal-Prädiktoren hergeleitet, das den Prädiktionsgewinn maximiert. Die Eigenschaften eines solchen Interkanal-Prädiktors und ihre Abhängigkeit von den freien Prädiktorparametern wurden anhand von verschiedenen theoretischen Signalmodellen untersucht. Dabei wurden auch die Auswirkungen der Quantisierung der Prädiktorparameter und des Prädiktor-Eingangssignals auf den Prädiktionsgewinn betrachtet.

Mit Hilfe eines Simulationsprogramms, das von der Teilbandzerlegung und der zeitlichen Blockstruktur des ISO Audio-Codecs für Layer II ausgeht und für jedes Teilband und jeden Block einen eigenen Interkanal-Prädiktor verwendet, wurden

die bei unterschiedlichen realen Tonsignalen erzielbaren Prädiktionsgewinne in Abhängigkeit von den freien Prädiktorparametern berechnet. Dabei traten teilweise Prädiktionsgewinne von über 40 dB in einem Teilband auf. Es zeigte sich, daß die erzielbaren Prädiktionsgewinne sehr stark von den Eigenschaften des untersuchten Tonsignals abhängen.

Bei Berücksichtigung des Umfangs der bei der Codierung der freien Prädiktorparameter auftretenden Seiteninformation ergab sich, daß es sinnvoll ist, abhängig von den aktuellen Signaleigenschaften Prädiktoren unterschiedlicher Komplexität einzusetzen. Ausgehend von diesen Ergebnissen wurde ein Verfahren zur Codierung der Prädiktorparameter entwickelt, das die Verwendung von Interkanal-Prädiktoren unterschiedlicher Komplexität erlaubt.

Es wurde dann ein als C-Programm vorliegender Audio-Codec für Layer II um die Interkanal-Prädiktion erweitert. Dazu wurde ein Verfahren zur Ermittlung der optimalen Prädiktorkomplexität und der optimalen Prädiktorparameter entwickelt, das die genannten Quantisierungseffekte und die für die Prädiktorparameter benötigte Seiteninformation berücksichtigt. Das Verfahren zur Auswahl der Teilbandquantisierer wurde so modifiziert, daß auch die erzielten Prädiktionsgewinne ausgenutzt werden.

Mit Hilfe dieses modifizierten Codecs wurde untersucht, inwieweit die Interkanal-Prädiktion eine Verbesserung der objektiven und subjektiven Tonqualität der codierten Stereosignale ermöglicht. Die Mehrzahl dieser Untersuchungen wurde mit einer Bitrate von 128 kbit/s für ein Stereosignal durchgeführt. Die objektiven Untersuchungen haben ergeben, daß mit der Interkanal-Prädiktion teilweise Gewinne möglich sind, die einer zusätzlichen Bitrate von über 40 kbit/s entsprechen. Die durch den Einsatz der Interkanal-Prädiktion erzielten Verbesserungen hängen aber stark von den Eigenschaften des zu codierenden Tonsignals ab. Bei allen untersuchten Tonsignalen war bei Einsatz der Interkanal-Prädiktion immer mindestens eine leichte objektive Qualitätsverbesserungen möglich.

Diese objektiven Aussagen stimmen weitgehend mit den subjektiven Beurteilungen der Tonqualität der codierten Signale überein. So war mit Hilfe der Interkanal-Prädiktion bei mehreren untersuchten Tonsignalen eine deutlich hörbare Verbesserung der Tonqualität möglich. Entgegen den Ergebnissen der objektiven Untersuchungen konnte bei einer Bitrate von 128 kbit/s auch bei Einsatz der Interkanal-Prädiktion

bei keinem der untersuchten Tonsignale eine transparente Codierung ohne hörbare Störungen erzielt werden.

Abschließend wurden Ansätze aufgezeigt, mit denen die Bestimmung der optimalen Prädiktoren weiter verbessert werden kann. Daneben wurden auch verschiedene, von der Interkanal-Prädiktion unabhängige Optimierungsvorschläge für den ISO Audio-Codec vorgestellt. Außerdem ist davon auszugehen, daß bei einer Kombination der Interkanal-Prädiktion mit Verfahren zur Ausnutzung der Stereo-Irrelevanz die Codierung von Stereosignalen weiter verbessert werden kann.

7 Literaturverzeichnis

- /1/ ISO Draft Standard:
Coding of moving pictures and associated audio for digital storage media
at up to about 1.5 Mbit/s; Part 3, Audio.
CD 11172-3, 1992.
- /2/ CCIR Draft New Recommendation (Doc. 10/11):
Multi-channel stereophonic sound system with and without accompanying
picture.
1992.
- /3/ Bauer, D.; Seitzer, D.:
Statistical Properties of High Quality Stereo Signals in the Time Domain.
Proc. ICASSP, 1989,
S. 2045 - 2048.
- /4/ Feldmann, D.:
Repetitorium der Ingenieur-Mathematik, Teil 2 H.
Hannover, 1989,
S. 34 - 40.
- /5/ Fuchs, H.:
Die Hörtests im Rahmen der ISO/MPEG-Audio Standardisierung.
Bericht der 17. Tonmeistertagung, Karlsruhe, November 1992.
- /6/ Johnston, J.D.:
Perceptual Transform Coding of Wideband Stereo Signals.
Proc. ICASSP, 1989,
S. 1993 - 1996.
- /7/ Makhoul, J.:
Linear Prediction: A Tutorial Review.
Proc. IEEE, vol. 63, pp 561-580, Apr. 1975.

- /8/ Musmann, H.G.:
Informationstheorie.
Universität Hannover, Institut für Theoretische Nachrichtentechnik und
Informationsverarbeitung.
Vorlesungsmanuskript und Arbeitsblätter zur Übung.
- /9/ Musmann, H.G.:
The ISO Audio Coding Standard.
IEEE Global Telecommunications Conference
San Diego, Calif., December 2-7, 1990,
S. 511 - 517.
- /10/ Oppenheim, A.V.; Schafer R.W.:
Digital Signal Processing.
Prentice-Hall, 1975,
S. 391 - 395.
- /11/ Seifert, M.:
Analyse und Verifikation zweier psychoakustischer Modelle für die
Codierung von Audiosignalen.
Universität Hannover, Institut für Theoretische Nachrichtentechnik und
Informationsverarbeitung.
Studienarbeit, 1992.
- /12/ Theile, G.; Stoll, G.; Silzle, A.:
MUSICAM-Surround: A Multi-Channel Stereo Coding Method.
92nd AES Convention, March 1992, Vienna, Preprint No. 3337.
- /13/ van der Waal, R.G.; Veldhuis, R.N.J.:
Subband Coding of Stereophonic Digital Audio Signals.
Proc. ICASSP, 1991,
S. 3601 - 3604.
- /14/ Wannamaker, R.A.:
Psychoacoustically Optimal Noise Shaping.
J. Audio Eng. Soc.; Vol. 40, No. 7/8, July/August 1992,
S. 611 - 620.

- /15/ Zwickler, E.; Zollner, M.:
Elektroakustik.
Berlin: Springer-Verlag, 2. Aufl., 1987,
S. 240 - 245.

Anhang

A.1	Mögliche Quantisierer für die verschiedenen Teilbänder	A 2
A.2	Kenngrößen der Quantisierer für die Teilbandsample	A 3
A.3	Mögliche Modi bei der Codierung der Prädiktorparameter	A 4
A.4	Verzeichnis der verwendeten Tonsignale	A 6
A.5	Prädiktionsgewinn bei realen Tonsignalen (Diagramme)	A 7
A.5.1	Prädiktionsgewinn G_{opt} vs. Zeit	A 7
A.5.2	Mittlerer Prädiktionsgewinn G_{opt} vs. Verzögerung d	A 11
A.5.3	Mittlerer Prädiktionsgewinn G_{opt} vs. maximale Verzögerung d_{max}	A 12
A.5.4	Mittlerer Prädiktionsgewinn G_{opt} vs. Prädiktorgrad p	A 13
A.5.5	Wahrscheinlichkeitsdichte $p(a_k)$ vs. Prädiktorkoeffizient a_k	A 14
A.5.6	Prädiktionsgewinne G_{opt} & G_{peak} & G_{quant} vs. Zeit	A 16
A.6	Prädiktionsgewinn und Seiteninformation (Diagramme)	A 18
A.6.1	Verbleibender Prädiktionsgewinn & Seiteninformation vs. Zeit	A 18
A.6.2	Mittlerer verbleibender Prädiktionsgewinn vs. Teilband	A 22
A.7	Objektive Qualität der codierten Tonsignale (Diagramme)	A 23
A.7.1	Objektive Qualität mit und ohne Interkanal-Prädiktion	A 23
A.7.2	Vergleich von SNR-Summe und SQR-Summe	A 28
A.7.3	Codierung eines Blocks des Tonsignals "Klarinette"	A 30

A.1 Mögliche Quantisierer für die verschiedenen Teilbänder

Teil- band	Anz. Bits	Anzahl der Quantisierungsstufen															
		(für Bitzuteilungs-Index 0 bis 15)															
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	4	-	3	7	15	31	63	127	255	511	1023	2047	4095	8191	16383	32767	65535
1	4	-	3	7	15	31	63	127	255	511	1023	2047	4095	8191	16383	32767	65535
2	4	-	3	7	15	31	63	127	255	511	1023	2047	4095	8191	16383	32767	65535
3	4	-	3	5	7	9	15	31	63	127	255	511	1023	2047	4095	8191	65535
4	4	-	3	5	7	9	15	31	63	127	255	511	1023	2047	4095	8191	65535
5	4	-	3	5	7	9	15	31	63	127	255	511	1023	2047	4095	8191	65535
6	4	-	3	5	7	9	15	31	63	127	255	511	1023	2047	4095	8191	65535
7	4	-	3	5	7	9	15	31	63	127	255	511	1023	2047	4095	8191	65535
8	4	-	3	5	7	9	15	31	63	127	255	511	1023	2047	4095	8191	65535
9	4	-	3	5	7	9	15	31	63	127	255	511	1023	2047	4095	8191	65535
10	4	-	3	5	7	9	15	31	63	127	255	511	1023	2047	4095	8191	65535
11	3	-	3	5	7	9	15	31	65535								
12	3	-	3	5	7	9	15	31	65535								
13	3	-	3	5	7	9	15	31	65535								
14	3	-	3	5	7	9	15	31	65535								
15	3	-	3	5	7	9	15	31	65535								
16	3	-	3	5	7	9	15	31	65535								
17	3	-	3	5	7	9	15	31	65535								
18	3	-	3	5	7	9	15	31	65535								
19	3	-	3	5	7	9	15	31	65535								
20	3	-	3	5	7	9	15	31	65535								
21	3	-	3	5	7	9	15	31	65535								
22	3	-	3	5	7	9	15	31	65535								
23	2	-	3	5	65535												
24	2	-	3	5	65535												
25	2	-	3	5	65535												
26	2	-	3	5	65535												
27	0	-															
28	0	-															
29	0	-															
30	0	-															
31	0	-															

Diese Tabelle gilt unter anderem für Stereosignalen mit 48 kHz Abtastfrequenz und Bitraten von 112 kbit/s bis 384 kbit/s. Sie führt die bei Layer II in den einzelnen Teilbändern erlaubten Quantisierer mit der Anzahl ihrer Quantisierungsstufen auf. Die Spalte "Anz. Bits" gibt an, wieviele Bits für die Codierung der gewählten Bitzuteilung benutzt werden. Insgesamt werden pro Kanal dazu 88 Bits benötigt. (aus /1/)

A.2 Kenngrößen der Quantisierer für die Teilbandsample

Anzahl der Quantisierungs- stufen	Abtastwerte pro Codewort	Bits pro Codewort	Signal/Rausch- Abstand (SNR / dB)
-	-	0	0.00
3	3	5	7.00
5	3	7	11.00
7	1	3	16.00
9	3	10	20.84
15	1	4	25.28
31	1	5	31.59
63	1	6	37.75
127	1	7	43.84
255	1	8	49.89
511	1	9	55.93
1023	1	10	61.96
2047	1	11	67.98
4095	1	12	74.01
8191	1	13	80.03
16383	1	14	86.05
32767	1	15	92.01
65535	1	16	98.01

Diese Tabelle gibt für die unterschiedlichen Quantisierer von Layer II an, wieviele Bits für die Codierung eines Teilbandsamples benötigt werden. Bei 3, 5 oder 9 Quantisierungsstufen werden jeweils 3 aufeinanderfolgende Teilbandsample gemeinsam in einem Codewort codiert. Die letzte Spalte gibt den bei der Bitzuteilung angenommenen Signal/Rausch-Abstand (SNR) in dB an. (aus /1/)

A.3 Mögliche Modi bei der Codierung der Prädiktorparameter

pmode	p	d	pd	Anz. Bits	SI (dB)	Code für Bitstrom	Kommentar
-	-	-	-	-	-	0 0	reserviert
0	-	-	-	2	0,3	1 1	keine Prädiktion
1	0	0	0..1	6	1,0	1 0 pd a ₀ *	
2	0	0	0..1	14	2,3	0 1 0 0 pd sf a ₀	
3	0	1..16	0..1	18	3,0	0 1 0 1 pd d sf a ₀	
4	1	0	0..1	22	3,7	0 1 1 0 0 0 pd sf a ₀ a ₁	
5	1	1..16	0..1	26	4,3	0 1 1 0 0 1 pd d sf a ₀ a ₁	
6	2	0	0..1	28	4,7	0 1 1 0 1 0 pd sf a ₀ .. a ₂	
7	2	1..16	0..1	32	5,4	0 1 1 0 1 1 pd d sf a ₀ .. a ₂	
8	3	0	0..1	34	5,7	0 1 1 1 0 0 pd sf a ₀ .. a ₃	
9	3	1..15	0..1	38	6,4	0 1 1 1 0 1 pd d sf a ₀ .. a ₃	
10	4	0	0..1	41	6,9	0 1 1 1 1 0 0 pd sf a ₀ .. a ₄	
11	5	0	0..1	47	7,9	0 1 1 1 1 0 1 pd sf a ₀ .. a ₅	
12	6	0	0..1	53	8,9	0 1 1 1 1 1 0 pd sf a ₀ .. a ₆	
13	7	0	0..1	59	9,9	0 1 1 1 1 1 1 pd sf a ₀ .. a ₇	

Für die Codierung des gewählten Prädiktormodus pmode wird ein Präfixcode mit variabler Codewortlänge verwendet. Für jeden der möglichen Modi ist der Prädiktordergrad p und der für die Verzögerung d zulässige Bereich festgelegt. Die weiteren Bestandteile des Codes für einen Prädiktor sind unten erläutert. "Anz. Bits" gibt die insgesamt für die Codierung eines Prädiktors benötigten Bits an. "SI" (Seiteninformation) gibt an, um wieviel dB der Prädiktionsgewinn bei Berücksichtigung der für die Codierung des Prädiktors benötigten Bits reduziert werden muß. Dabei wird von einer Blocklänge von N = 36 Teilbandsample sowie einem Signal/Rausch-Abstand (SNR) von 6,02 dB pro für die Teilbandsamplequantisierung benutztem Bit ausgegangen.

Name	Codierte Information	Länge	Codierung
pd	Prädiktionsrichtung pdir	1 Bit	0 x = Rechts y = Links 1 x = Links y = Rechts
d	Verzögerung d	4 Bit	0000 d = 1 0001 d = 2 ... 1111 d = 16
sf	Skalierungsfaktor (für alle Koeffizienten eines Prädiktors)	3 Bit	000 sf = 0,25 001 sf = 0,5 010 sf = 0,75 011 sf = 1,0 100 sf = 1,5 101 sf = 2,0 110 sf = 4,0 111 sf = 8,0
a ₀ *	Prädiktorkoeffizient a ₀ (ungleichförmige Quantisierung) (ohne Skalierungsfaktor) (p=0, d=0, a ₀ =0 -> keine Prädiktion)	3 Bit	000 a ₀ = -1,0 001 a ₀ = -0,5 010 a ₀ = 0,25 011 a ₀ = 0,5 100 a ₀ = 0,75 101 a ₀ = 1,0 110 a ₀ = 1,25 111 a ₀ = 1,5
a _k	Prädiktorkoeffizient a _k (gleichförmige Quantisierung) (mit Skalierungsfaktor sf)	6 Bit	000000 a _k = -31(sf/32) 000001 a _k = -30(sf/32) ... 011110 a _k = -1(sf/32) 011111 a _k = 0 100000 a _k = 1(sf/32) ... 111110 a _k = 31(sf/32) 111111 nicht erlaubt !

