

Fortschritt-Berichte VDI

VDI

Reihe 10

Informatik/
Kommunikation

Nr. 854

Dipl.-Ing. Stephan Preihs,
Hannover

Fehlerrobuste Audiocodierung mit geringer Latenz



Institut für Informationsverarbeitung
www.tnt.uni-hannover.de

Fehlerrobuste Audiocodierung mit geringer Latenz

Von der Fakultät für Elektrotechnik und Informatik
der Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des akademischen Grades

Doktor-Ingenieur

(abgekürzt: Dr.-Ing.)

genehmigte

Dissertation

von

Dipl.-Ing. Stephan Preihs

geboren am 02. September 1983 in Hannover.

2016

Referent: Prof. Dr.-Ing. J. Ostermann
Korreferent: Prof. Dr.-Ing. G. Schuller
Vorsitzender: Prof. Dr.-Ing. B. Rosenhahn
Tag der Promotion: 06.09.2016

Fortschritt-Berichte VDI

Reihe 10

Informatik/
Kommunikation

Dipl.-Ing. Stephan Preihs,
Hannover

Nr. 854

Fehlerrobuste
Audiocodierung
mit geringer Latenz



Institut für Informationsverarbeitung
www.tnt.uni-hannover.de

Preihs, Stephan

Fehlerrobuste Audiocodierung mit geringer Latenz

Fortschr.-Ber. VDI Reihe 10 Nr. 854. Düsseldorf: VDI Verlag 2017.

128 Seiten, 36 Bilder, 4 Tabellen.

ISBN 978-3-18-385410-3, ISSN 0178-9627,

€ 48,00/VDI-Mitgliederpreis € 43,20.

Für die Dokumentation: Audiocodierung mit geringer Latenz – fehlerrobuste Subband-ADPCM
– Filterbankdesign – dynamische Bit-Allokation – Audiocodec Optimierung

In der vorliegenden Arbeit wird das Ziel der Entwicklung eines Verfahrens zur fehler robusten Audiocodierung mit geringer Latenz verfolgt. Hierzu wird der erfolgversprechende Ansatz einer Subband-ADPCM um Maßnahmen zum Erreichen einer algorithmischen Fehlerrate hinzu ergänzt, die Codierstrukturen hinsichtlich einer Anpassung an wechselnde Signaleigenschaften erweitert und diese bezüglich ihrer Parameter global optimiert. Die Ermittlung der Leistungsfähigkeit des entwickelten Codierverfahrens und der Vergleich mit dem Stand der Technik erfolgt auf Basis des PEAQ-Algorithmus. Die Ergebnisse werden zusätzlich durch einen Hörversuch mit im kritischen Hören geübten Probanden validiert. Gemäß der in dieser Arbeit erfolgten Untersuchungen ist es mit dem entwickelten fehler robusten Audiocodec möglich, bei moderaten Bitraten von bis zu 200 kbit/s und einer algorithmischen Latenz von nur 0,5 ms eine Audioqualität zu erreichen, die selbst erhöhten Ansprüchen gerecht wird.

Bibliographische Information der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet unter <http://dnb.ddb.de> abrufbar.

Bibliographic information published by the Deutsche Bibliothek

(German National Library)

The Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliographie
(German National Bibliography); detailed bibliographic data is available via Internet at
<http://dnb.ddb.de>.

© VDI Verlag GmbH · Düsseldorf 2017

Alle Rechte, auch das des auszugsweisen Nachdruckes, der auszugsweisen oder vollständigen Wiedergabe (Fotokopie, Mikrokopie), der Speicherung in Datenverarbeitungsanlagen, im Internet und das der Übersetzung, vorbehalten.

Als Manuskript gedruckt. Printed in Germany.

ISSN 0178-9627

ISBN 978-3-18-385410-3

Vorwort

Die vorliegende Arbeit entstand während meiner Zeit als Wissenschaftlicher Mitarbeiter am Institut für Informationsverarbeitung (TNT) der Gottfried Wilhelm Leibniz Universität Hannover.

Mein besonderer Dank gilt meinem Doktorvater Prof. Dr.-Ing. Jörn Ostermann, der durch meine Einstellung am Institut, die Betreuung der Dissertation und die Unterstützung bei der Bearbeitung von Projekten sowie durch angeregte Diskussionen und kritisches Hinterfragen maßgeblich zum Gelingen dieser Arbeit beigetragen hat. Die Vielzahl an Entfaltungsmöglichkeiten, die über das rein wissenschaftliche Arbeiten hinausgingen sowie das stets entgegengebrachte Vertrauen und Verständnis wusste ich jederzeit sehr zu schätzen.

Mein Dank gilt auch Herrn Prof. Dr.-Ing. Gerald Schuller für die Annahme des Korreferates und das damit verbundene Anfertigen eines Gutachtens. Des Weiteren möchte ich Herrn Prof. Dr.-Ing. Bodo Rosenhahn für die Übernahme des Prüfungsvorsitzes, das Leiten des Promotionsverfahrens sowie für den unkomplizierten Umgang bei der Betreuung von Vorlesungen, Laboren oder institutsinternen Angelegenheiten danken.

Ein großer Dank geht an die Kollegen des TNT. Die durchweg gute Atmosphäre am Institut mit stets geöffneten Türen und Ohren und die vielfältigen Möglichkeiten des fachlichen und privaten Austausches haben mich immer motiviert. Besonders erwähnen möchte ich dabei die „Freunde aus dem eNIFE“ Björn Scheuermann, Florian Baumann, Arne Ehlers, Kai Cordes und Matthias Reso. Den aufeinander folgenden Bürokollegen Gerard Pons-Moll und besonders Thomas Krause möchte ich für die Bereicherung des Büroalltages durch angeregte Diskussionen, Hilfe bei fachlichen Fragestellungen und die aufmunternden Worte in schwierigen Zeiten danken.

Mein aufrichtiger Dank gebührt auch den Kolleginnen und Kollegen aus dem Sekretariat und der Administration des TNT namentlich Pia Bank, Hilke Brodersen und Doris Jaspers-Göring sowie Martin Pahl und Thomas Wehberg. Die hervorragende Unterstützung bei der Koordination von Institutsangelegenheiten, Planung und Durchführung von Dienstreisen und Bewältigung von vertraglichen Belangen hat mir immer sehr geholfen. Zusätzlich danke ich Herrn Matthias Schuh für die Unterstützung bei technischen Angelegenheiten und der Beschaffung von Hardware sowie den dabei herrschenden ungezwungenen Umgang.

Bedanken möchte ich mich auch bei den zahlreichen Studenten, die im Rahmen von Abschlussarbeiten und Tätigkeiten als hilfswissenschaftliche Kräfte durch ihre engagierte und zuverlässige Zusammenarbeit viele Beiträge zu meinen Forschungsaktivitäten geleistet haben. Mein Dank gilt zudem den Probanden des Hörversuches für ihre Bereitschaft zur Teilnahme sowie die Zeit und Geduld bei dessen Durchführung. Herrn Nikolaus Meine danke ich zusätzlich für die Bereitstellung des Quellcodes für die Benutzeroberfläche zur Realisierung der Hörversuche.

Ein großer Dank gilt meinen Eltern Ruth und Wolfgang Preihs. Ohne die Unterstützung im beruflichen und privaten Werdegang und die Rückendeckung in allen Lebenslagen wäre mein Studium und die Promotion nicht denkbar gewesen.

Mein spezieller Dank gilt meiner Frau Katha, die mir immer den Rücken frei hält und mit ihrer Geduld, dem vielen guten Zureden sowie dem akribischen Korrekturlesen der Arbeit deren Fertigstellung überhaupt erst ermöglicht hat.

für Oma Henny (1916-2016)

Inhaltsverzeichnis

Abkürzungen und Formelzeichen	VI
Kurzfassung/Abstract	XII
1 Einleitung	1
1.1 Stand der Technik: Audiodcodierung mit geringer Latenz	4
1.2 Ansatz ADPCM-basierter Subband-Codierung	7
1.3 Ziele, Beiträge und Aufbau der Arbeit	8
2 Filterbänke mit geringer Latenz	10
2.1 Filterbänke in der Audiodcodierung	10
2.2 <i>M</i> -Kanal Filterbankentwurfsverfahren	13
2.3 Vergleich verschiedener Filterbankentwürfe	21
2.4 Globale Optimierung von Entwurfsparametern	26
3 Fehlerrobuste ADPCM-Codierung der Subband-Signale	30
3.1 Grundlagen zur ADPCM	30
3.2 Fehlerrobuste adaptive Prädiktion	35
3.3 Fehlerrobuste adaptive Optimalquantisierung	45
3.4 Maßnahme zur Verbesserung der Subband-Codierung	64
4 Audiocodec Optimierung	72
4.1 Audiocodec Bewertungsgrößen	72
4.2 Methode zur globalen Parameteroptimierung	75
4.3 Ergebnisse von globalen Optimierungen der Audiocodec Parameter .	79
5 Evaluierung	83
5.1 Objektive Ergebnisse	83
5.2 Ergebnisse subjektiver Tests	88
5.3 Einfluss von Kanalübertragungsfehlern	94
6 Zusammenfassung und Ausblick	98
A Anhang: Tabellarische Ergebnisse	102
A.1 Ergebnisse der objektiven Evaluierung	102
A.2 Ergebnisse der subjektiven Evaluierung	105
Literaturverzeichnis	108

Abkürzungen und Formelzeichen

Abkürzungen

3D	<u>3-Dimensional</u>
AAC	<u>Advanced Audio Coding</u>
AD/DA	<u>Analog-Digital/Digital-Analog</u>
ADPCM	<u>Adaptive Differential Pulse Code Modulation</u>
AFB	<u>Analysefilterbank</u>
AKF	<u>Autokorrelationsfolge</u>
BER	<u>Bit Error Rate</u>
BP	<u>Bandpass</u>
CELP	<u>Code Excited Linear Prediction</u>
CELT	<u>Constrained Energy Lapped Transform</u>
dB	<u>dezi-Bel</u>
dBA	<u>dynamische Bit-Allokation</u>
De-MUX	<u>De-Multiplexer</u>
DI	<u>Distortion Index</u>
DSP	<u>Digital Signal Processor</u>
ELD	<u>Enhanced Low Delay</u>
EVS	<u>Enhanced Voice Service</u>
FFT	<u>Fast Fourier Transform</u>
FIR	<u>Finite Impulse Response</u>
FLAC	<u>Free Lossless Audio Codec</u>
GAL	<u>Gradient Adaptive Lattice</u>
GSM	<u>Global System for Mobile Communication</u>
HP	<u>Hochpass</u>
IETF	<u>Internet Engineering Task Force</u>
IIR	<u>Infinite Impulse Response</u>
ISO	<u>International Standardization Organization</u>
LSB	<u>Least Significant Bit</u>
MDCT	<u>Modifizierte Diskrete Cosinus Transformation</u>
MOV	<u>Model Output Variable</u>
MP3	<u>MPEG-I Audio Layer 3</u>
MPEG	<u>Moving Pictures Expert Group</u>
MSB	<u>Most Significant Bit</u>
MUX	<u>Multiplexer</u>

NMP	Network <u>M</u> usic <u>P</u> erformance
NMR	Noise to <u>M</u> ask <u>R</u> atio
ODG	Objective <u>D</u> ifference <u>G</u> rade
PEAQ	Perceptual <u>E</u> valuation of <u>A</u> udio <u>Q</u> uality
PQM F	Pseudo <u>Q</u> uadrature <u>M</u> irror <u>F</u> ilter
QMF	Quadrature <u>M</u> irror <u>F</u> ilter
segSNR	segmental <u>S</u> ignal to <u>NR</u> atio
SFB	Synthesefilterbank
SLQ	Sphärisch <u>L</u> ogarithmische <u>Q</u> uantisierung
SNR	Signal to <u>NR</u> atio
SNR _r	Signal-to-reconstruction <u>N</u> oise <u>R</u> atio
TP	Tiefpass
ULD	Ultra <u>L</u> ow <u>D</u> elay

Formelzeichen

Allgemeine Formelzeichen, Parameter und Bezeichner

A	A-Law Parameter
a_i	Prädiktionskoeffizient des Direktstruktur-Prädiktionsfilters
α	Leakage-Faktor im Lattice-Filter bzw. GAL-Algorithmus
$\alpha(n)$	Zeitkonstante der Pegelschätzung in der dynamischen Bit-Allokation
b	Vektor aus Ergebnissen der dynamischen Bit-Allokation
$b_m(n)$	Rückwärtsprädiktionsfehler in der m -ten Stufe des Lattice-Filters
β	Leakage-Faktor des GAL-Algorithmus
C_k	k -tes Codewort bzw. k -ter Repräsentativwert des Quantisierers
d_k	k -te Entscheidungsschwelle bzw. Maximal- und Minimalwert des Quantisierers
Δ_k	Breite der k -ten Stufe des skalaren Quantisierers
$e(n)$	Prädiktionsfehlersignal
$\tilde{e}(n)$	Signal des rekonstruierten Prädiktionsfehlers
$e_{\text{norm}}(n)$	Normalisiertes Prädiktionsfehlersignal
$\tilde{e}_{\text{norm}}(n)$	Rekonstruiertes normalisiertes Prädiktionsfehlersignal
$F(z)$	Übertragungsfunktion eines Direktstruktur-Prädiktionsfehlerfilters
$F_{\text{Alias.,max.}}$	Maximale Aliasingverzerrung einer Filterbank
$F_{\text{Ampl.,max.}}$	Maximale Amplitudenverzerrung einer Filterbank
$f_m(n)$	Vorwärtsprädiktionsfehler in der m -ten Stufe des Lattice-Filters
$G_i(z)$	Synthesefilter Übertragungsfunktion des i -ten Bandes
$g_i(n)$	Impulsantwort des i -ten Synthesefilters einer Filterbank
$G_{S,\min.}$	Miminale Sperrdämpfung der Synthesefilter einer Filterbank
$G_{S,\text{mtl.}}$	Mittlere Sperrdämpfung der Synthesefilter einer Filterbank