Der Code 111111 für einen Prädiktorkoeffizienten wird nicht verwendet, um so möglichst Folgen von 12 oder mehr "1" Bits zu verhindern, die mit dem Synchronisationswort "1111 1111 1111" verwechselt werden könnten.

A.4 Verzeichnis der verwendeten Tonsignale

Für die objektiven und subjektiven Untersuchungen der Interkanal-Prädiktion und des erweiterten Audio-Codecs für Layer II wurden die folgenden fünf Tonsignale verwendet:

- <1> Clarinet (arpeggio) 0'08"
(Klarinette: Folge von 7 Einzeltönen mit steigender Tonhöhe)
- <2> Harpsichord (arpeggio) 0'08"
(Cembalo: Folge von 10 Einzeltönen mit steigender Tonhöhe)
- <3> Castanets 0'08"
(Kastagnetten: Kurzer Solo-Rhythmus)
- <4> Wind ensemble (Stravinsky) 0'13"
(Ausschnitt aus einer Aufnahme eines Blasinstrumenten-Orchesters)
- <5> Suzanne Vega: Tom's Diner 0'21"
(Anfang des a-cappella-Stückes "Tom's Diner")

Alle fünf Tonsignale sind Stereosignale. Das Signal <5> hat jedoch einen starken Mono-Charakter. Die Abtastrate der im Rahmen dieser Arbeit verwendeten Tonsignale war durch Umcodierung von 44,1 kHz auf 48 kHz geändert worden.

Die Tonsignale <1> bis <4> kommen von folgender CD:

European Broadcasting Union (EBU):
Sound Quality Assessment Material (SQAM)
(Recordings for subjective tests)

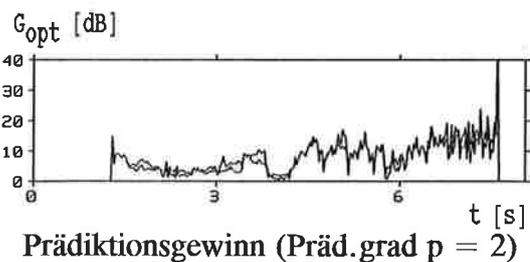
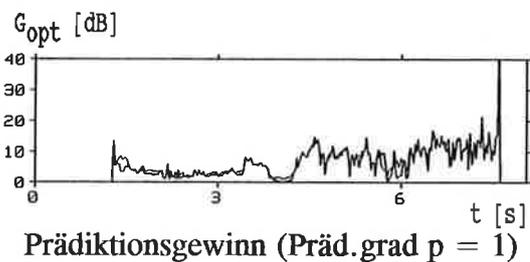
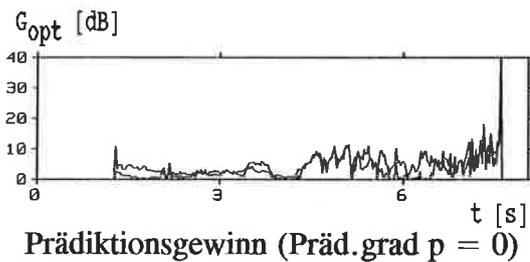
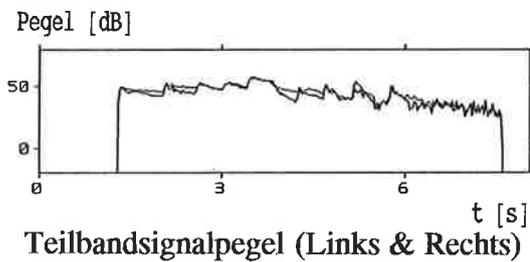
Das Tonsignal <5> kommt von folgender CD:

Suzanne Vega:
Solitude Standing

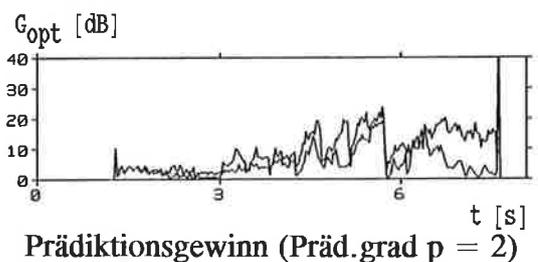
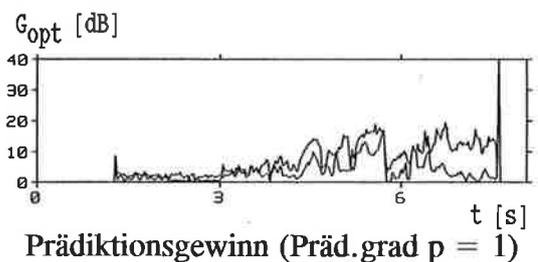
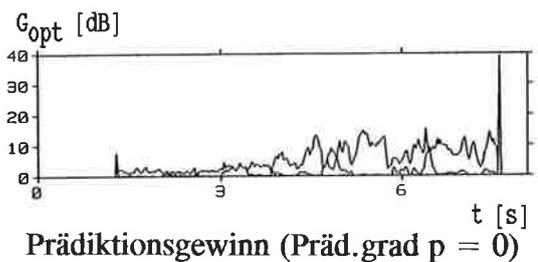
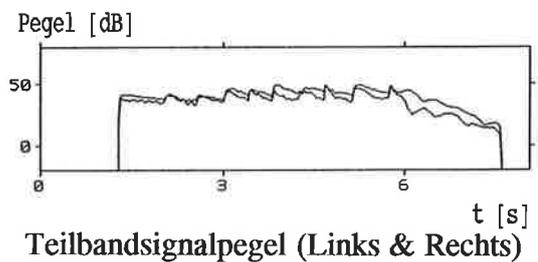
A.5 Prädiktionsgewinn bei realen Tonsignalen (Diagramme)

A.5.1 Prädiktionsgewinn G_{opt} vs. Zeit

Tonsignal "Cembalo"
Teilband 0 (0 Hz - 750 Hz)



Tonsignal "Cembalo"
Teilband 1 (750 Hz - 1500 Hz)



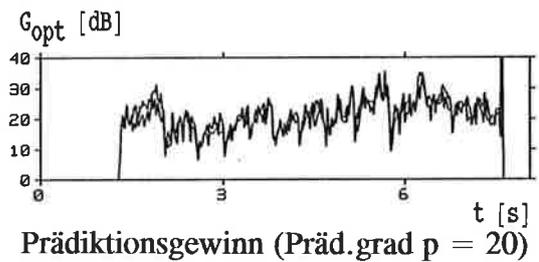
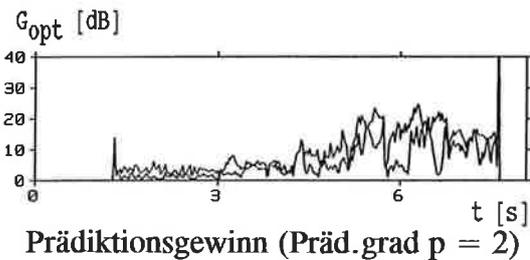
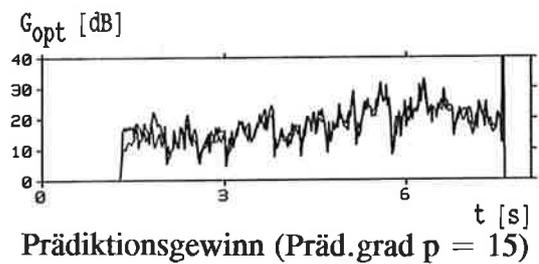
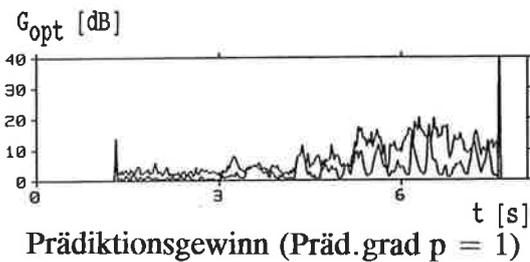
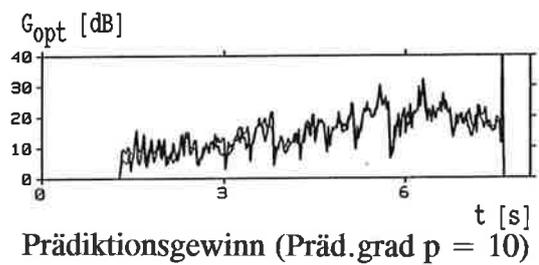
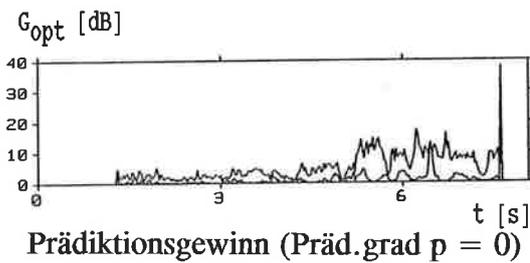
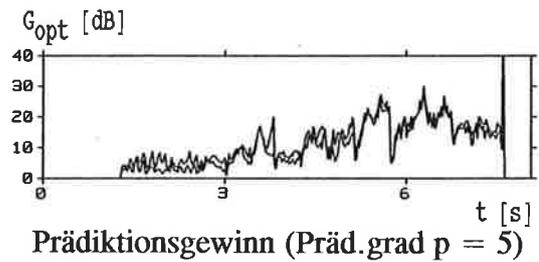
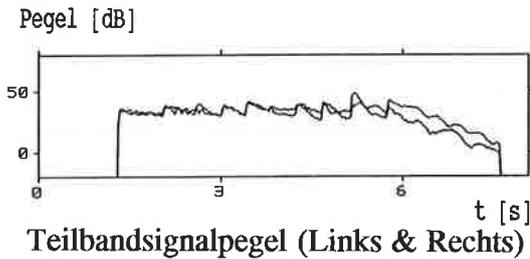
Bei den Prädiktionsgewinnen gilt:

obere Kurve: optimale Verzögerung d ($0 \leq d \leq d_{max} = 16$) / optimale Prädiktionsrichtung

untere Kurve: feste Verzögerung $d = 0$ / optimale Prädiktionsrichtung

Blocklänge: $N = 36$ Teilbandsample (24 ms)

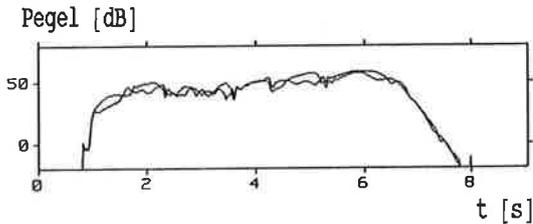
Tonsignal "Cembalo"
Teilband 2 (1500 Hz - 2250 Hz)



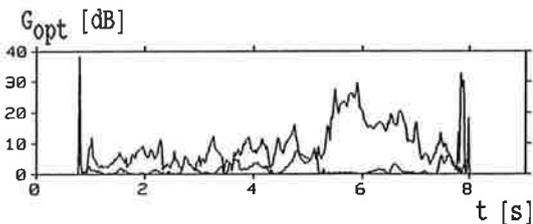
Bei den Prädiktionsgewinnen gilt:
obere Kurve: optimale Verzögerung d ($0 \leq d \leq d_{max} = 16$) / optimale Prädiktionsrichtung
untere Kurve: feste Verzögerung $d = 0$ / optimale Prädiktionsrichtung
Blocklänge: $N = 36$ Teilbandsample (24 ms)

Tonsignal "Klarinette"

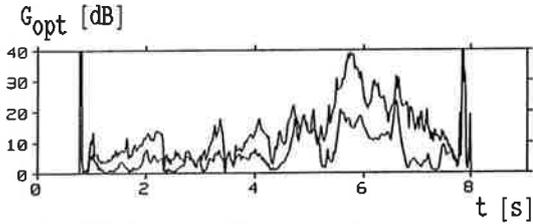
Teilband 2 (1500 Hz - 2250 Hz)



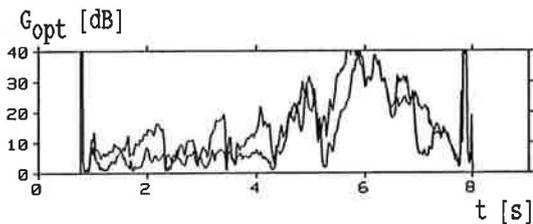
Teilbandsignalpegel (Links & Rechts)



Prädiktionsgewinn (Präd.grad p = 0)



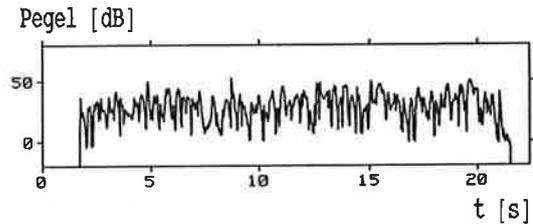
Prädiktionsgewinn (Präd.grad p = 1)



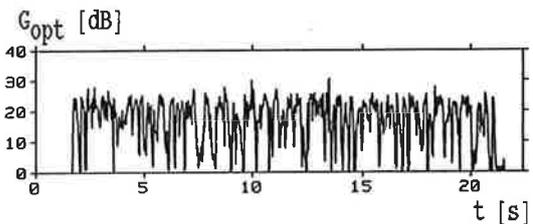
Prädiktionsgewinn (Präd.grad p = 2)

Tonsignal "Suzanne Vega"

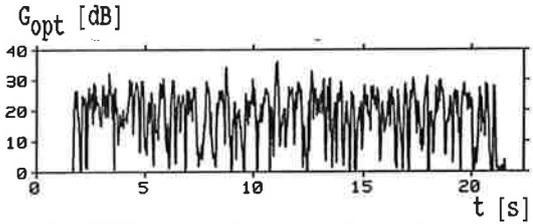
Teilband 2 (1500 Hz - 2250 Hz)



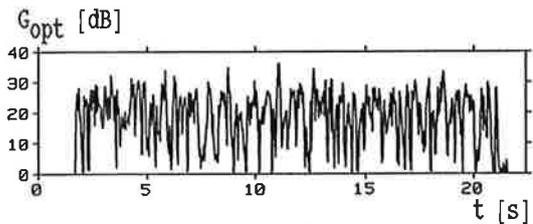
Teilbandsignalpegel (Links & Rechts)



Prädiktionsgewinn (Präd.grad p = 0)



Prädiktionsgewinn (Präd.grad p = 1)



Prädiktionsgewinn (Präd.grad p = 2)

Bei den Prädiktionsgewinnen gilt:

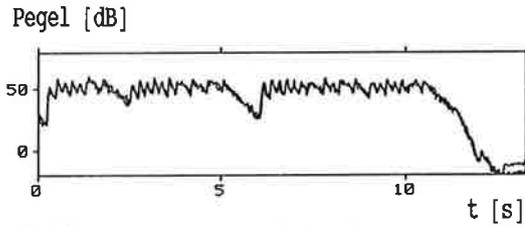
obere Kurve: optimale Verzögerung d ($0 \leq d \leq d_{\max} = 16$) / optimale Prädiktionsrichtung

untere Kurve: feste Verzögerung $d = 0$ / optimale Prädiktionsrichtung

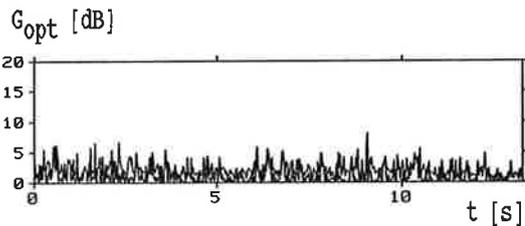
Blocklänge: $N = 36$ Teilbandsample (24 ms)

Tonsignal "Stravinsky"

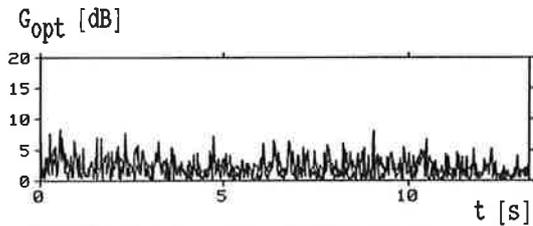
Teilband 2 (1500 Hz - 2250 Hz)



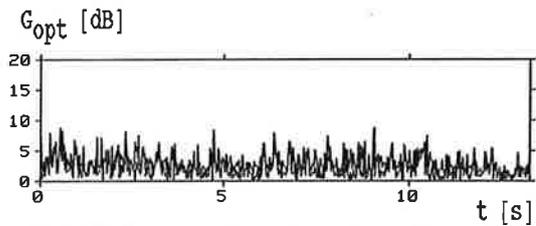
Teilbandsignalpegel (Links & Rechts)



Prädiktionsgewinn (Präd.grad p = 0)



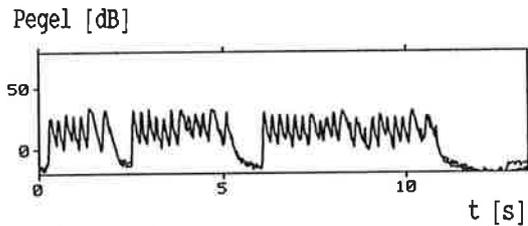
Prädiktionsgewinn (Präd.grad p = 1)



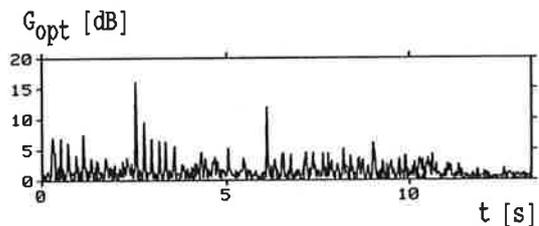
Prädiktionsgewinn (Präd.grad p = 2)

Tonsignal "Stravinsky"

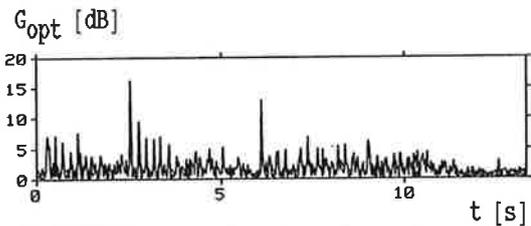
Teilband 15 (11250 Hz - 12000 Hz)



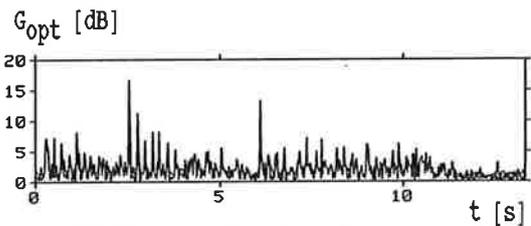
Teilbandsignalpegel (Links & Rechts)



Prädiktionsgewinn (Präd.grad p = 0)



Prädiktionsgewinn (Präd.grad p = 1)

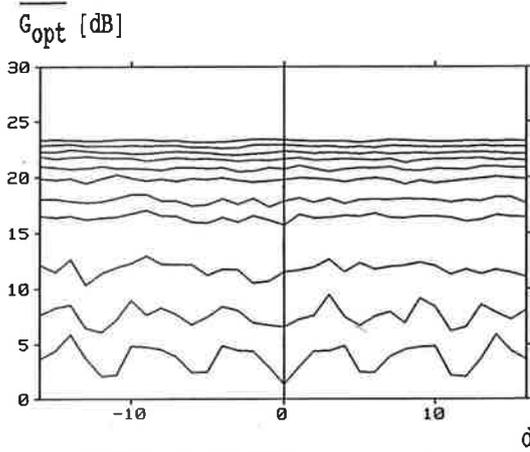


Prädiktionsgewinn (Präd.grad p = 2)

Bei den Prädiktionsgewinnen gilt:
obere Kurve: optimale Verzögerung d ($0 \leq d \leq d_{\max} = 16$) / optimale Prädiktionsrichtung
untere Kurve: feste Verzögerung $d = 0$ / optimale Prädiktionsrichtung
Blocklänge: $N = 36$ Teilbandsample (24 ms)

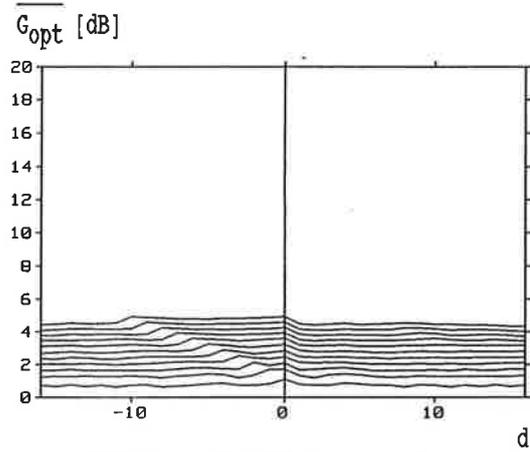
A.5.2 Mittlerer Prädiktionsgewinn G_{opt} vs. Verzögerung d

Tonsignal "Klarinette"
Teilband 2



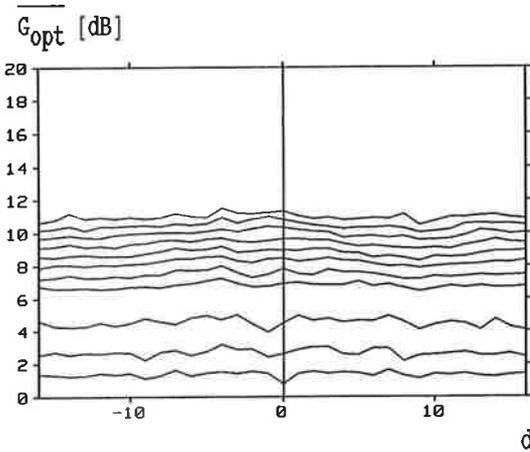
Mittlerer Prädiktionsgewinn

Tonsignal "Stravinsky"
Teilband 2



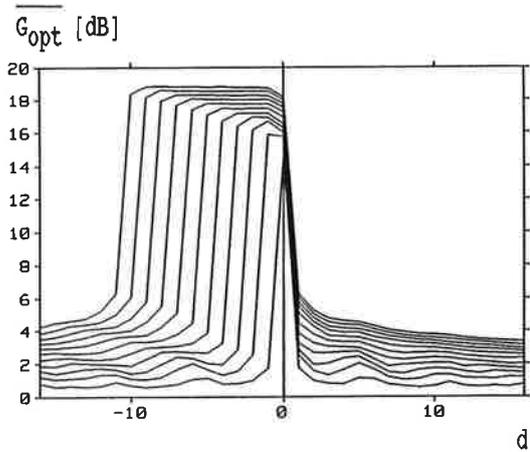
Mittlerer Prädiktionsgewinn

Tonsignal "Cembalo"
Teilband 2



Mittlerer Prädiktionsgewinn

Tonsignal "Suzanne Vega"
Teilband 2



Mittlerer Prädiktionsgewinn

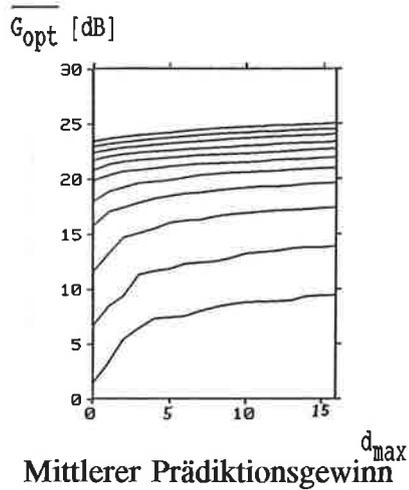
Parameter: Prädiktorgrad $p = 0, 1, 2, \dots, 10$
(unterste Kurve: $p = 0$ / oberste Kurve: $p = 10$)

feste Verzögerung d ($-16 \leq d \leq 16$) / optimale Prädiktionsrichtung / Blocklänge $N = 36$

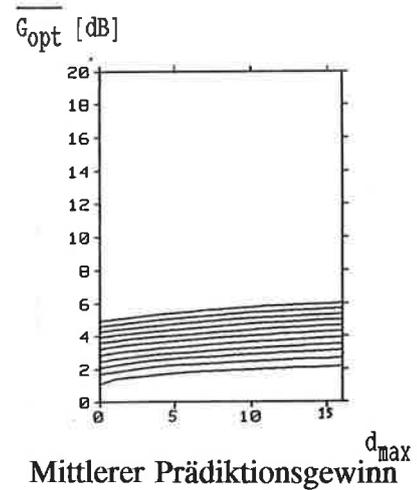
Die G_{opt} in dB werden über die gesamte Signaldauer gemittelt.
Beim Signal "Klarinette" werden dabei Vor- und Nachspann nicht berücksichtigt.

A.5.3 Mittlerer Prädiktionsgewinn G_{opt} vs. maximale Verzögerung d_{max}

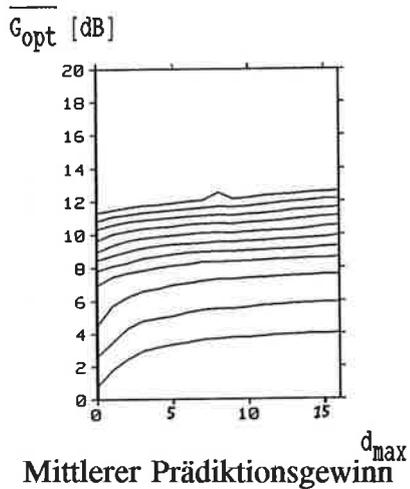
Tonsignal "Klarinette"
Teilband 2



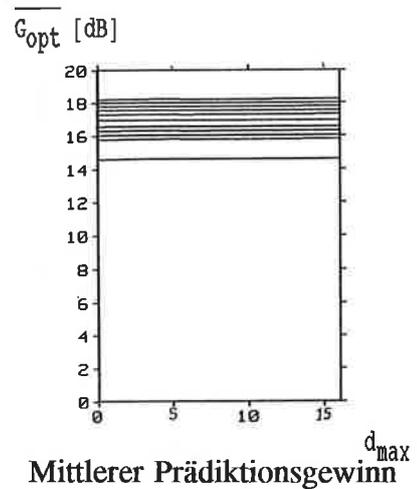
Tonsignal "Stravinsky"
Teilband 2



Tonsignal "Cembalo"
Teilband 2



Tonsignal "Suzanne Vega"
Teilband 2



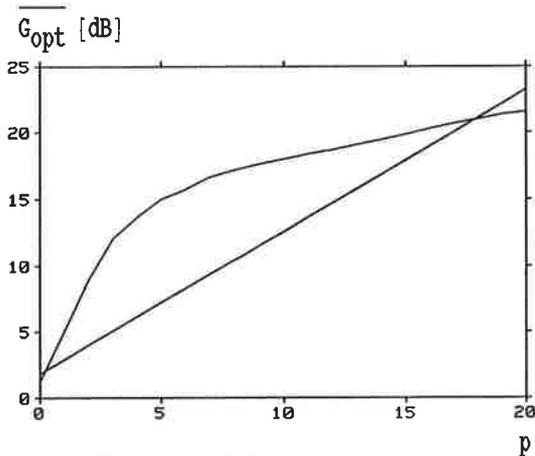
Parameter: Prädiktorgrad $p = 0, 1, 2, \dots, 10$
(unterste Kurve: $p = 0$ / oberste Kurve: $p = 10$)

optimale Verzögerung d ($0 \leq d \leq d_{max}$) ($0 \leq d_{max} \leq 16$) / optimale Prädiktionsrichtung / $N = 36$

Die G_{opt} in dB werden über die gesamte Signaldauer gemittelt.
Beim Signal "Klarinette" werden dabei Vor- und Nachspann nicht berücksichtigt.

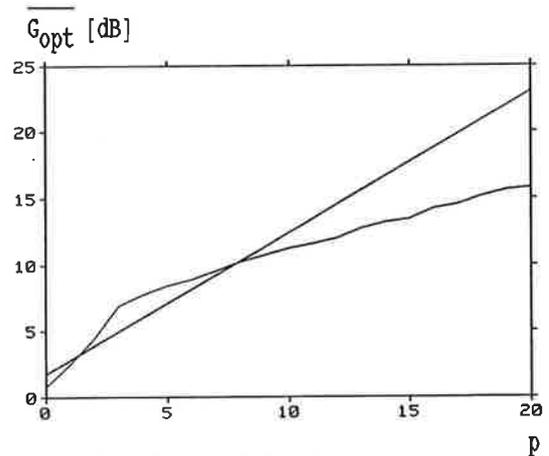
A.5.4 Mittlerer Prädiktionsgewinn G_{opt} vs. Prädiktorgrad p

Tonsignal "Klarinette"
Teilband 2

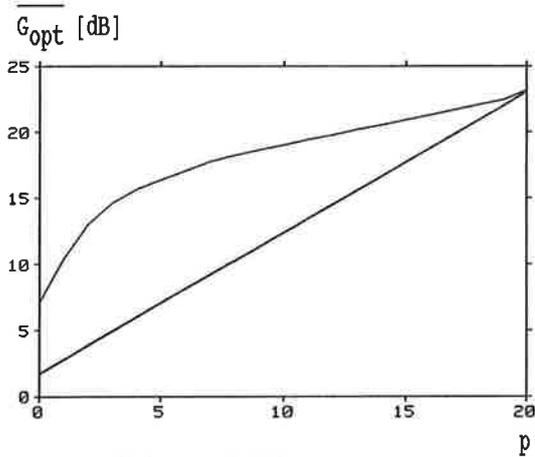


Mittlerer Prädiktionsgewinn
(feste Verzögerung, $d = 0$)

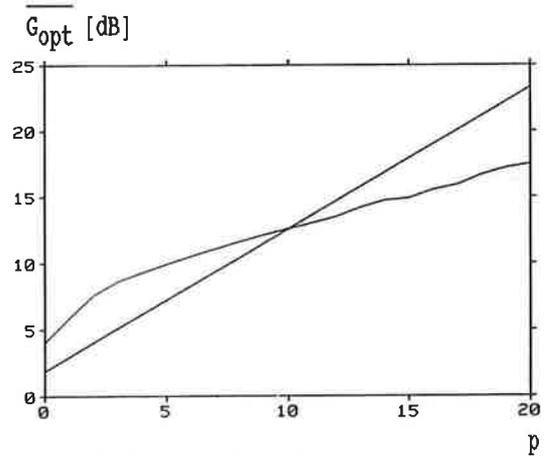
Tonsignal "Cembalo"
Teilband 2



Mittlerer Prädiktionsgewinn
(feste Verzögerung, $d = 0$)



Mittlerer Prädiktionsgewinn
(optimale Verzögerung, $d_{max} = 16$)



Mittlerer Prädiktionsgewinn
(optimale Verzögerung, $d_{max} = 16$)

Oben: feste Verzögerung $d = 0$ / optimale Prädiktionsrichtung / $N = 36$

Unten: optimale Verzögerung d ($0 \leq d \leq d_{max} = 16$) / optimale Prädiktionsrichtung / $N = 36$

Die G_{opt} in dB werden über die gesamte Signaldauer gemittelt.

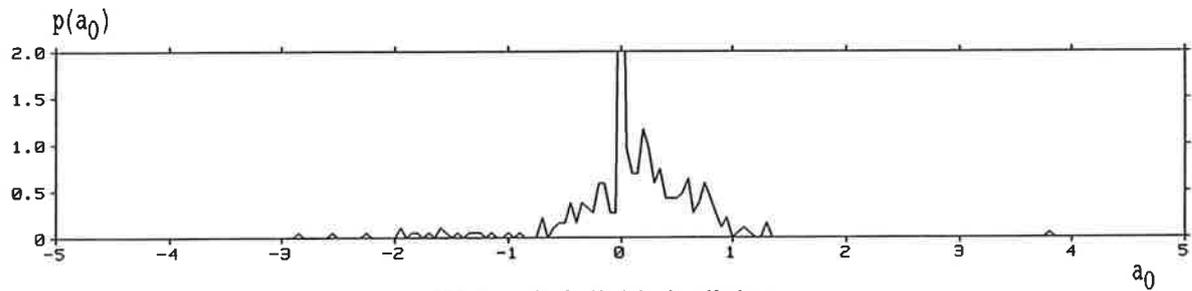
Die eingezeichnete Gerade gibt die ungefähr benötigte Seiteninformation SI [dB] bei Koeffizientenquantisierung mit 64 Stufen (6 Bit) an.