G_p	Prädiktionsgewinn
$H_i(z)$	Analysefilter Übertragungsfunktion des i -ten Bandes
$h_i(n)$	Impulsantwort des i -ten Analysefilters einer Filterbank
$H_{i,S,\text{mtl.}}$	Mittlere Sperrdämpfung des i -ten Analysefilters einer Filterbank
$H_{S,\text{min.}}$	Minimale Sperrdämpfung der Analysefilter einer Filterbank
$H_{S,\text{mtl.}}$	Mittlere Sperrdämpfung der Analysefilter einer Filterbank
i, j, k, l	Allgemeine Laufvariablen
J	Blockgröße der dynamischen Bit-Allokation
K_C	Anzahl der Codewörter und Repräsentativwerte des Quantisierers
$K(\cdot)$	Operator zum Berechnen des Kostenwertes der globalen Parameteroptimierung
κ	Parameter der Pegelreduktion in der dynamischen Bit-Allokation
k_m	Lattice-Koeffizient in der m -ten Stufe des Lattice-Filters
L	Länge des Eingangsblockes zur Lattice-Prädiktoradaption
$\lambda(n)$	Parameter für die Hüllkurvenschätzung zur Signalnormalisierung
λ_{AT}	Parameter für die Hüllkurvenschätzung im „Ansprech-Fall“
λ_{RT}	Parameter für die Hüllkurvenschätzung im „Rücklauf-Fall“
M	Anzahl der Frequenzbänder des Subband-Codecs bzw. der Filterbank sowie sich daraus ebenfalls ergebender Unterabtastungs-, Überabtastungs- und Kompensationsfaktor
m	Stufenindex des Lattice-Filters
$\downarrow M$	Unterabtastung mit dem Faktor M
$\uparrow M$	Überabtastung mit dem Faktor M
$\hat{\mu}$	Grundschriftweite im GAL-Algorithmus
N	Anzahl der Koeffizienten der Subband-Filter einer Filterbank
$P(z)$	Übertragungsfunktion eines Direktstruktur-Prädiktionsfilters
$p(n)$	Prädiktionssignal
$\mathbf{P}_a(z)$	Polyphasen-Matrix der Analysefilterbank beim Entwurfsverfahren nach Schuller
$p_{\Delta_n}(n)$	Zeitlich verschobene Impulsantwort des Prototypfilters beim Entwurfsverfahren nach Keiler/Nguyen
$p_l(n)$	Impulsantwort des linearphasigen Prototypfilters beim Entwurfsverfahren nach Keiler/Nguyen
$p_{\text{opt.}}(n)$	Numerisch optimierte Impulsantwort des Prototypfilters beim Entwurfsverfahren nach Keiler/Nguyen
$\mathbf{P}_s(z)$	Polyphasen-Matrix der Synthesefilterbank beim Entwurfsverfahren nach Schuller
ϕ	Parameter der Pegelreduktion in der dynamischen Bit-Allokation
p	Ordnung des Lattice- oder Direktstruktur-Filters
$Q(\cdot)$	Operator zum Durchführen der skalaren Quantisierung

$\hat{q}(n)$	Quantisierungsindex
$\hat{q}_i(m)$	Subband-Quantisierungsindex des i -ten Bandes
$Q^{-1}(\cdot)$	Operator zum Abbilden des Quantisierungsindexes auf den Repräsentativwert
$Q_a(\cdot)$	Operator der adaptiven skalaren Quantisierung
$Q_a^{-1}(\cdot)$	Operator zum Abbilden des Quantisierungsindexes auf den Repräsentativwert und zur anschließenden Denormalisierung
$r_k(n)$	k -ter Eintrag für die Lookup-Tabelle des adaptiven Quantisierers
$r(\cdot)$	Operator zum Zugriff auf die Einträge der Lookup-Tabelle des adaptiven Quantisierers
$r_{xx}(k)$	Autokorrelationsfolge
r_v	Periodenlänge der Prädiktionsfehlernormalisierungsfaktorübertragung bei der Breitband-Codierung
ρ	Parameter der Pegelreduktion in der dynamischen Bit-Allokation
$S(z)$	Übertragungsfunktion eines Direktstruktur-Synthesefilters
s	Verzögerungsparameter beim Filterbankentwurf nach Keiler/Nguyen
$s_i(n)$	Geschätzter Pegel des i -ten Subbandes in der dynamischen Bit-Allokation
$\sigma_m^2(n)$	(Geschätzte) Fehlerenergie der m -ten Stufe im Lattice-Filter
σ_{\min}^2	Begrenzung des Normalisierungsfaktors im GAL-Algorithmus
T	Anzahl der Signale eines Testsets für die globale Parameteroptimierung
τ	Gruppenlaufzeit einer Filterbank in Samples
$u(n)$	Gefensteretes Eingangssignal
$v(n)$	Normalisierungssignal der adaptiven Quantisierung
v_{\min}	Minimalwert für das Normalisierungssignal der adaptiven Quantisierung
$w(n)$	Fensterfunktion
ω	Normierte Kreisfrequenz
$\omega_{S,i}$	Als Sperrbereich definierter Teil der normierten Kreisfrequenz
X	Pegelreduktion in der Iterationsschleife der Bit-Allokation
χ	Parametervektor bei der globalen Parameteroptimierung
$x(n)$	Eingangssignal des Subband-Codecs bzw. der ADPCM oder exemplarisches Eingangssignal des Prädiktions- bzw. Lattice-Filters
\mathbf{x}	Allgemeiner Signal- oder Filterkoeffizientenvektor
\mathbf{x}_0	Initialer Filterkoeffizientenvektor beim Entwurfsverfahren nach Schuller
$x_i(m)$	Unterabgetastetes Subband-Signal des i -ten Bandes
$x_i(n)$	Subband-Signal des i -ten Bandes

$\tilde{x}(n)$	Ausgangssignal des Subband-Codecs bzw. der ADPCM sowie Rekonstruktionssignal im Encoder
$\tilde{x}_i(m)$	Rekonstruiertes unterabgetastetes Subband-Signal
ζ	Leakage-Faktor zum Erreichen der Fehlerrobustheit in der adaptiven Quantisierung

Formelzeichen des Filterbankentwurfsverfahrens nach Nayebi

A_0	Matrix mit Prototypen-Koeffizienten der Analysefilterbank
A_k	Matrix mit den Koeffizienten der Analysefilterbank in der k -ten Iteration
A	Matrix mit den Koeffizienten der Analysefilterbank
B	Matrix mit verschobener Antidiagonalmatrix zum Einstellen der gewünschten Gruppenlaufzeit des Filterbankentwurfs
δ_{A_k}	Matrix zur Modifikation der Analysefilterbank-Koeffizienten in der k -ten Iteration
E	Wert der Nayebi-Kostenfunktion
E_{global}	Wert der Kostenfunktion zur globalen Optimierung der Nayebi-Gewichtsparameter
$E_{\min.}$	Schwelle für die Nayebi-Kostenfunktion zum Abbruch der Optimierung
ϵ_j	Allgemeiner Fehler der Übertragungsfunktion der Subband-Filter
ϵ_{pb}	Fehler der Übertragungsfunktion eines Subband-Filters im Durchlassbereich
ϵ_r	Rekonstruktionsfehler der Filterbank
ϵ_{sb}	Fehler der Übertragungsfunktion eines Subband-Filters im Sperrbereich
S	Matrix mit Koeffizienten der Synthesefilterbank
S_{LS}	Matrix mit Koeffizienten einer Least Squares-Lösung für die Synthesefilterbank
w_j	Gewichtsfaktor der Nayebi-Kostenfunktion

Formelzeichen der adaptiven Vor- und Nachfilterung

$A_k(z)$	Übertragungsfunktion des k -ten Filters zur Allpasstransformation
G_k	Grenzfrequenz des k -ten Prototypfilters
$g_k(n)$	Verstärkungsfaktor des k -ten Filters der Vorfilterkaskade
$g_k^{-1}(n)$	Verstärkungsfaktor des k -ten Filters der Nachfilterkaskade
$\check{H}_{\text{bpk}}(z)$	Übertragungsfunktion des k -ten Bandpasses zur Leistungsschätzung
$\check{H}_k(z)$	Übertragungsfunktion des k -ten Bandes der Vorfilterkaskade
$\check{H}_{k,\text{prot.}}(z)$	Übertragungsfunktion des k -ten Prototypfilters
$\check{H}_{\text{pre}}(z)$	Resultierende Übertragungsfunktion der Vorfilterkaskade
K	Anzahl der Subbänder der Vor- und Nachfilterkaskade

λ_p	Glättungsparameter des rekursiven Filters zur Leistungsschätzung
ν	Parameter zum Abbilden auf den Verstärkungsfaktor
$\omega_{c,k}$	Mittenfrequenz des k -ten Bandes
$\omega_{l,k}$	Untere Grenzfrequenz des k -ten Bandes
$\omega_{u,k}$	Obere Grenzfrequenz des k -ten Bandes
$\omega_{w,k}$	Breite des k -ten Bandes
$\bar{p}(n)$	Mittelwert der geschätzten Leistungen
p_{\min}	Minimalwert für die geschätzte Leistung
$p_k(n)$	Geschätzte Leistung des k -ten Bandes
$\tilde{x}_{j,\text{bp}k}(n)$	Rekonstruiertes überabgetastetes bandpassgefiltertes Subband-Signal
$y(n)$	Eingangssignal der Vorfilterkaskade
$\hat{y}(n)$	Ausgangssignal der Nachfilterkaskade