(Annahme: $SI/dB = 1 + (p+1)$)

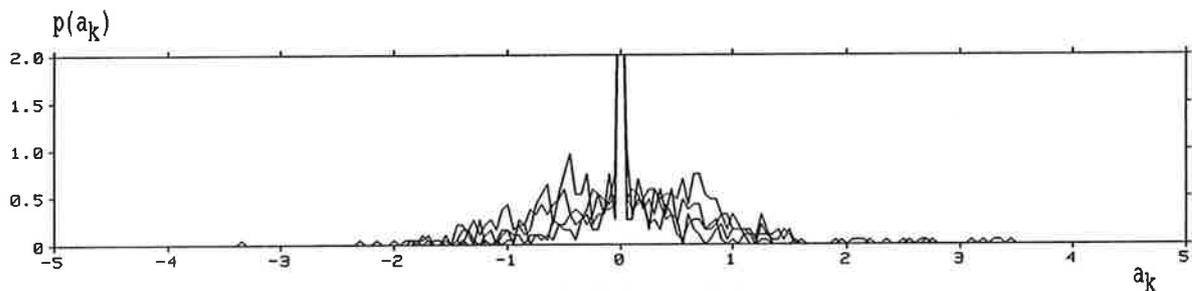
A.5.5 Wahrscheinlichkeitsdichte $p(a_k)$ vs. Prädiktorkoeffizient a_k

Tonsignal "Klarinette"

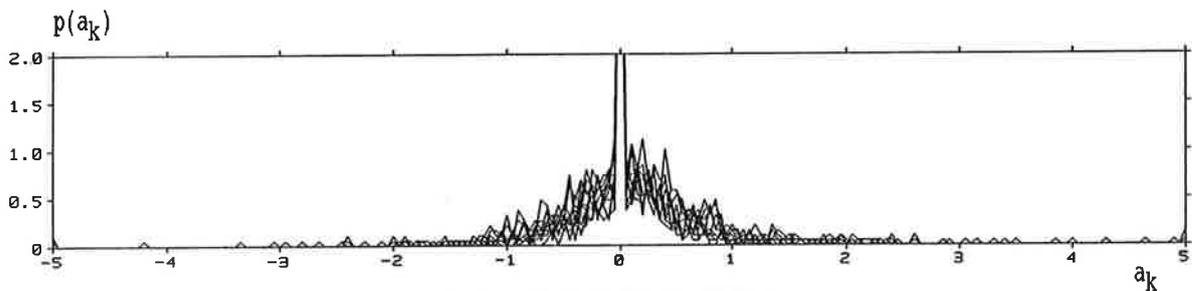
Teilband 2



Wahrscheinlichkeitsdichte
(Prädiktorgrad $p = 0$ / Koeffizient a_0)



Wahrscheinlichkeitsdichte
(Prädiktorgrad $p = 3$ / Koeffizienten $a_0 \dots a_3$)



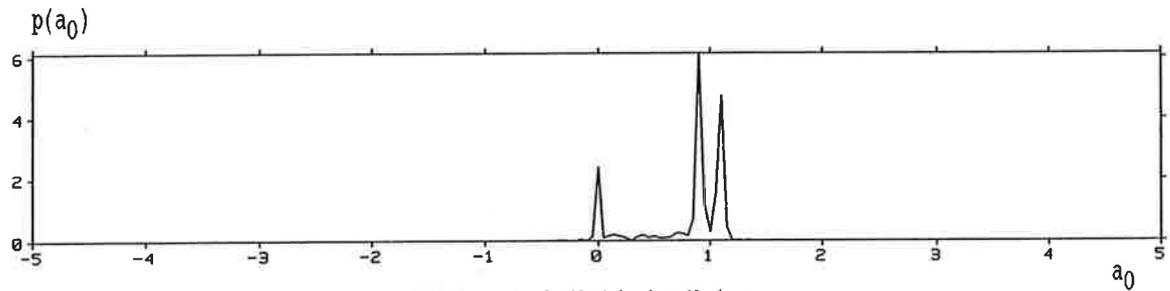
Wahrscheinlichkeitsdichte
(Prädiktorgrad $p = 9$ / Koeffizienten $a_0 \dots a_9$)

festе Verzögerung $d = 0$ / optimale Prädiktionsrichtung / Blocklänge $N = 36$

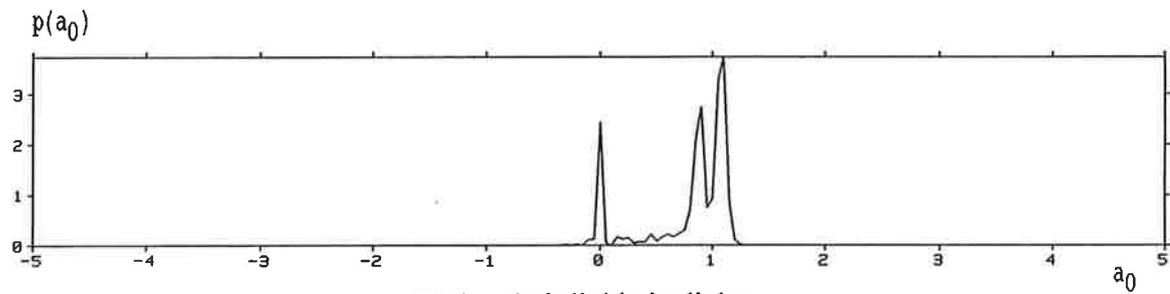
Die Wahrscheinlichkeitsdichten werden über die gesamte Signaldauer ermittelt.

Tonsignal "Suzanne Vega"

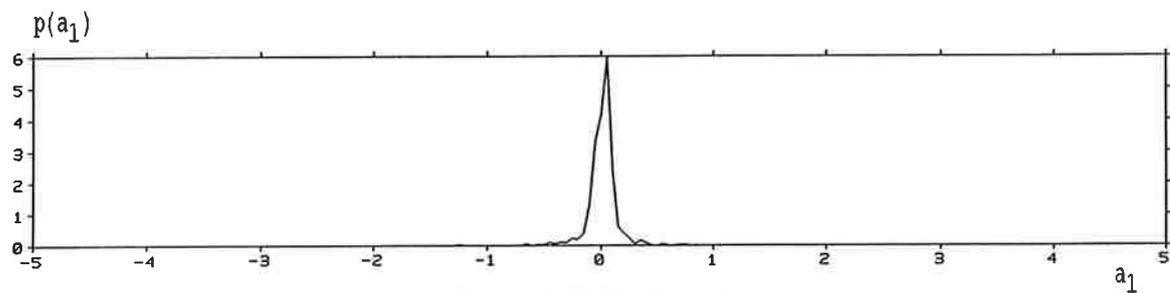
Teilband 2



Wahrscheinlichkeitsdichte
(Prädiktorgrad $p = 0$ / Koeffizient a_0)



Wahrscheinlichkeitsdichte
(Prädiktorgrad $p = 1$ / Koeffizient a_0)



Wahrscheinlichkeitsdichte
(Prädiktorgrad $p = 1$ / Koeffizient a_1)

feste Verzögerung $d = 0$ / optimale Prädiktionsrichtung / Blocklänge $N = 36$

Die Wahrscheinlichkeitsdichten werden über die gesamte Signaldauer ermittelt.

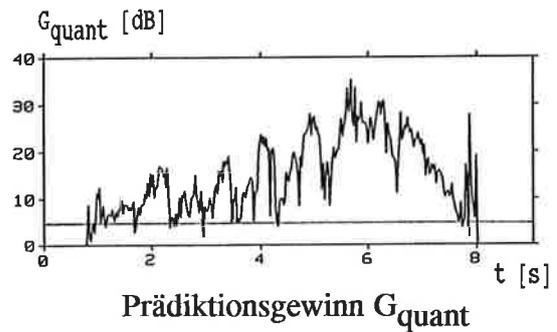
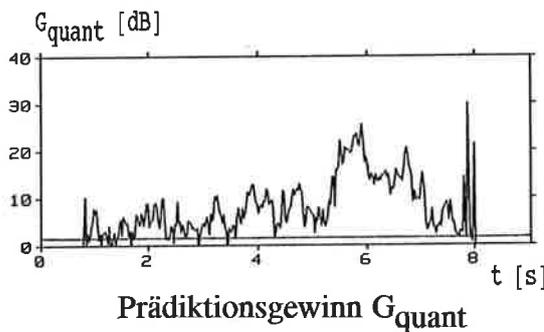
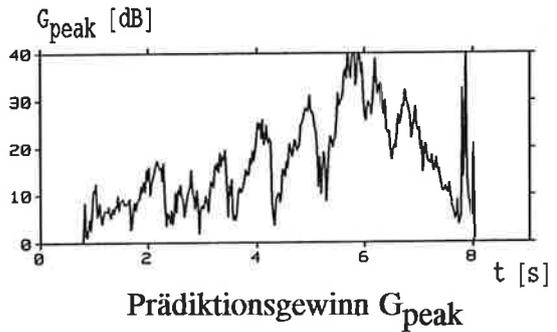
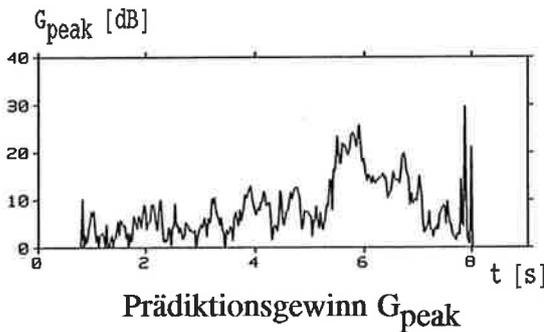
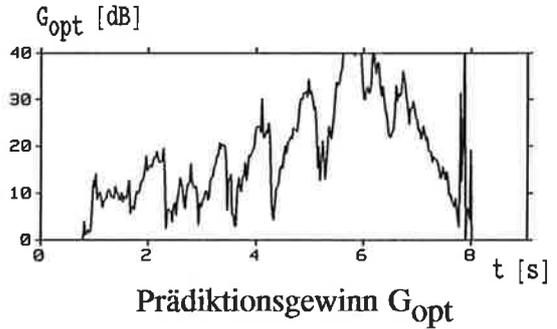
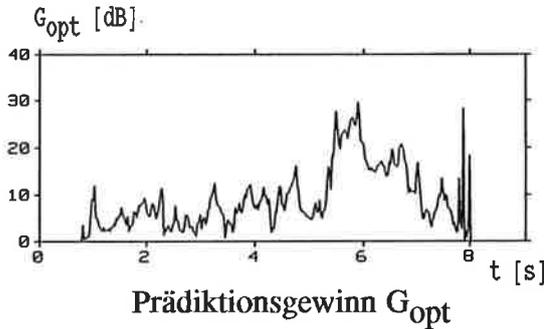
A.5.6 Prädiktionsgewinne G_{opt} & G_{peak} & G_{quant} vs. Zeit

Tonsignal "Klarinette"

Teilband 2
(Prädiktorgrad $p = 0$)
(optimale Verzögerung, $d_{max} = 16$)

Tonsignal "Klarinette"

Teilband 2
(Prädiktorgrad $p = 3$)
(optimale Verzögerung, $d_{max} = 16$)



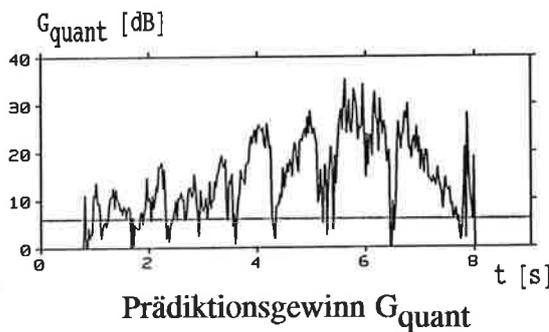
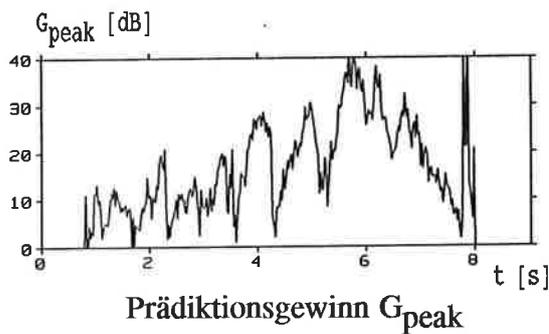
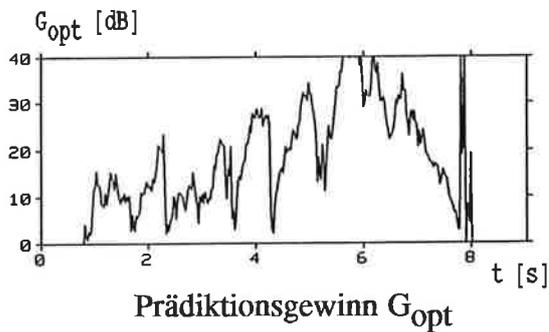
optimale Verzögerung d ($0 \leq d \leq d_{max} = 16$) / optimale Prädiktionsrichtung / Blocklänge $N = 36$

Quantisierung der Prädiktorkoeffizienten: gleichförmig im Bereich $[-2,2]$ mit 64 Stufen

Die bei G_{quant} eingezeichnete Gerade gibt die ungefähr benötigte Seiteninformation SI [dB] bei der gewählten Koeffizientenquantisierung mit 64 Stufen (6 Bit) an.
(Links: $SI = 11 \text{ Bit} = 1,8 \text{ dB}$ / Rechts: $SI = 29 \text{ Bit} = 4,8 \text{ dB}$)

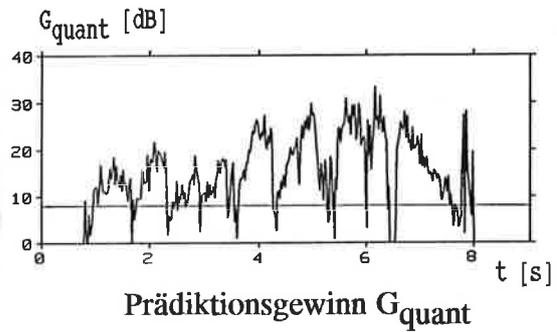
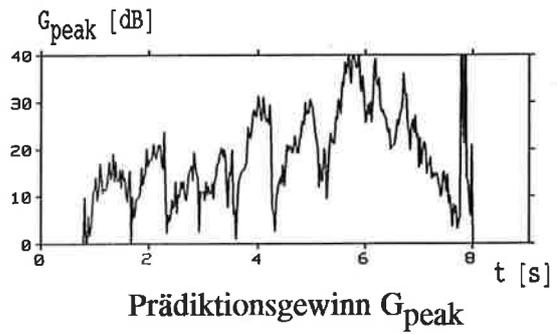
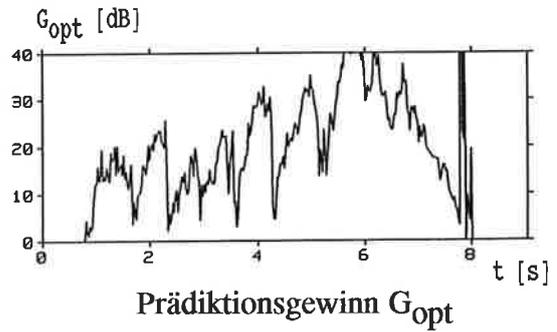
Tonsignal "Klarinette"

Teilband 2
(Prädiktorgrad $p = 5$)
(feste Verzögerung, $d = 0$)



Tonsignal "Klarinette"