Formelzeichen der Sphärisch Logarithmischen Quantisierung

D	Zahl der verwendeten Dimensionen
Δ	Kantenlänge der Quantisierungszellen
$e_{\text{ADPCM}}(n)$	Fehlersignal der ADPCM-Codierung
η	Zahl der Durchläufe bei der Radiusschätzung
N_φ	Winkelindex
N_r	Radiusindex
N_{SLQ}	SLQ-Index
R	Resultierende Anzahl an Bits/Sample
r	Radius des Vektors
\hat{r}	Quantisierter Radius
\tilde{r}	Geschätzter Radius
$\hat{\mathbf{x}}$	Vektor mit quantisierten Signalwerten

Kurzfassung

Fehlerrobuste Audiocodierung mit geringer Latenz hat eine Vielzahl von Anwendungsfeldern. So kann diese durch Einsatz in drahtlosen Mikrofon- und Kopfhörersystemen, Audiostreaming-Applikationen und modernen Hörhilfen für effizientere Übertragungsstrecken, erweiterte Funktionsumfänge und kostengünstigere Produkte sorgen. Allerdings eignen sich bestehende standardisierte, quelloffene oder in der Literatur zu findende Verfahren aufgrund zu hoher Latzenzen, zu geringer Audioqualität oder nicht vorliegender Fehlerrobustheit nur eingeschränkt für die Verwendung in solchen Systemen. Aus diesem Grund wird bei vielen praktischen Anwendungen auf proprietäre Lösungen zurückgegriffen.

In der vorliegenden Arbeit wird daher das Ziel der Entwicklung eines Verfahrens zur fehlerrobusten Audiocodierung mit geringer Latenz verfolgt, das diese Nachteile überwindet. Hierzu wird der erfolgversprechende Ansatz einer Subband-ADPCM um Maßnahmen zum Erreichen einer algorithmischen Fehlerrobustheit ergänzt, die Codierstrukturen hinsichtlich einer Anpassung an wechselnde Signaleigenschaften erweitert und diese bezüglich ihrer Parameter global optimiert. Die Ermittlung der Leistungsfähigkeit des entwickelten Codierverfahrens und der Vergleich mit dem Stand der Technik erfolgt auf Basis des PEAQ-Algorithmus. Die Ergebnisse werden zusätzlich durch einen Hörversuch mit im kritischen Hören geübten Probanden validiert.

Gemäß der in dieser Arbeit erfolgten Untersuchungen ist es mit dem entwickelten Audiocodec möglich, bei moderaten Bitraten von bis zu 200 kbit/s und einer algorithmischen Latenz von nur 0,5 ms eine Audioqualität zu erreichen, die selbst erhöhten Ansprüchen gerecht wird, was durch die Ergebnisse des Hörversuches Bestätigung fand. Hierbei wird aufgrund der globalen Parameteroptimierung für das SQAM-Testset eine PEAQ-Bewertung ohne signifikante Ausreißer erreicht. Die Codierrobustheit bei auftretenden Kanalübertragungsfehlern ist dabei algorithmisch sichergestellt, wobei für geringe Fehlerraten eine Abschätzung der Audioqualität zu Ergebnissen führt, die mit denen von aktuellen fehlerrobusten Verfahren aus der Literatur vergleichbar sind.

Stichworte: Audiocodierung mit geringer Latenz, fehlerrobuste Subband-ADPCM, Filterbankdesign, dynamische Bit-Allokation, Audiocodec Optimierung.

Abstract

Error robust low delay audio coding has numerous applications. When utilized in technologies like wireless microphones and headsets, audiostreaming solutions or modern hearing aids it can lead to a more efficient transmission, enhanced functionalities and inexpensive products. Unfortunately most of the well known standardized, open source and published approaches are of limited usability in such scenarios because of too high delays, a not sufficiently high audio quality or a missing robustness against transmission errors. Thus for practical solutions in many cases proprietary codecs are used.

This thesis therefore targets at developing an audio coding scheme which can overcome the drawbacks of existing methods. For this, the promising approach of a subband-ADPCM is extended by methods for achieving a robustness against transmission errors and the codec structures are enhanced by an adaption to changing signal characteristics as well as globally optimized regarding their parameters. The evaluation of the codec performance and comparison to the state of the art is done by means of the PEAQ-algorithm. The tendencies of this evaluation are verified by a listening test with subjects that are trained in critical listening.

According to the results obtained with the research conducted in this work, the developed audio codec, with its algorithmic delay of only 0,5 ms, provides an audio quality that even meets higher demands at a bitrate of up to 200 kbit/s which was confirmed by the results of the listening test. Due to the applied global parameter optimization, a PEAQ-evaluation of the codec leads to results without significant outliers. The robustness against channel transmission errors is algorithmically ensured and for low error rates an audio quality is achieved that is comparable to the results of error robust state of the art approaches.

Keywords: Low delay audio coding, error robust subband-ADPCM, filter bank design, dynamic bit-allocation, audio codec optimization.

Kapitel

1

Einleitung

Digitale Audiosignalverarbeitung und im Speziellen Audiocodierung hat zweifelsohne in der heutigen Kommunikations- und Unterhaltungselektronik einen hohen Stellenwert eingenommen. So verfügt jedes moderne Smartphone neben der Funktionalität zum GSM-basierten Telefonieren über die Möglichkeit der Internet-basierten Sprachkommunikation sowie die Unterstützung zum Abspielen komprimierter Musikstücke in Formaten etablierter Standards wie MPEG-1 Audio Layer III (MP3) [ISO93] oder MPEG-2 (Part 7) Advanced Audio Coding (AAC) [ISO97].

Während die aktuellen Trends der Codecentwicklung hin zu effizienten Algorithmen zur objektorientierten 3D-Audio Codierung, dem so genannten *Spatial Audio Object Coding* [HD07] gehen, findet bei der Archivierung und dem Download von Musikinhalten eher eine Kehrtwende hin zu unkomprimierter Verarbeitung statt [Pau15]. Dennoch zeigen nicht zuletzt die aktuellen Aktivitäten in der Standardisierung [FHM⁺15], dass im Bereich der mobilen Echtzeitkommunikation weiterhin sehr viel aktive Forschungsarbeit betrieben wird.

Ums so überraschender ist es, dass es für Echtzeitanwendungen mit höheren Ansprüchen an die Audioqualität und Latenz keine standardisierten Codierungsverfahren gibt und auch keine quelloffenen Codecs vorliegen, die den Anforderungen entsprechen, so dass daher in kommerziellen Anwendungen meist auf proprietäre Lösungen zurückgegriffen wird.

Dabei sind die Anwendungsfelder von Audiocodierung mit geringer Latenz vielseitig. Abbildung 1.1 zeigt drei typische Szenarien, in denen ein Bedarf an Audiocodierung mit geringer Latenz und hoher Audioqualität besteht.

Teilabbildung (a) beschreibt mit einer Live-Bühnenperformance inklusive Rückübertragung über so genanntes *In-Ear-Monitoring* einen der kritischsten Anwendungsfälle. Das Szenario sieht die codierte Übertragung des Signales eines drahtlosen Mikrofonsystems an ein üblicherweise verwendetes Mischpult sowie die ebenfalls erfolgende Rückübertragung auf die, vom Musiker getragenen, Kopfhörer vor. Untersuchungen wie [LB07] haben gezeigt, dass mit der Überschreitung einer gewissen Verzögerung (hier exemplarisch 8 ms) zwischen Erzeugung und Wahrnehmung eines Lautes durch einen Musiker, ein Verlust der Stimmkontrolle sowie das Wahrnehmen deutlicher Artefakte einhergehen. Wird berücksichtigt, dass zusätzlich eine AD/DA-Wandlung des Signales sowie dessen Übertragung stattfinden muss, verbleiben je nach Auslegung des Systems für die Latenz der Codierung weniger als 2 ms.