Teilband 2
(Prädiktorgrad $p = 7$)
(feste Verzögerung, $d = 0$)



feste Verzögerung $d = 0$ / optimale Prädiktionsrichtung / Blocklänge $N = 36$

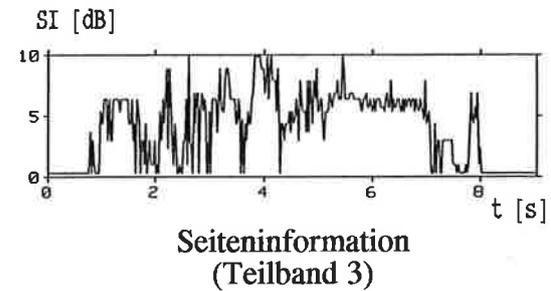
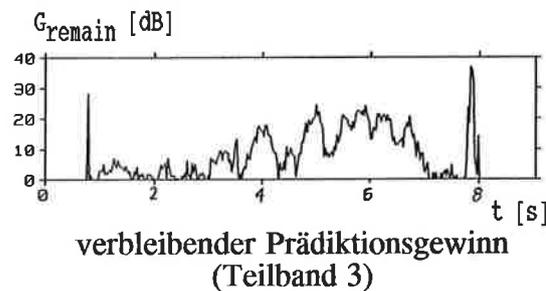
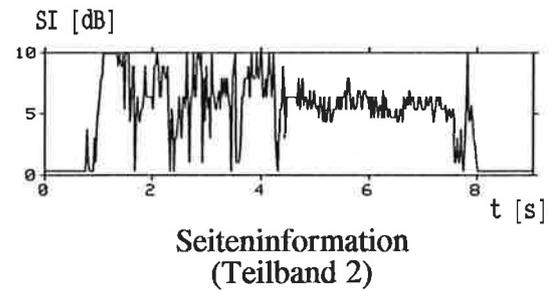
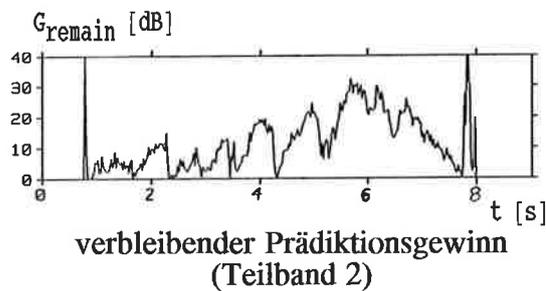
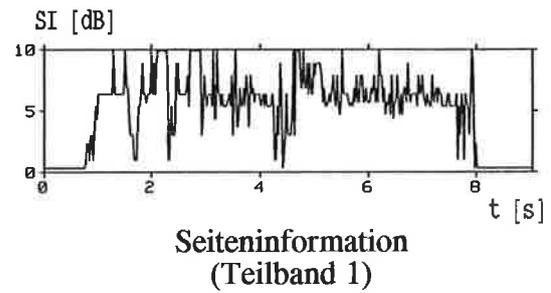
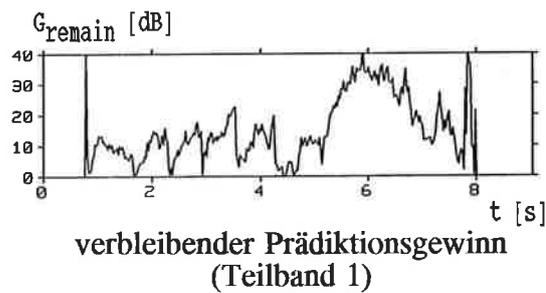
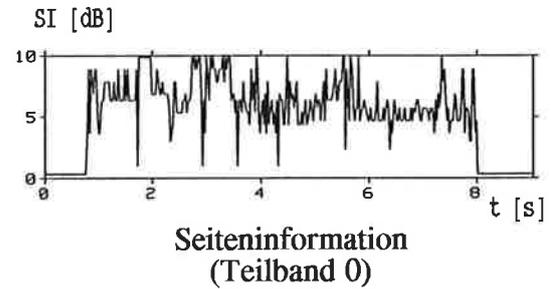
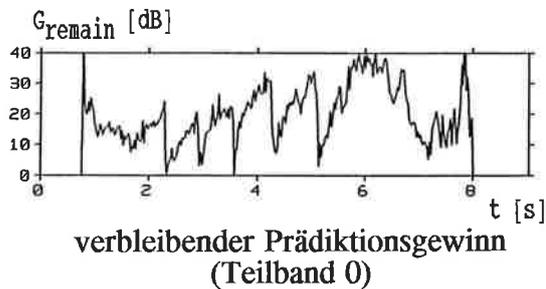
Quantisierung der Prädiktorkoeffizienten: gleichförmig im Bereich $[-2,2]$ mit 64 Stufen

Die bei G_{quant} eingezeichnete Gerade gibt die ungefähr benötigte Seiteninformation SI [dB] bei der gewählten Koeffizientenquantisierung mit 64 Stufen (6 Bit) an.
(Links: $SI = 37$ Bit = 6,2 dB / Rechts: $SI = 49$ Bit = 8,2 dB)

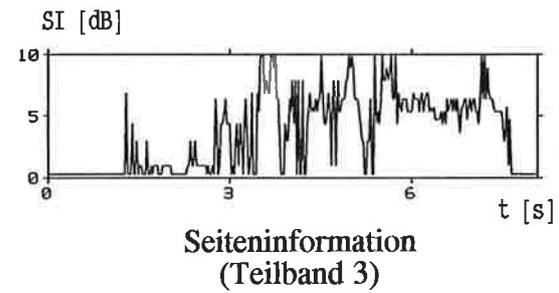
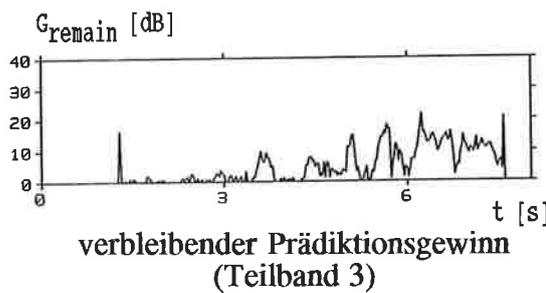
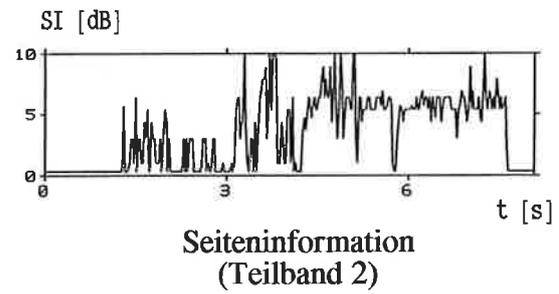
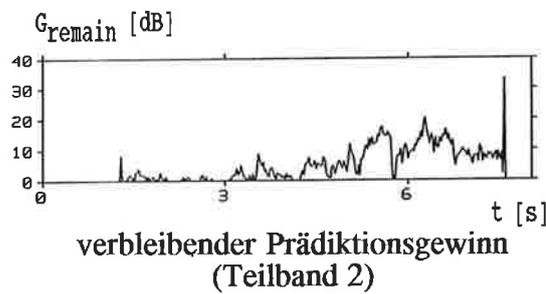
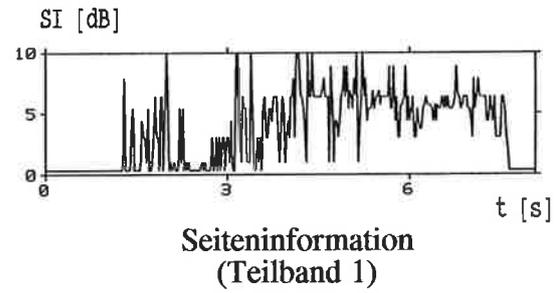
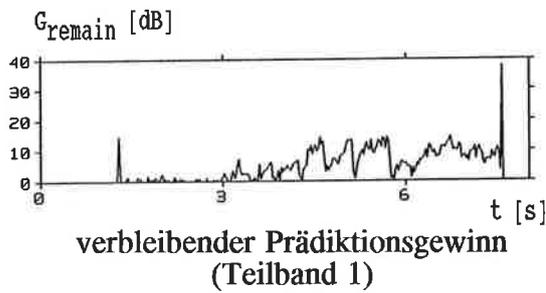
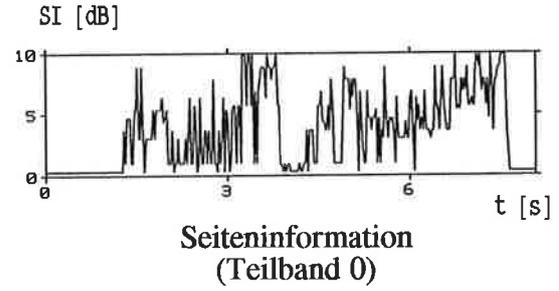
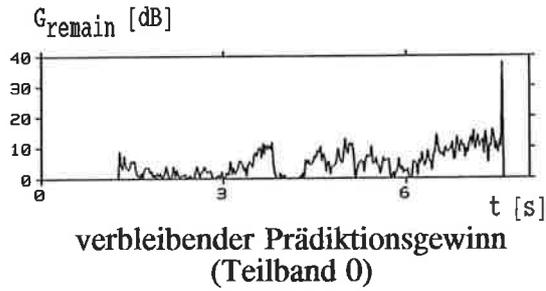
A.6 Prädiktionsgewinn und Seiteninformation (Diagramme)

A.6.1 Verbleibender Prädiktionsgewinn & Seiteninformation vs. Zeit

Tonsignal "Klarinette"

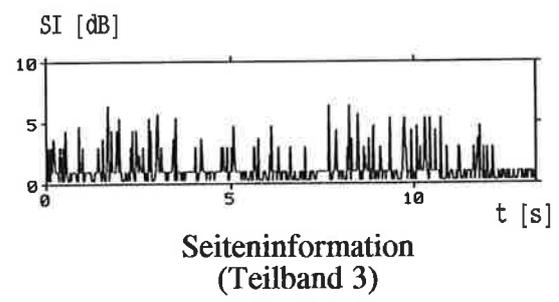
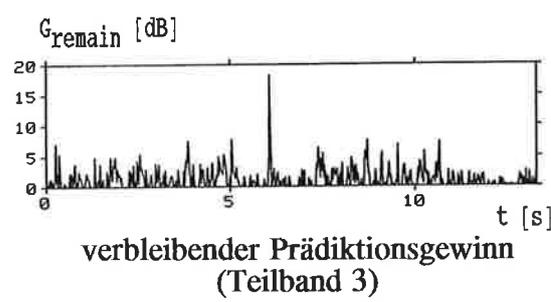
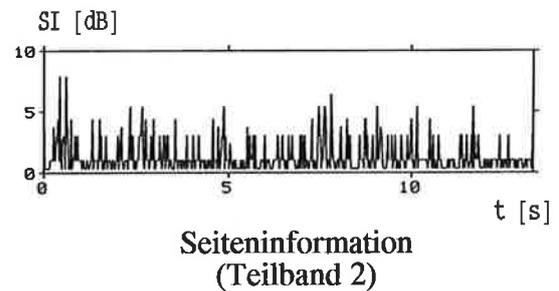
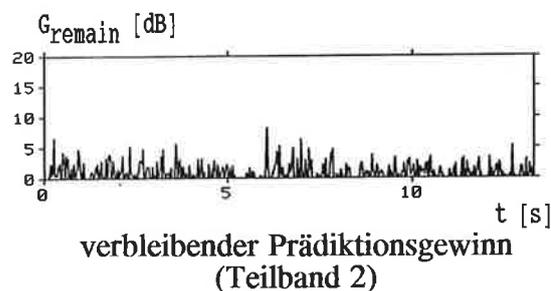
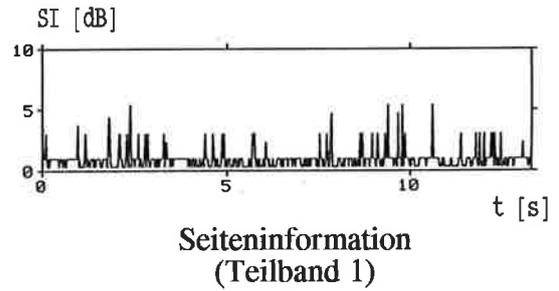
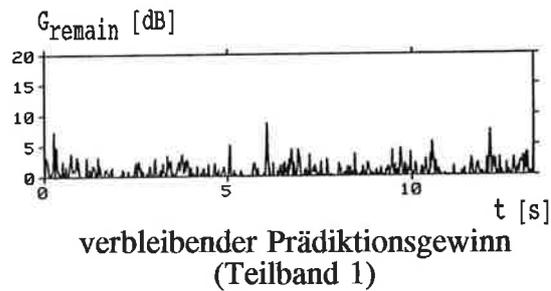
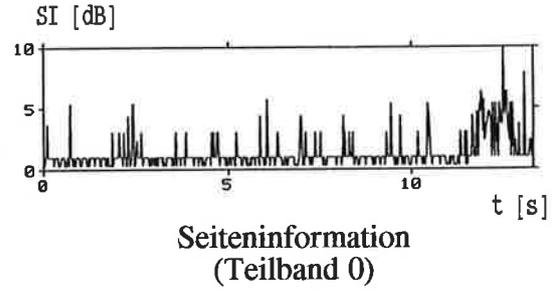
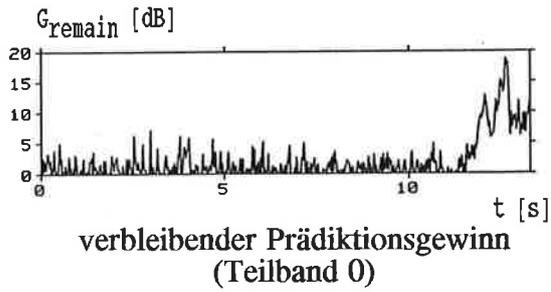


Tonsignal "Cembalo"



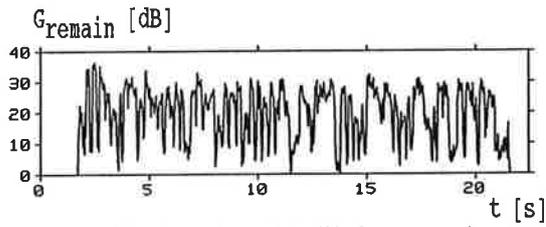
optimaler Prädiktormodus / optimale Prädiktionsrichtung / Blocklänge N = 36

Tonsignal "Stravinsky"

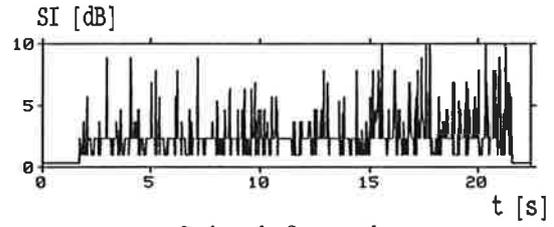


optimaler Prädiktormodus / optimale Prädiktionsrichtung / Blocklänge $N = 36$

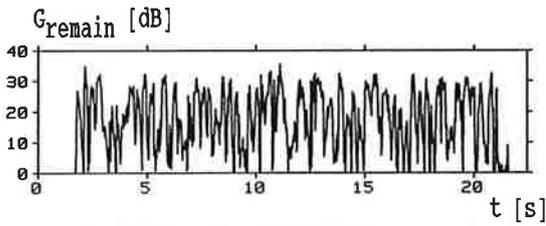
Tonsignal "Suzanne Vega"



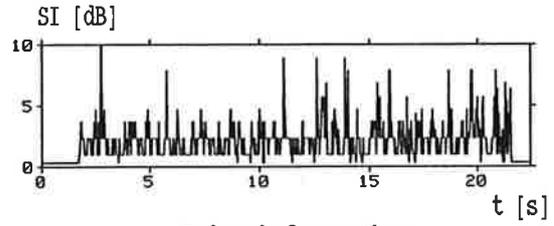
**verbleibender Prädiktionsgewinn
(Teilband 0)**



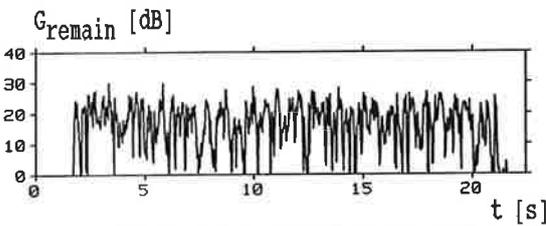
**Seiteninformation
(Teilband 0)**



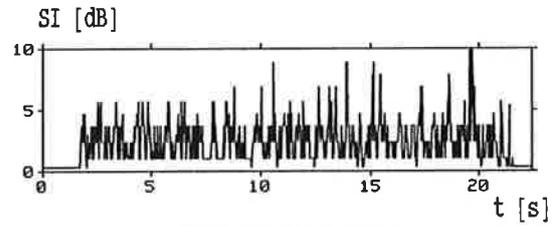
**verbleibender Prädiktionsgewinn
(Teilband 1)**



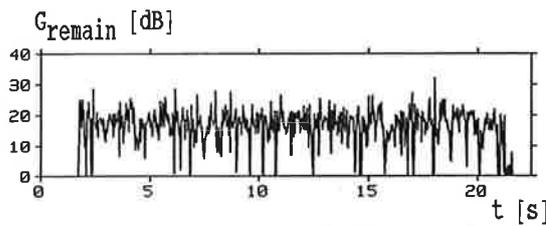
**Seiteninformation
(Teilband 1)**



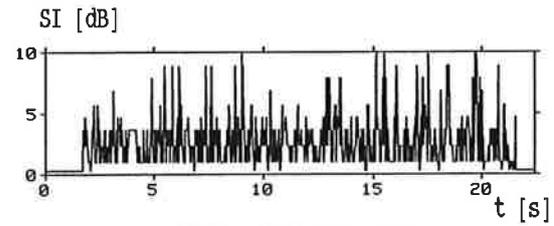
**verbleibender Prädiktionsgewinn
(Teilband 2)**



**Seiteninformation
(Teilband 2)**



**verbleibender Prädiktionsgewinn
(Teilband 3)**

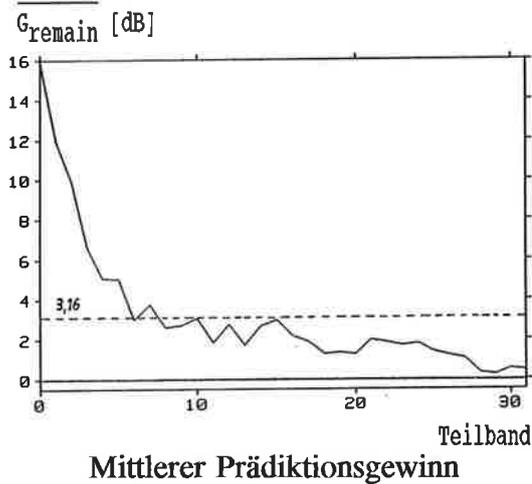


**Seiteninformation
(Teilband 3)**

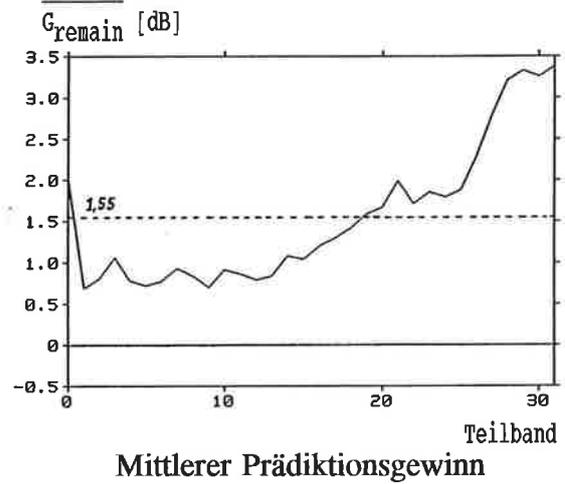
optimaler Prädiktormodus / optimale Prädiktionsrichtung / Blocklänge $N = 36$

A.6.2 Mittlerer verbleibender Prädiktionsgewinn vs. Teilband

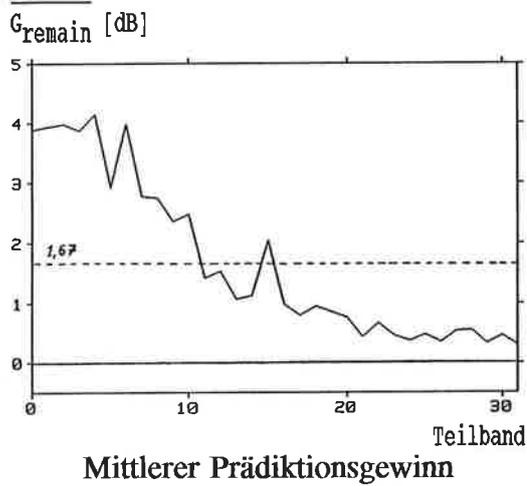
Tonsignal "Klarinette"



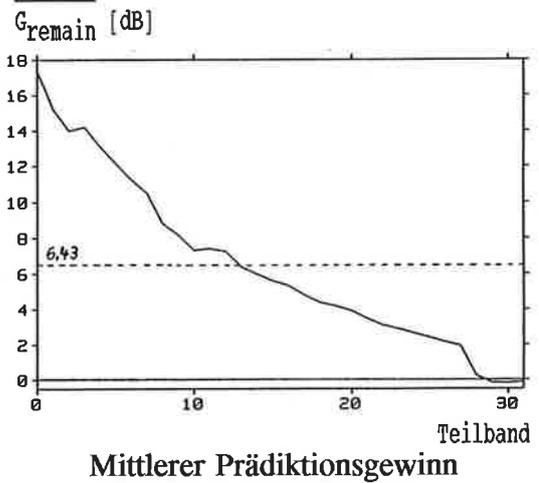
Tonsignal "Stravinsky"



Tonsignal "Cembalo"



Tonsignal "Suzanne Vega"



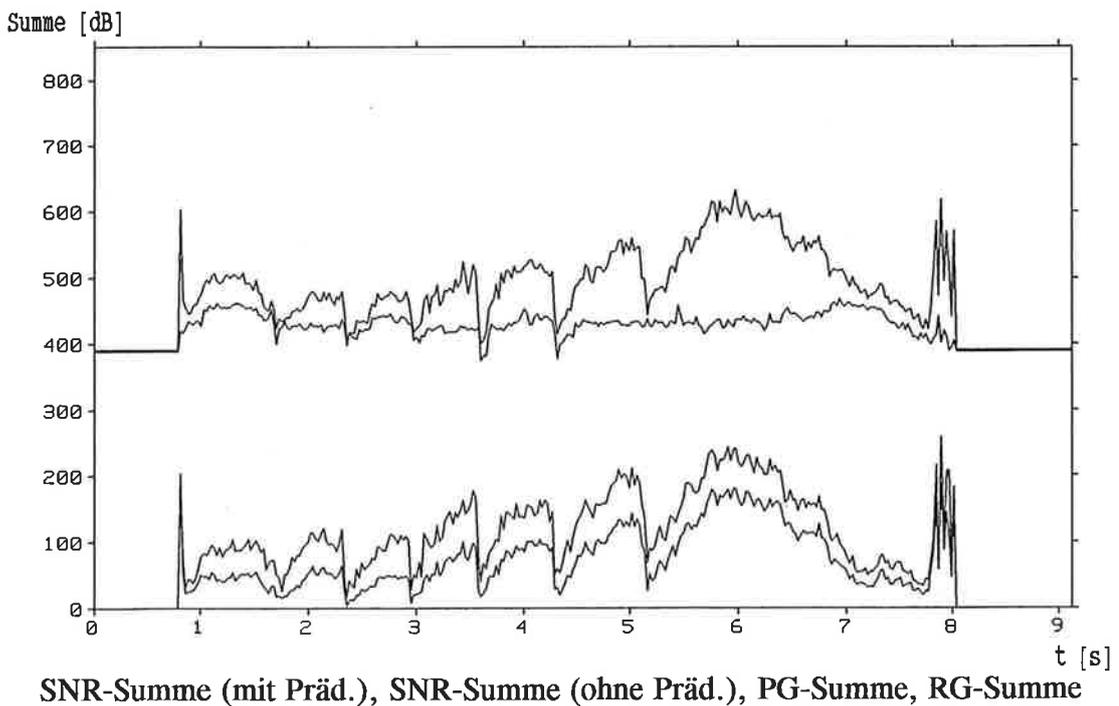
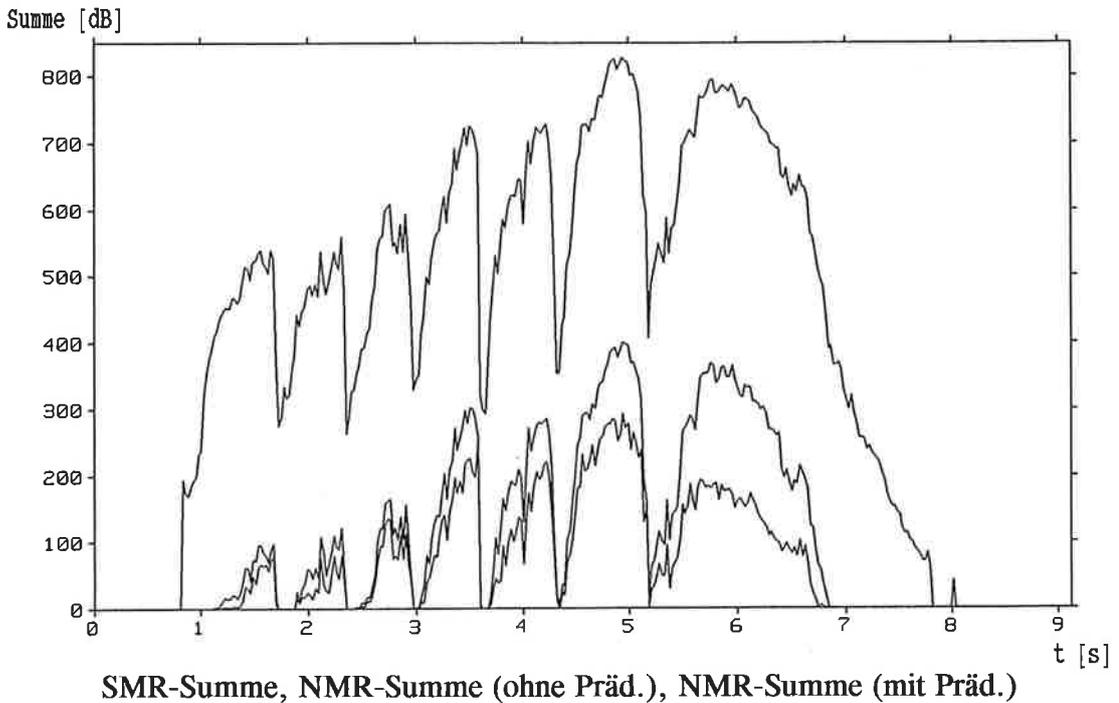
optimaler Prädiktormodus / optimale Prädiktionsrichtung / Blocklänge $N = 36$

Die G_{remain} in dB werden über die gesamte Signaldauer gemittelt.
Die gestrichelte Linie gibt den Mittelwert von G_{remain} für die 32 Teilbänder an.

A.7 Objektive Qualität der codierten Tonsignale (Diagramme)

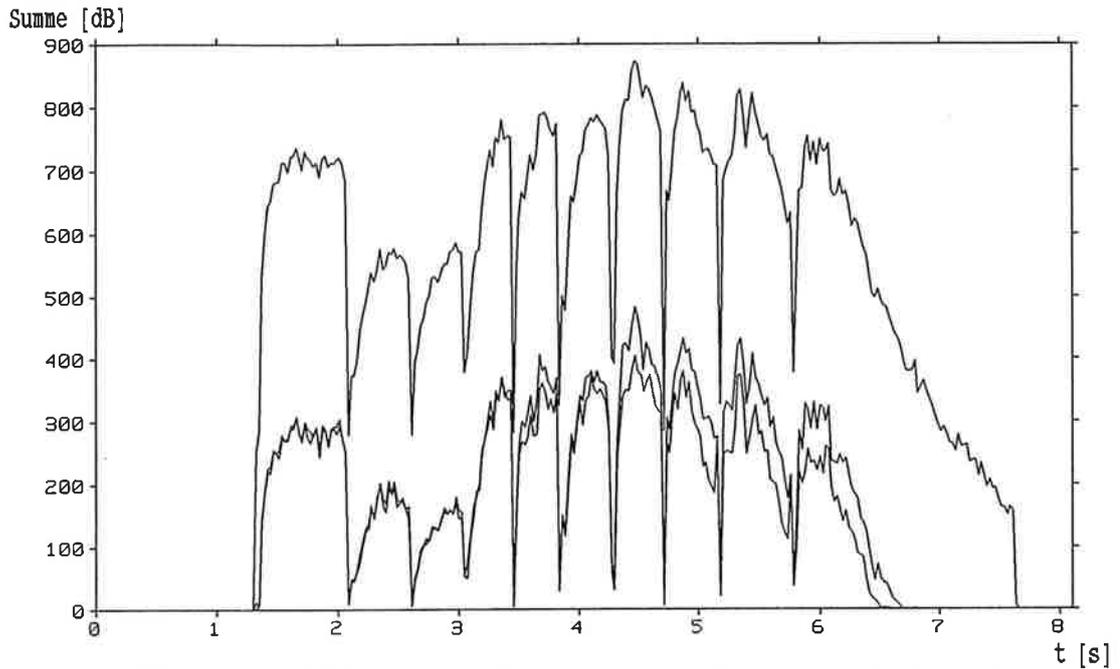
A.7.1 Objektive Qualität mit und ohne Interkanal-Prädiktion

Tonsignal "Klarinette" (Layer II, 128 kbit/s)

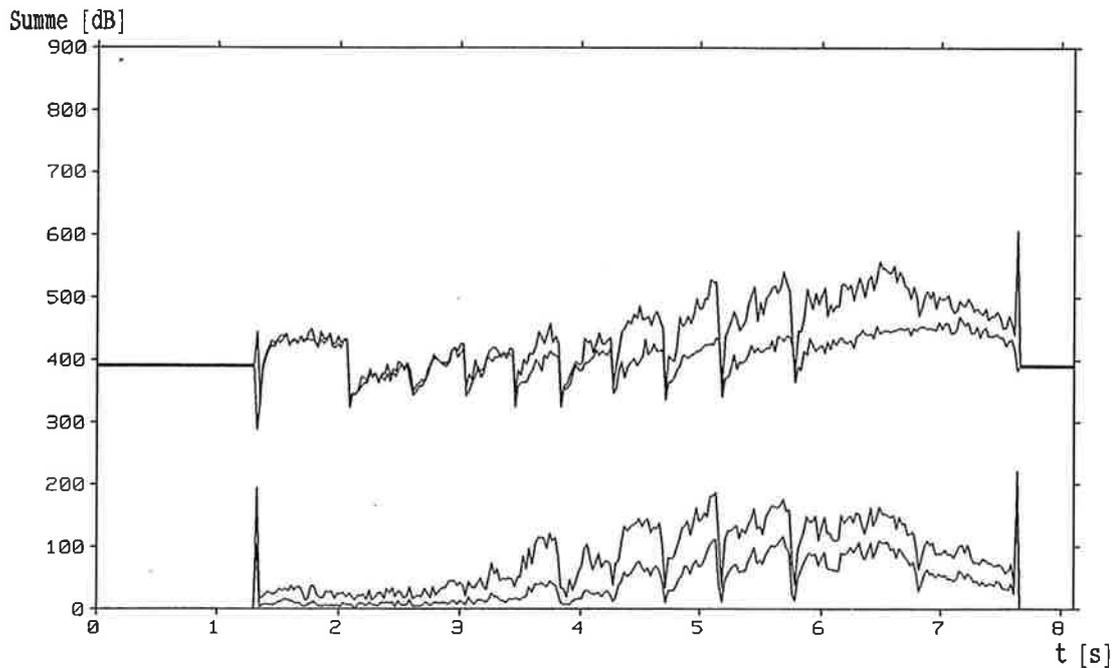


Die Kurven werden in dieser Reihenfolge angegeben: oberste Kurve, ... , unterste Kurve

Tonsignal "Cembalo" (Layer II, 128 kbit/s)