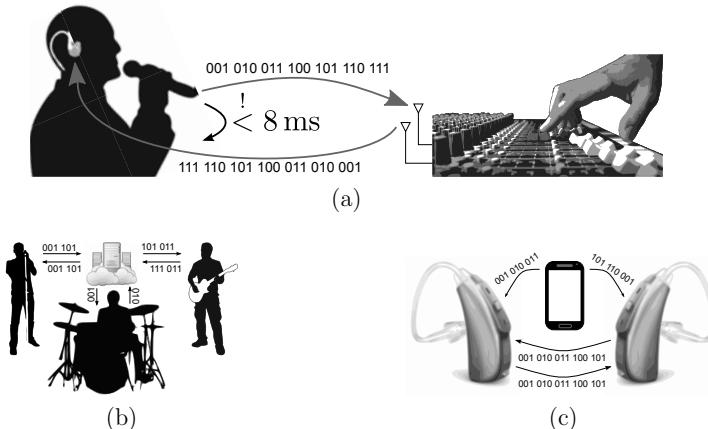


Abbildung 1.1: Anwendungsszenarien für eine Audiocodierung mit geringer Latenz:
 (a) Live-Bühnenperformance mit Rückübertragung über Kopfhörer
 (b) Audiostreaming für *Network Music Performance*-Systeme
 (c) Übertragung von Audiosignalen auf und zwischen Hörhilfen.

Teilabbildung (b) verdeutlicht einen in den letzten Jahren immer mehr in den Fokus der Forschung geratenen Anwendungsfall von Audiocodierung in Systemen zum gemeinsamen Musizieren über das Internet, der so genannten *Network Music Performance* (NMP) [RCR07]. Bedingt durch die Tatsache, dass durch die Übertragung der Audiodaten über ein Netzwerk schon eine verhältnismäßig hohe Verzögerung der Signale vorliegt, werden für die Codierung der Audiodaten Algorithmen mit besonders geringer Latenz benötigt [KW08]. Zwar ist der Anspruch an die Audioqualität unter Umständen etwas geringer als bei dem Fall der Live-Bühnenperformance, aber dennoch liegt er immer noch deutlich über der von heutigen Sprachcodecs bereitgestellten, da das Spielerlebnis nicht beeinflusst werden soll.

Der dritte, in der Teilabbildung (c) illustrierte, Anwendungsfall ergibt sich aus der Tatsache, dass moderne Hörhilfen zunehmend mit Technologien zur Kommunikation untereinander oder mit externen Geräten wie Smartphones ausgestattet werden [Pho12]. Während bei dem Übertragen der Audiosignale wiedergegebener Inhalte von Smartphones oder Fernsehern oft nur die visuelle Synchronität von Bedeutung ist, ergeben sich für die Kommunikation der Geräte untereinander deutlich striktere Latenz- und Qualitätsanforderungen, um fortschrittliche Signalverarbeitung, z.B. zur Erhöhung der Sprachverständlichkeit durch so genanntes *binaurales Beamforming*, zu ermöglichen. Die bei der Übertragung im Kopfumfeld eingeschränkten Datenraten sowie begrenzte Energieressourcen machen hierbei eine Codierung der Audiodaten notwendig. Auch wenn sich bisherige Anwendungen auf die Frequenzbandbreiten menschlicher Sprache konzentrieren, kann es für zukünftige Gerätege-

nerationen von Interesse sein, Vollband-Audio zu übertragen, um das Hörerlebnis und die Sprachverständlichkeit weiter zu verbessern.

Aus den beschriebenen Anwendungsszenarien ergeben sich die Rahmenbedingungen, die dieser Arbeit zugrunde gelegt wurden und die eine genauere Einordnung des Themenfeldes der *fehlerrobusten Audiocodierung mit geringer Latenz* ermöglichen.

Hiernach leitet sich aus den vorgestellten Anwendungsfällen eine obere Grenze für die Latenz¹ von 2 ms ab, wobei in dieser Arbeit ein Verfahren zur Codierung angestrebt wird, das eine Verzögerung von deutlich unter 1 ms erlaubt. Für die Datenrate werden die Aussagen aus [Pho12] sowie Abschätzungen aus [Kei06] zugrunde gelegt. In [Pho12] wird die Codec-Datenrate für die Übertragung zwischen Hörhilfen mit maximal 300 kbit/s angegeben. Gemäß [Kei06] ist für Anwendungen wie in Abbildung 1.1 (a) mit Bandbreiten von 200 kHz und einer Bandbreiteneffizienz von 1 Bit/s/Hz und demnach einer möglichen Bruttodatenrate von 200 kbit/s zu rechnen. Für die vorliegende Arbeit wird daher von einer zulässigen Nettodatenrate² von ungefähr 140-190 kbit/s ausgegangen.

Die Anforderungen an die Audioqualität des, in der vorliegenden Arbeit entwickelten, Codierungsverfahrens in einer Kennzahl festzuhalten, gestaltet sich als vorab schwierig. Zwar wird mit dem in Kapitel 4 vorgestellten PEAQ-Verfahren ein Algorithmus benutzt, der das Abschätzen der Audioqualität eines Audiocodierungssystems ermöglicht, letzten Endes müssen diese Ergebnisse jedoch immer durch Hörversuche bestätigt werden. Als Anforderung an die Audioqualität des zu entwickelnden Systems kann jedoch festgehalten werden, dass diese so hoch sein soll, dass ein codiertes Signal selbst für geübte Probanden kaum vom Original zu unterscheiden ist.

Die letzte Anforderung betrifft die Robustheit des Codieralgorithmus gegenüber Kanalübertragungsfehlern sowie die Anpassung an sich ändernde Kanalzustände. Hierbei gilt es, Auswirkungen von auf dem Kanal auftretenden Einzelbit- oder Bündelfehlern in ihren Auswirkungen auf das decodierte Audiosignal zu begrenzen sowie eine Fortpflanzung eingekoppelter Fehler auszuschließen. Des Weiteren wäre es denkbar, eine Möglichkeit zur Skalierbarkeit des Codecs zur Laufzeit vorzusehen, um beim Vorhandensein eines entsprechenden Steuersignales eine Anpassung des Codecs an den jeweiligen Kanalzustand vornehmen zu können.

Die oben definierten Stellgrößen und Anforderungen stehen in einer engen Wechselwirkung zueinander (vgl. Abb. 1.2), die sich mit den gesetzten Rahmenbedingungen deutlich verschärft. Dies liegt unter anderem darin begründet, dass viele der bekannten Methoden zur verlustfreien und verlustbehafteten Audiocodierung aufgrund der strikten Latenzanforderungen nicht anwendbar sind. Damit besteht eine besondere Herausforderung in der vorliegenden Problemstellung, wie sich im Zuge der folgenden Kapitel zeigen wird.

¹In dieser Arbeit werden die Begriffe Latenz und Verzögerung synonym verwendet und bezeichnen die zeitliche Differenz zwischen dem Vorliegen einer Wellenform am Eingang des Encoders und die Ausgabe der entsprechend rekonstruierten Wellenform am Ausgang des Decoders.

²Alle im Folgenden für das entwickelte Verfahren angegebenen Bitraten sind damit, soweit nicht anders erwähnt, Nettodatenraten und beziehen sich auf eine Abtastrate von 48 kHz.