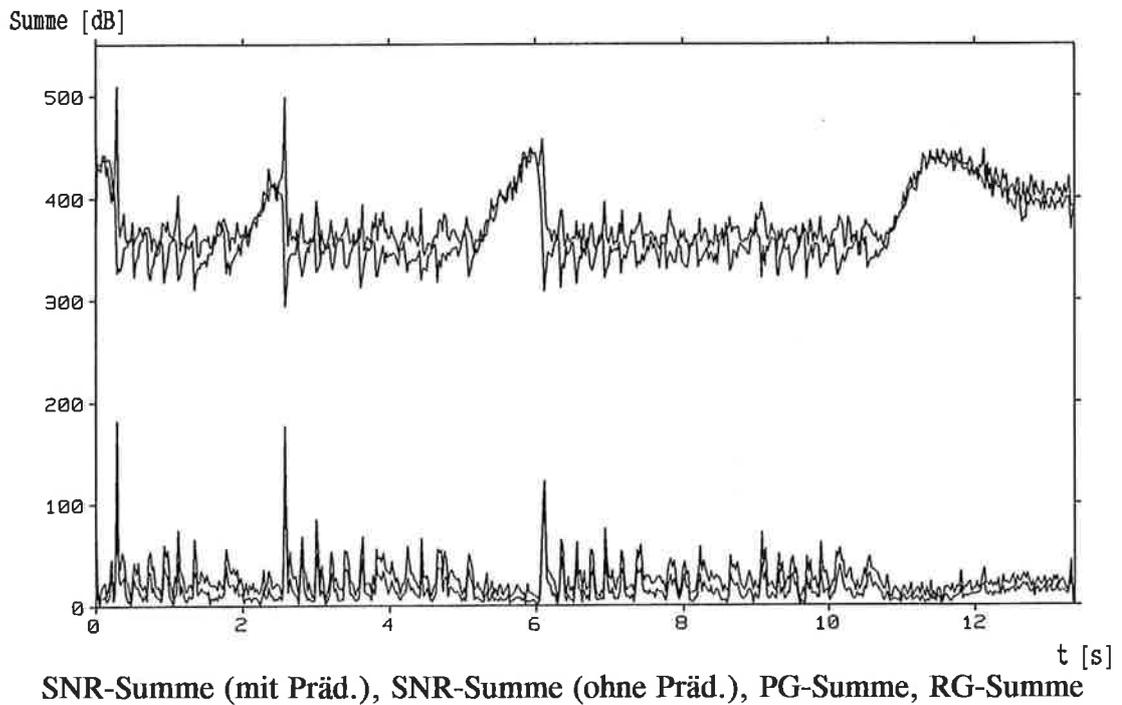
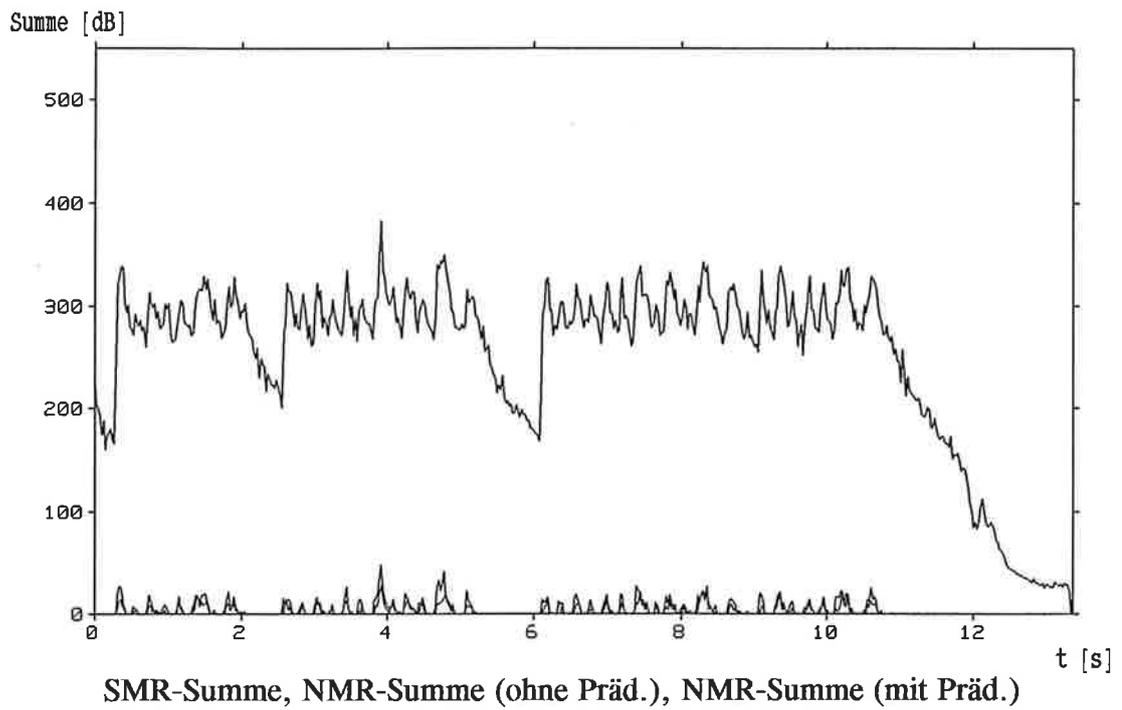
SMR-Summe, NMR-Summe (ohne Präd.), NMR-Summe (mit Präd.)



SNR-Summe (mit Präd.), SNR-Summe (ohne Präd.), PG-Summe, RG-Summe

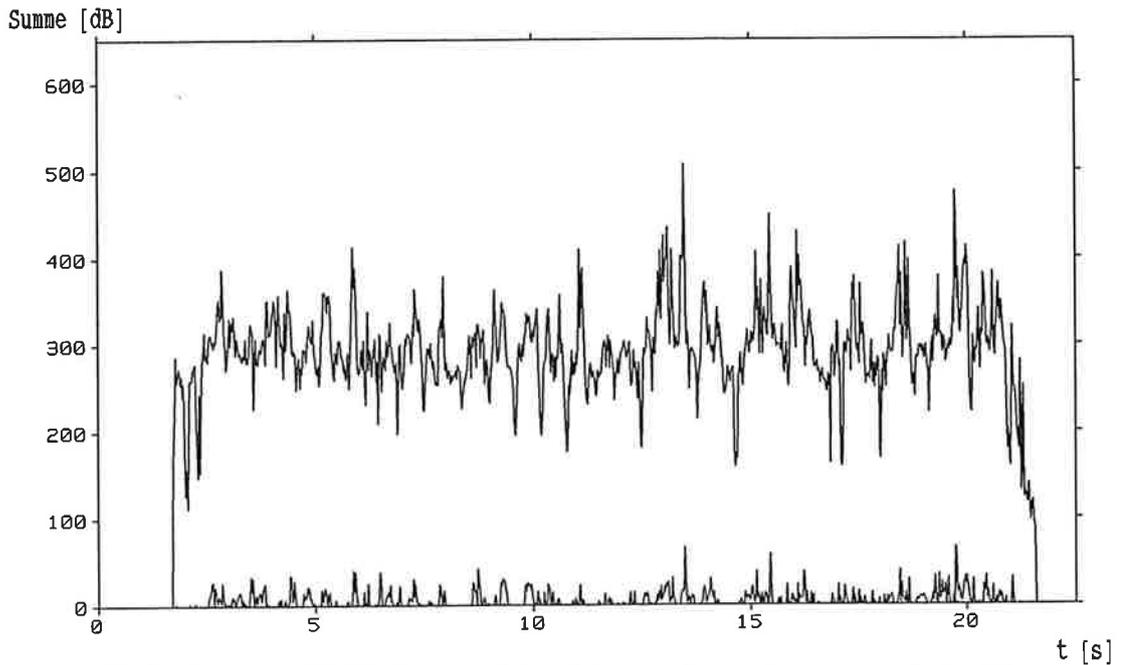
Die Kurven werden in dieser Reihenfolge angegeben: oberste Kurve, ... , unterste Kurve

Tonsignal "Stravinsky" (Layer II, 128 kbit/s)

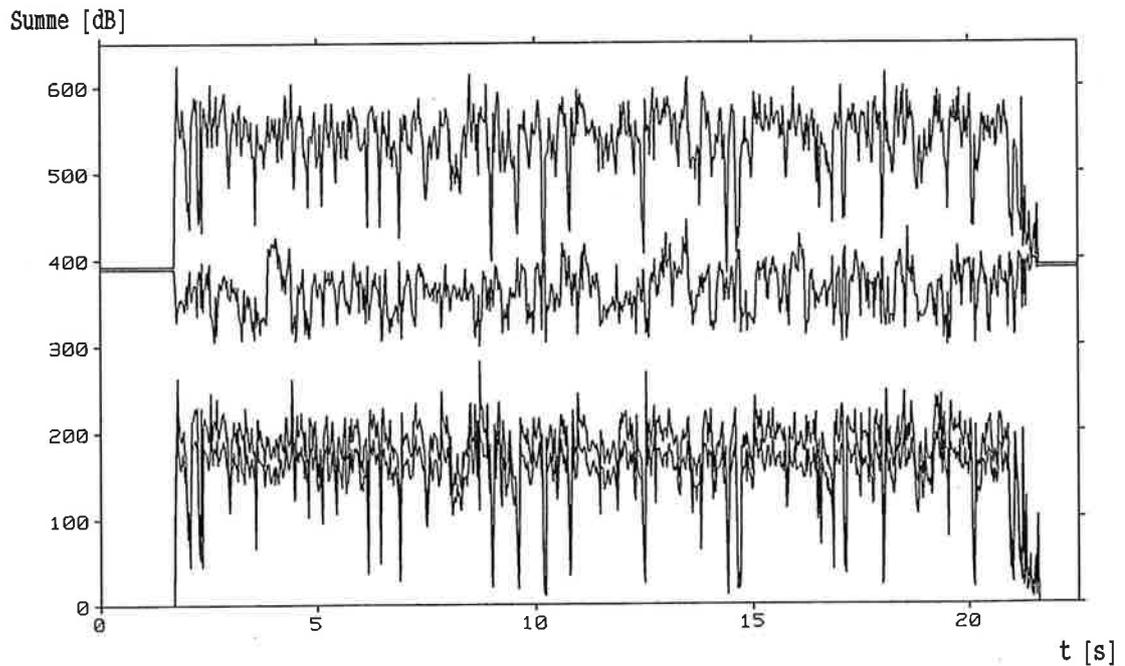


Die Kurven werden in dieser Reihenfolge angegeben: oberste Kurve, ... , unterste Kurve

Tonsignal "Suzanne Vega" (Layer II, 128 kbit/s)



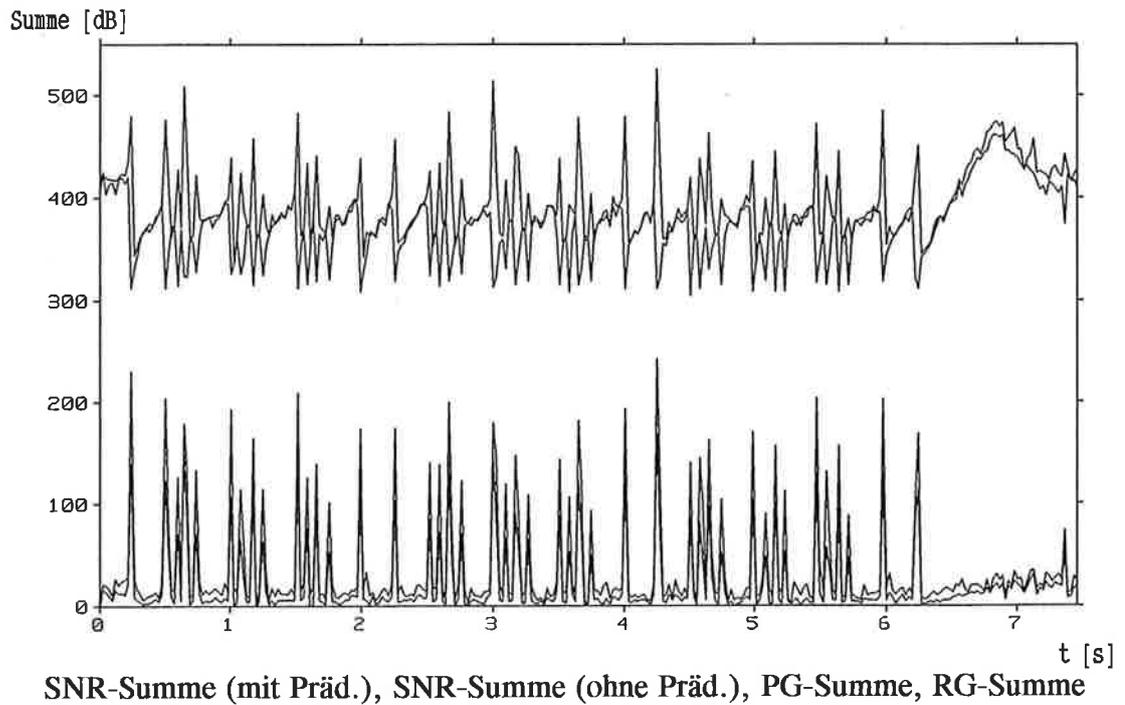
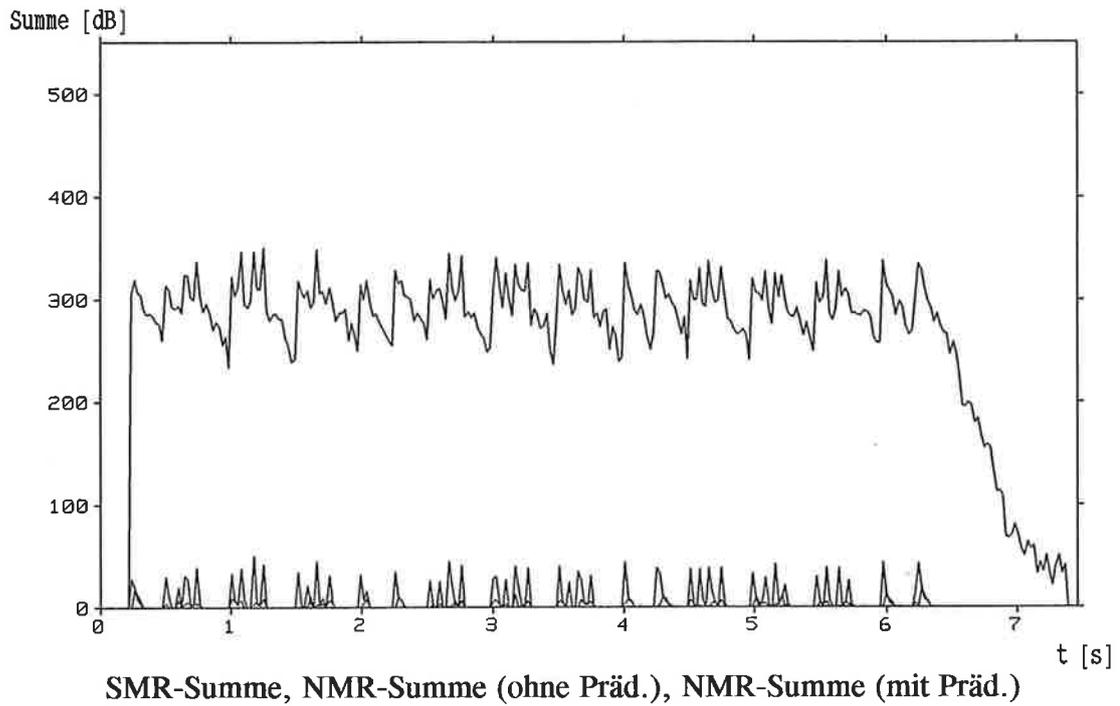
SMR-Summe, NMR-Summe (ohne Präd.), NMR-Summe (mit Präd.)



SNR-Summe (mit Präd.), SNR-Summe (ohne Präd.), PG-Summe, RG-Summe

Die Kurven werden in dieser Reihenfolge angegeben: oberste Kurve, ... , unterste Kurve

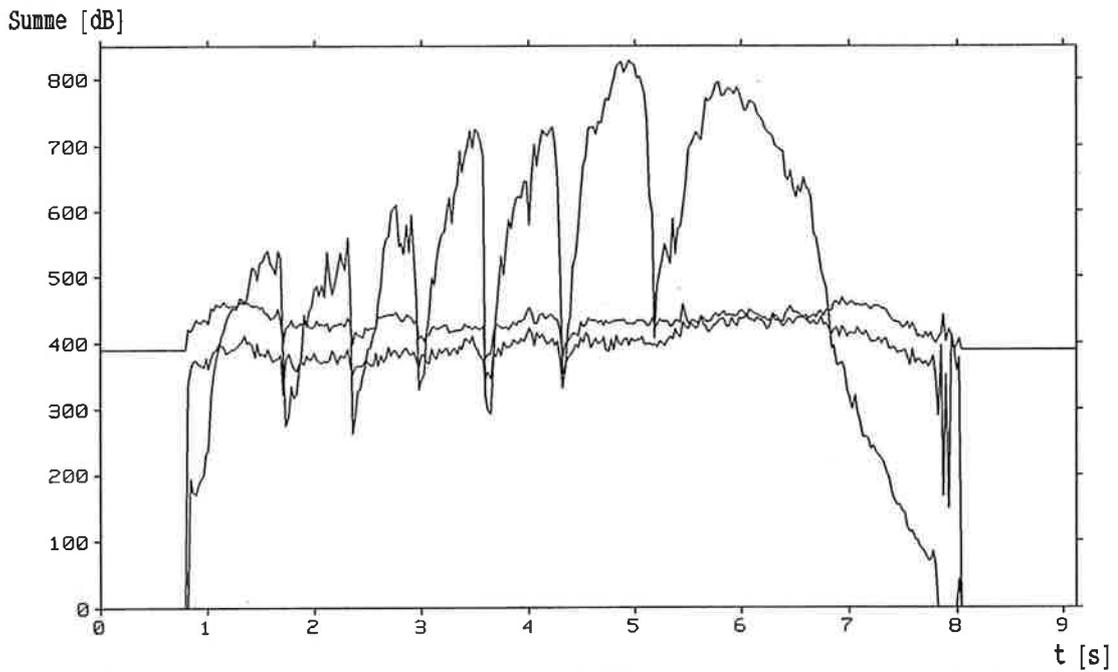
Tonsignal "Kastagnetten" (Layer II, 128 kbit/s)



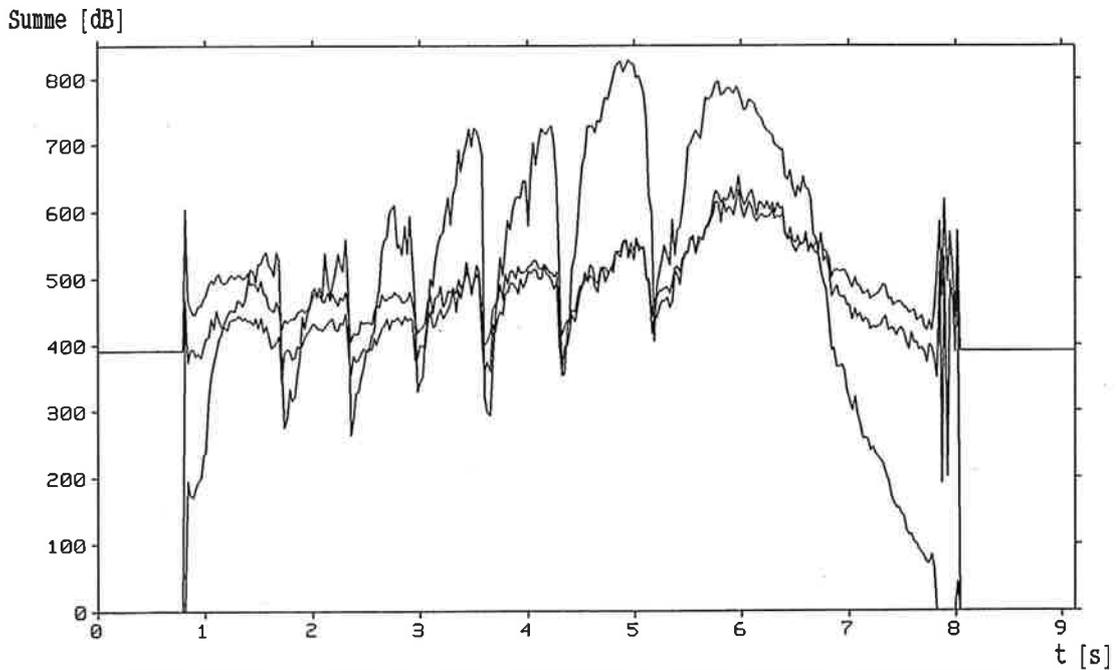
Die Kurven werden in dieser Reihenfolge angegeben: oberste Kurve, ... , unterste Kurve

A.7.2 Vergleich von SNR-Summe und SQR-Summe

Tonsignal "Klarinette" (Layer II, 128 kbit/s)



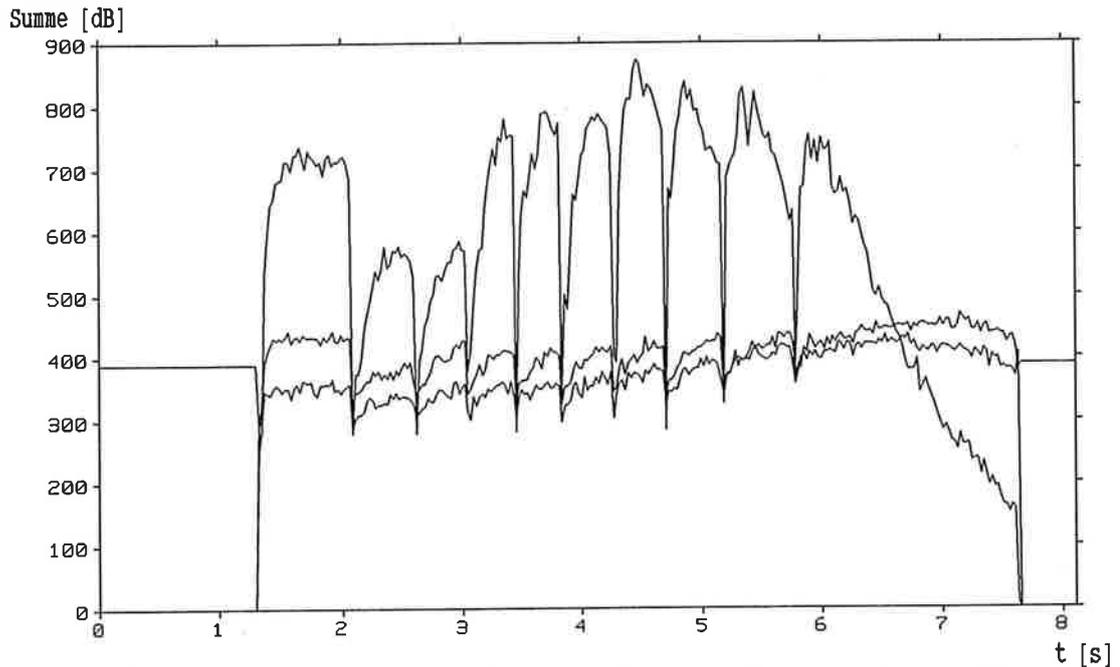
SMR-Summe, SNR-Summe (ohne Präd.), SQR-Summe (ohne Präd.)



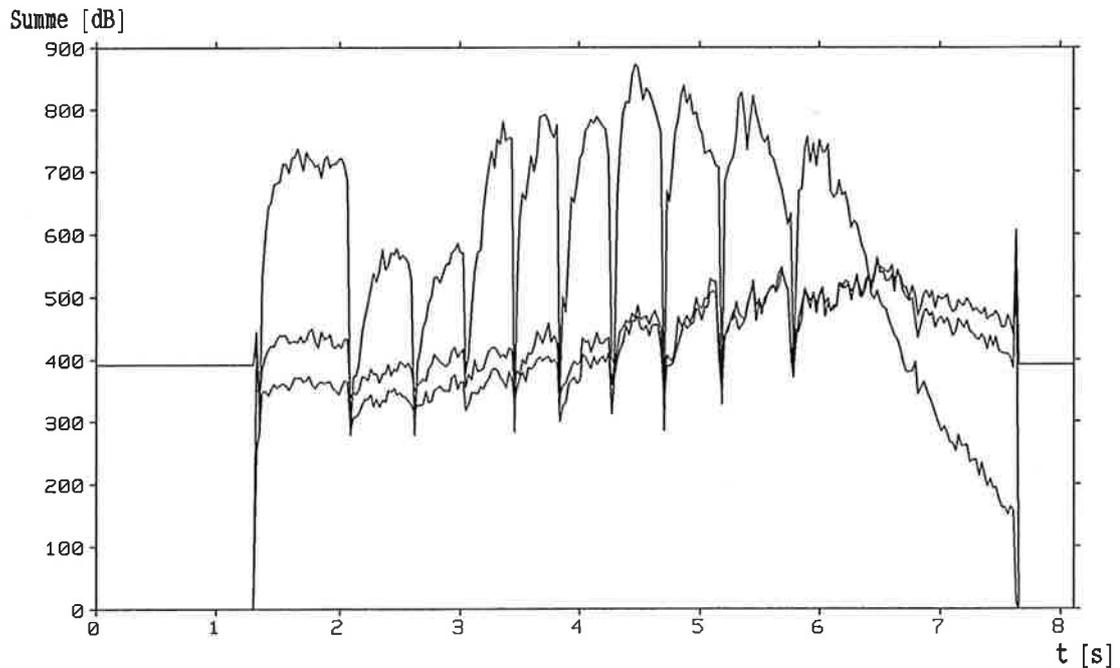
SMR-Summe, SNR-Summe (mit Präd.), SQR-Summe (mit Präd.)

Die Kurven werden in dieser Reihenfolge angegeben: oberste Kurve, ... , unterste Kurve

Tonsignal "Cembalo" (Layer II, 128 kbit/s)



SMR-Summe, SNR-Summe (ohne Präd.), SQR-Summe (ohne Präd.)



SMR-Summe, SNR-Summe (mit Präd.), SQR-Summe (mit Präd.)

Die Kurven werden in dieser Reihenfolge angegeben: oberste Kurve, ... , unterste Kurve

A.7.3 Codierung eines Blocks des Tonsignals "Klarinette"

Layer II ohne Interkanal-Prädiktion

```

{ 241}
noisy_sbs=30
sb 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
sig0 76 73 66 54 55 40 38 28 30 26 30 17 18 17 17 18 11 4 8 13 4 4 9 8 7
sig1 78 74 68 56 54 45 29 35 30 19 30 11 16 14 17 20 14 10 10 11 8 0 -1 8 3
smr0 68 44 45 20 35 32 34 26 27 23 25 10 9 3 0 -2 --- --- --- --- --- --- ---
smr1 68 48 46 27 33 34 27 32 27 15 25 5 6 1 0 0 -8 --- --- --- --- --- --- ---
snr0 50 32 32 7 21 16 21 11 11 7 7 0 0 0 0 0 0 0 0 0 0 0 0 0 0
snr1 56 32 32 11 21 21 11 16 11 0 7 0 0 0 0 0 0 0 0 0 0 0 0 0 0
sqr0 50 31 30 8 18 16 20 13 12 9 11 0 0 0 0 0 0 0 0 0 0 0 0 0 0
sqr1 55 30 30 11 20 19 13 16 16 0 11 0 0 0 0 0 0 0 0 0 0 0 0 0 0
qnr0 0 1 2 -1 3 0 1 -2 -1 -2 -4 0 0 0 0 0 0 0 0 0 0 0 0 0 0
qnr1 1 1 2 0 1 1 -2 0 -5 0 -4 0 0 0 0 0 0 0 0 0 0 0 0 0 0
nmr0 18 13 13 13 14 16 13 15 16 16 18 10 9 3 0 -2 --- --- --- --- --- --- ---
nmr1 12 17 15 16 12 13 16 16 16 15 18 5 6 1 0 0 -8 --- --- --- --- --- --- ---
bal0 7 4 4 1 4 3 4 2 2 1 1 - - - - - - - - - - - - - - -
bal1 8 4 4 2 4 4 2 3 2 - 1 - - - - - - - - - - - - - - -
    
```

Layer II mit Interkanal-Prädiktion

```

{ 241}
old tot snr= 614.85   old aver nmr= 7.00
eba0 9 5 5 2 6 5 5 4 4 3 3 1 - - - - - - - - - - - - - - -
eba1 9 6 5 4 5 5 4 5 4 1 3 - - - - - - - - - - - - - - -
noisy_sbs=30
sb 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
sig0 76 73 66 54 55 40 38 28 30 26 30 17 18 17 17 18 11 4 8 13 4 4 9 8 7
sig1 78 74 68 56 54 45 29 35 30 19 30 11 16 14 17 20 14 10 10 11 8 0 -1 8 3
smr0 68 44 45 20 35 32 34 26 27 23 25 10 9 3 0 -2 --- --- --- --- --- --- ---
smr1 68 48 46 27 33 34 27 32 27 15 25 5 6 1 0 0 -8 --- --- --- --- --- --- ---
snr0 66 30 38 18 29 21 25 19 22 16 16 7 0 0 0 0 0 0 0 0 0 0 0 0 0
snr1 62 38 40 21 25 23 18 25 21 12 16 4 0 0 0 0 0 0 0 0 0 0 0 0 0
sqr0 65 32 37 20 32 18 23 21 26 16 18 10 0 0 0 0 0 0 0 0 0 0 0 0 0
sqr1 61 36 42 18 25 26 20 24 19 11 19 7 0 0 0 0 0 0 0 0 0 0 0 0 0
qnr0 1 -2 0 -2 -4 3 2 -2 -3 0 -2 -3 0 0 0 0 0 0 0 0 0 0 0 0 0
qnr1 1 2 -2 3 0 -3 -2 1 1 1 -3 -3 0 0 0 0 0 0 0 0 0 0 0 0 0
nmr0 2 15 7 2 6 11 9 6 5 7 9 3 9 3 0 -2 --- --- --- --- --- --- ---
nmr1 6 11 6 7 8 11 10 7 6 4 9 1 6 1 0 0 -8 --- --- --- --- --- --- ---
bal0 3 - 5 - 1 4 5 1 1 3 3 1 - - - - - - - - - - - - - - -
bal1 9 5 1 4 5 1 1 5 4 - - - - - - - - - - - - - - -
pmr0 27 11 45 2 13 32 34 13 12 23 25 10 9 3 0 -2 --- --- --- --- --- --- ---
pmr1 68 48 13 27 33 17 17 32 27 4 9 1 6 1 0 0 -8 --- --- --- --- --- --- ---
pg0 40 30 --- 18 22 --- --- 12 15 --- --- --- --- --- --- --- --- ---
pg1 --- --- 33 --- --- 16 11 --- --- 12 16 4 --- --- --- --- --- --- ---
m 8 8 5 7 7 9 3 3 6 5 7 3 - - - - - - - - - - - - - - -
pd 0 0 1 0 0 1 1 0 0 1 1 1 - - - - - - - - - - - - - - -
d 0 0 9 7 4 13 15 10 0 6 4 11 - - - - - - - - - - - - - - -
rg 35 24 28 13 16 10 8 9 11 7 10 1 - - - - - - - - - - - - - - -
opg 40 33 33 18 22 18 11 12 15 12 16 4 - - - - - - - - - - - - - - -
    
```

Blocknummer 241 (5,784 s - 5,808 s)

0 = Linker Kanal / 1 = Rechter Kanal

- sb = Teilbandnummer
- sig = Teilbandsignalpegel [dB]
- smr = Signal/Mask-Ratio [dB]
- snr = Signal/Noise-Ratio [dB]
- sqr = Signal/Quantization_Noise-Ratio [dB]
- qnr = snr - sqr
- nmr = smr - snr
- bal = Bitzuteilungs-Index (siehe A.1)

- eba = geschätzter Bitzuteilungs-Index
- pmr = snr - $G_{quant,dB}$ bei y-Kanal
- pg = Prädiktionsgewinn $G_{quant,dB}$
- m = Prädiktormodus pmode (siehe A.3)
- pd = Prädiktionsrichtung pdir (siehe A.3)
- d = Verzögerung d
- rg = verbleibender Präd.gewinn $G_{remain,dB}$
- opg = $G_{quant,dB}$ bei Bitzuteilung