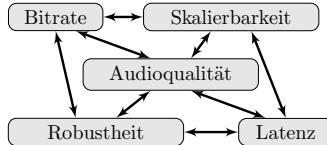


Abbildung 1.2: Wechselwirkungen im Zielkonflikt der Audiocodierung.

1.1 Stand der Technik: Audiocodierung mit geringer Latenz

Grundsätzlich lassen sich die Verfahren für die Codierung von, zur Wahrnehmung durch den Menschen verarbeiteten, Daten in die zwei Gruppen verlustlos und verlustbehaftet einteilen. Verlustlose Codierung bezeichnet eine Form der Datenreduktion, die das bitexakte Wiederherstellen des ursprünglichen digitalen Datenstroms erlaubt. Bei einer verlustbehafteten Codierung werden, unter Ausnutzung menschlicher kognitiver Eigenschaften, nicht wahrgenommene Anteile eines Signales bei der Codierung weggelassen. Beim Einsatz verlustbehafteter Codierung wird in vielen Fällen zwar auch noch eine hinterher geschaltete verlustlose Codierung verwandt, das Verfahren wird damit aber immer noch als verlustbehaftet klassifiziert.

Für die vorliegende Problemstellung scheidet eine reine verlustlose Codierung wie z.B. eine Huffman Codierung oder arithmetische Codierung aus, da die Kompressionsrate meist zu gering ist und in vielen Fällen das gesamte Signal oder große Teilstücke zum Erstellen von Signalstatistiken bekannt sein müssen. Zusätzlich entsteht durch die dabei auftretenden variablen Codewortlängen bei Übertragung über einen Kanal konstanter Bitrate die Notwendigkeit ein Bitreservoir einzufügen und damit zusätzliche Verzögerung in Kauf zu nehmen. Bekannte Open Source Verfahren für verlustlose Codierung von Audiosignalen sind der *Free Lossless Audio Codec* (FLAC) [FLA15] und der *WavPack*-Codec [Wav98]. Letzterer erlaubt auch eine Konfiguration zur nicht vollständig verlustfreien Codierung, die zumindest für Musiksignale eine ausreichende Audioqualität bei den hier anvisierten Bitraten ermöglicht, jedoch in der Latenz mit mindestens 2,6 ms noch über der in dieser Arbeit angestrebten Grenze liegt [KW08].

Gegen übliche Codierungsverfahren zur verlustbehafteten Audiodatenkompression wie MP3 sprechen in dem hier skizzierten Anwendungsfall rein algorithmische Gründe. Hierbei verursachen die, zur Analyse gemäß eines psychoakustischen Modells und für die Codierung verwendeten, Filter und Blockbildungen für Transformationen eine unausweichliche Verzögerung, die mit mindestens 20 ms in ihrer Größenordnung deutlich über der für Echtzeitanwendungen nötigen liegt.

Daher gibt es neben dem, in der mobilen Kommunikation verwendeten und im Standard G.728 aufgenommenen, Verfahren zur Sprachcodierung mit geringer Latenz *Low Delay Code Excited Linear Prediction* (LD-CELP) [ITU92] auch diverse Verfahren zur Audiocodierung mit gegenüber MP3 deutlich verringelter Verzögerung.

Im Folgenden wird ein kleiner Überblick über diese Verfahren gegeben und die jeweilige Eignung für den vorliegenden Anwendungsfall diskutiert. Hierbei kann und soll kein Anspruch auf Vollständigkeit erhoben werden. In [Kei06] (S. 7-12) lassen sich zum Beispiel noch weitere zum Teil deutlich ältere Ansätze finden, die an dieser Stelle daher nicht weiter aufgegriffen werden, zumal es sich zum Teil nur um erweiterte Sprachcodecs handelt.

Ein prominenter Vertreter zur Codierung von Vollband-Audiosignalen³ ist der im Bluetooth Standard [Blu12] aufgenommene, ursprünglich von Philips entwickelte, *Subband-Codec* [dBG095]. Dieser erlaubt eine nahezu freie Skalierung der Bitrate und Latenz mit, für eine hohe Audioqualität im Standard empfohlenen, moderaten Bitraten von ca. 190 kbit/s bei einer Verzögerung im Bereich von 5 ms. Da für die Aufnahme in den Bluetooth Standard die geringe Komplexität und freie Verfügbarkeit im Vordergrund standen, ist bei einer Verringerung der Latenz auf Werte unter 2 ms mit einer deutlich verminderten Audioqualität zu rechnen (vgl. [HH09]).

Ein weiteres bekanntes Verfahren zur Audiodcodierung für Echtzeitanwendungen ist aus Arbeiten wie [ES00] und [SYHE02] hervorgegangen und wird von der Fraunhofer Gesellschaft unter dem Namen *Ultra Low Delay Audio Codec* (ULD) [KSWH04] vermarktet. Mit der Standardkonfiguration, die eine Latenz von 5,3 ms und Datenrate von 96 kbit/s vorsieht, erzielt der ULD-Codec für viele Signalklassen eine sehr gute Audioqualität (vgl. [ITU09]). Bei Benutzung des optionalen *Low Delay Psychoacoustic Module*, das eine Verringerung der Latenz auf 1,3 ms ermöglicht, fällt die Audioqualität, gemäß informeller Tests, trotz einer dann auf 192 kbit/s zu erhögenden Bitrate jedoch deutlich ab.

Ein lizenzzfreier und quelloffener Audicodec mit geringer Latenz ist der mittlerweile im IETF Standard des so genannten *Opus-Codecs* [IET12] aufgegangene CELT-Codec [VTMM10]. Dieser basiert auf einer Verarbeitung des Eingangssignales in kurzen Blöcken über eine initiale Transformation mittels der *Modifizierten Diskreten Cosinus Transformation* (MDCT) und ermöglicht eine minimale Latenz von 2,5 ms bei einer im Prinzip frei einstellbaren Bitrate [VMTV13]. Auch der Opus-Codec erreicht für Latenzen im Bereich von 5 ms und eine Bitrate von ca. 100 kbit/s eine für die meisten Anwendungen ausreichende Audioqualität. Eine Reduktion der Latenz führt jedoch auch für dieses Verfahren zu einer deutlichen Verschlechterung in der Audioqualität, die sich allerdings durch eine Anhebung der Bitrate in den Bereich von 200 kbit/s kompensieren lässt.

Als weitere aus der Standardisierung hervorgegangene Codecs sind der mit der *Enhanced Low Delay*-Konfiguration (AAC-ELD) erweiterte, bereits erwähnte, AAC-LD Codec [SSJ⁺08] sowie der von der ITU als G.722.2 standardisierte *Enhanced Voice Service* (EVS) [FHM⁺15] zu nennen. Beide dieser Codecs zielen darauf ab, Echtzeitkommunikation mit erweiterter Audiobandbreite zu ermöglichen. Die erreichte Verzögerung ist jedoch in beiden Fällen mit über 7,5 ms deutlich zu hoch und die Audioqualität liegt bei den üblicherweise verwendeten Datenraten deutlich unter-

³Als Vollband-Audiosignal werden in dieser Arbeit digitale Tonsignale bezeichnet, deren Abtastrate bei mindestens 44,1 kHz liegt, so dass der abgebildete Frequenzbereich die theoretisch mögliche Bandbreite des menschlichen Hörens vollständig abdeckt.

halb der hier anvisierten.

Als rein kommerzielles Verfahren zur Audiocodierung in Echtzeitanwendungen, u.a. bei Bluetooth Geräten, hat sich in den letzten Jahren der so genannte aptX® Live Codec etabliert. Dieser ging aus Arbeiten wie [SS91] hervor, ist aber seitdem in Bezug auf die verbesserte Technologie und einen Vergleich zu anderen Verfahren wenig dokumentiert. Er ermöglicht Latenzen im Bereich von 1,5 ms bei einer Bitrate von ca. 150 kbit/s und erreicht dabei, gemäß informeller Tests, eine Audioqualität, die die in dieser Arbeit zugrunde gelegten Qualitätsanforderungen in den meisten Fällen entsprechen sollte.

Weitere in der Literatur zu findende Ansätze sollen im Folgenden kurz Erwähnung finden, ohne aber im Detail diskutiert zu werden. Darunter sind zum einen Verfahren wie [MLRR99], [MRLR00] und [THS13] zu nennen, die viel versprechende Ansätze beinhalten, im Rahmen der Veröffentlichung jedoch unzureichend dokumentiert und evaluiert sind, so dass sich der Vergleich zu anderen Verfahren schwierig gestaltet. Zum anderen gibt es in der Literatur weitere Ansätze wie [Mat08] oder [Krü10], die mehr von rein wissenschaftlichem Interesse sind, da sie sehr hohe Anforderungen an die Rechenkapazität haben bzw. vermutlich keine ausreichende Audioqualität für den vorliegenden Anwendungsfall bereitstellen. Des Weiteren lassen sich Verfahren identifizieren, die neuartige Ansätze z.B. auf Basis von *2D Lifting Wavelet Transformation* verfolgen [Luo05], aber zu hohe Latenzanforderungen haben oder zur Zeit ihrer Entwicklung nicht für Vollband-Audio vorgesehen bzw. evaluiert wurden [CP88].

Zu guter Letzt finden sich in der Literatur wissenschaftliche Arbeiten, die auf einer Zeitbereichsverarbeitung über eine *Adaptive Differential Pulse Code Modulation* (ADPCM) basieren. Die Ansätze unterscheiden sich zwar in den initial getroffenen Annahmen zur Robustheit gegenüber Kanalübertragungsfehlern und mitunter des zulässigen Rechenaufwandes, bilden aber dennoch zu großen Teilen die Grundlage für die vorliegende Arbeit. Im Detail ist es zum einen der in [Kei06] beschriebene Codec, der bei einer Abtastfrequenz von 44,1 kHz eine minimale Latenz von 0,9 ms bei einer Datenrate von 176 kbit/s erlaubt, aber für viele der kritischen Signale keine ausreichende Audioqualität erreicht. Zum anderen stellt die Veröffentlichung [Hol09] wichtige Erweiterungen und Methoden für einen nicht fehlerrobusten Codec, der vollständig latenzfrei arbeitet, zur Verfügung, die es ermöglichen, selbst für kritische Signale in den meisten Fällen eine ausreichende Audioqualität bereitzustellen.

Parallel zu den in der vorliegenden Arbeit durchgeführten Untersuchungen wurden in [SHZ13a, SHZ14] die Ansätze aus [Hol09] um Maßnahmen erweitert, die zum Erreichen einer gewissen Fehlerrobustheit führen. Diese haben prinzipbedingt aber einen negativen Einfluss auf die Codierqualität, so dass nur bei höheren Bitraten für kritische Signale eine ausreichende Audioqualität erreicht wird und ansonsten Einschnitte bei dieser zu erwarten sind.

Zusammenfassend lässt sich festhalten, dass sich nach Kenntnis des Autors in der Literatur kein Verfahren finden lässt, das die Eigenschaften einer

- Vollband-Audiocodierung bei einer Abtastrate von mindestens 44,1 kHz
- mit einer sehr hohen Audioqualität, auch für kritische Signale,
- bei einer algorithmisch sichergestellten Fehlerrobustheit
- und, trotz verwendeter Filterbank, sehr geringer Latenz
- sowie einer Rechenbarkeit auf einem standard-DSP

auf sich vereint und somit die hier zu Grunde gelegten Anforderungen erfüllt.

1.2 Ansatz ADPCM-basierter Subband-Codierung

Gemäß den Erkenntnissen aus [Kei06] und [Hol09] sowie dem Erfolg der Verfahren aus [dBG095] und [SS91] wurde für die vorliegende Arbeit daher der Ansatz einer Subband-ADPCM-Codierung gewählt. Dies erfolgt mit dem Ziel, die Defizite dieser Verfahren bezüglich der Robustheit gegenüber Übertragungsfehlern bzw. der Latenz oder Audioqualität zu kompensieren, um die in dieser Arbeit zugrunde gelegten Anforderungen zu erfüllen.

Abbildung 1.3 zeigt den Flussgraphen des Basisalgorithmus. Das Eingangssignal $x(n)$ wird im Encoder durch eine Analysefilterbank (AFB) in M kritisch unterabgetastete Subband-Signale $x_i(m)$ zerlegt, die mittels einer adaptiven Prädiktion sowie Quantisierung des daraus ermittelten Prädiktionsfehlers (Blöcke $AdPQ_i$) in Subband-Quantisierungsindizes $\hat{q}_i(m)$ überführt werden. Im Decoder werden die mittels der Blöcke $AdPQ_i^{-1}$ rekonstruierten Subband-Signale $\tilde{x}_i(m)$ über eine Synthesefilterbank (SFB) zum Ausgangssignal des Codecs $\tilde{x}(n)$ rekonstruiert.

Da das Verarbeiten der Subband-Signale zu Quantisierungsindizes quasi ohne Verzögerung erfolgen kann, liegt die Latenz dieses Ansatzes maßgeblich in der Gruppenlaufzeit der Filterbank begründet.

Der Vorteil dieses Verfahrens ist, dass die Verarbeitung des Eingangssignales direkt im Zeitbereich erfolgt, so dass keine Verzögerung durch Blockbildung für etwaige Transformationen in Kauf genommen werden muss. Zusätzlich entfallen die, in der Transformationscodierung bei Bildung besonders kurzer Blöcke, nötigen Maßnahmen zur Kompensation der verminderten Frequenzauflösung (vgl. [VTMM10]). Des Weiteren ermöglicht das Verfahren, im Gegensatz zu einer breitbandigen ADPCM, durch die gewonnenen Freiheitsgrade eine freiere Skalierung von Größen wie der Bitrate und des Rechenaufwandes. Ein weiterer Aspekt, der für die Verwendung einer Subband-ADPCM spricht, ist der bereits in [SS91] diskutierte so genannte Subbandgewinn, der aus einer besser an die Eigenschaften der Subband-Signale angepassten Verarbeitung und einer indirekten psychoakustischen Rauschformung resultiert.

Trotz der genannten Vorteile einer Subband-Struktur ergeben sich bei der Entwicklung eines Codecs nach dem gewählten Ansatz auch vielseitige Herausforderungen. So findet der Entwurf von Filterbänken mit Gruppenlaufzeiten unter 1 ms und hinreichenden Frequenzgangs- und Rekonstruktionseigenschaften an den Grenzen der Leistungsfähigkeit gängiger Entwurfsverfahren statt. Zusätzlich ist das Erreichen einer ausreichenden Audioqualität bei gleichzeitiger geringer Bitrate und Latenz eine bislang nicht vollständig gelöste Problemstellung.

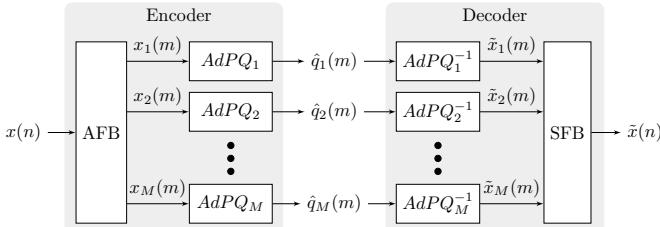


Abbildung 1.3: Flussgraph der ADPCM-basierten Subband-Codierung.

1.3 Ziele, Beiträge und Aufbau der Arbeit

Aus den in den vorherigen Abschnitten genannten Aspekten ergibt sich für diese Arbeit das Ziel der Entwicklung eines Audiocodecs

- mit einer sehr hohen Rekonstruktionsqualität, trotz möglichst geringer Latenz,
- der aufgrund seiner Fehlerrobustheit in einem realen System operieren könnte
- und in einem sinnvollen Arbeitspunkt durch globale Optimierung seiner Parameter arbeitet
- sowie durch optionale Vor- und Nachverarbeitung zur Verbesserung der Audioqualität erweitert werden kann.

Zum Erreichen dieser Ziele wurden die einzelnen Bausteine des Subband-ADPCM-Codecs durch Maßnahmen zum Erreichen einer Fehlerrobustheit ergänzt, auf potentielle Verbesserungsmöglichkeiten untersucht sowie das Gesamtsystem durch weitere Funktionalitäten ergänzt. Des Weiteren wurde eine globale Optimierung der Parameter der Subband-Codierung nach psychoakustischen Gesichtspunkten vorgenommen. Die dabei geleisteten und veröffentlichten Beiträge werden im Folgenden im Rahmen einer Beschreibung der übergeordneten Gliederung dieser Arbeit unter Angabe eines Verweises auf den jeweiligen Abschnitt explizit aufgelistet.

Im folgenden Kapitel 2 werden die untersuchten Verfahren zum Entwurf von Filterbänken mit geringer Latenz vorgestellt. Abschnitt 2.3 beinhaltet eine, auch in [PKO12] veröffentlichte, Übersicht über Verfahren für das Design von Filterbän-

ken mit geringer Latenz sowie den Vergleich von mit ihnen erzielten Filterbankentwürfen über geeignete Bewertungsgrößen. Zusätzlich wird in Abschnitt 2.4 eine, auch in [PO15b] präsentierte, Erweiterung eines Entwurfsverfahrens für Filterbänke mit geringerer Latenz vorgestellt, mit der es möglich ist, die Sperrdämpfungen der Subband-Filter deutlich zu erhöhen.

Kapitel 3 beinhaltet in den Abschnitten 3.2 und 3.3 eine Beschreibung der zum Teil auch in [PO11] präsentierten Methoden zur fehlerrobusten adaptiven Prädiktion und Quantisierung. In Abschnitt 3.2.3 wird zudem eine Modifikation zur Prädiktoradaption auf das Prädiktionsfehlersignal präsentiert. Es werden außerdem weitere Maßnahmen zur Verbesserung der ADPCM-Codierung, zum Beispiel durch eine Erweiterung der Prädiktorstruktur um einen statischen Anteil, eingeführt.

In einem weiteren Abschnitt 3.3.3 des dritten Kapitels werden die, auch im Zuge von [PO15a] vorgestellten, Modifikationen und Erweiterungen einer Methode zur dynamischen Bit-Allokation beschrieben. Untersuchungen, die die Basis für mögliche weiterführende Untersuchungen zu alternativen Konzepten der adaptiven Quantisierung bilden, die auch unter [PLO16] einzusehen sind, lassen sich in Abschnitt 3.3.4 finden. Des Weiteren wurden in Abschnitt 3.4.1 Ergebnisse aufgenommen, die die Grundlage für die Veröffentlichung [PWO15] bildeten, wobei mittels einer Erweiterung des Subband-Codecs durch eine adaptive Vor- und Nachfilterung eine Verbesserung der Audioqualität ohne zusätzlichen Signalisierungsaufwand erreicht werden konnte.

Das Kapitel 4 dient der Präsentation der auch in den Veröffentlichungen [PO15a] und [PWO15] genutzten Methoden und Algorithmen zur globalen Optimierung der Audiocodec Parameter. Es werden die angewendeten Bewertungsgrößen eingeführt, das gewählte Vorgehen zur Optimierung bei dynamischer Bit-Allokation beschrieben sowie Resultate dieser vorgestellt.

In Kapitel 5 werden die Ergebnisse einer Evaluierung der Leistungsfähigkeit des entwickelten Verfahrens angegeben. Es werden Resultate von Messungen mit objektiven Messgrößen dargestellt sowie Ergebnisse subjektiver Tests analysiert und diskutiert. Zusätzlich erfolgt die Evaluation des Einflusses von Kanalübertragungsfehlern auf die Audioqualität des Codierungsverfahrens.

Kapitel 6 liefert eine Zusammenfassung der wichtigsten Ergebnisse und Erkenntnisse der vorliegenden Arbeit und gibt einen Ausblick auf mögliche weiterführende Untersuchungen und Anknüpfungspunkte.

Kapitel

2

Filterbänke mit geringer Latenz

In diesem Kapitel werden die im Rahmen der vorliegenden Arbeit untersuchten Verfahren zum Entwurf von Filterbänken mit geringer Latenz sowie die mit ihnen erzielten Filterbankentwürfe vorgestellt. Abschnitt 2.1 gibt eine allgemeine Einleitung in die Thematik der Filterbänke in der Audiocodierung und beinhaltet die nötigen Grundlagen und Begriffsdefinitionen für das Verständnis der folgenden Abschnitte. In Abschnitt 2.2 werden die betrachteten Entwurfsverfahren kurz vorgestellt und zusammengefasst sowie in Abschnitt 2.3 ein Vergleich verschiedener Filterbankentwürfe über geeignete Bewertungsgrößen durchgeführt.

Ziel ist es dabei, die Eignung für den vorliegenden Anwendungsfall zu ermitteln und die Grenzen des jeweiligen Entwurfsverfahrens unter den in dieser Arbeit zugrunde gelegten Rahmenbedingungen aufzuzeigen.

Des Weiteren wird mit der in 2.4 vorgestellten neuartigen globalen Optimierung der Parameter eines der Entwurfsverfahren eine Methodik präsentiert, die eine weitere Verbesserung der Entwurfsergebnisse eines der Verfahren ermöglicht.

2.1 Filterbänke in der Audiocodierung

Filterbänke spielen in der verlustbehafteten Audiocodierung seit jeher eine wichtige Rolle. So basieren etablierte Verfahren wie MP3 auf der Zerlegung des Eingangssignales durch Filterbankstrukturen für die weitere signalangepasste Verarbeitung der Subband-Signale mittels einer MDCT.

Eine Filterbank ist die Zusammenschaltung mehrerer Einzelfilter definierter Frequenzcharakteristik und dient der Zerlegung des Eingangssignales in Teilsignale, die jeweils nur die Energie bzw. Information eines bestimmten und sich zumeist nicht überlappenden Frequenzbereiches beinhalten. Die Zerlegung in dem so genannten Analyseteil erfolgt dabei mit dem Ziel, eine weitere Verarbeitung oder Analyse der erzeugten Signale durchzuführen. Soll nach der Verarbeitung der Teilsignale wiederum ein Gesamtsignal erzeugt werden, so kommt eine zur ersten Anordnung komplementäre Struktur zum Einsatz, die als Syntheseteil bezeichnet wird.

Abbildung 2.1 zeigt den Flussgraphen einer gleichförmigen kritisch unterabgetas-

teten M-Kanal Filterbank in Analyse/Synthese-Struktur. Das Eingangssignal $x(n)$ wird im Analyseteil durch die Analysefilter $H_i(z)$ in die Subband-Signale $x_i(n)$ zerlegt. Diese werden mit dem Faktor M kritisch unterabgetastet, woraus die Subband-Signale $x_i(m)$ resultieren. Im Syntheseteil der Filterbank werden die Subband-Signale zunächst um den Faktor M überabgetastet und danach mit den Synthesefiltern $G_i(z)$ gefiltert. Die Ausgangssignale der Synthesefilter werden zur Kompensation der Unter- und Überabtastung sowie Synthesefilterung mit dem Faktor M multipliziert¹ und zum Rekonstruktionssignal $\tilde{x}(n)$ aufaddiert.

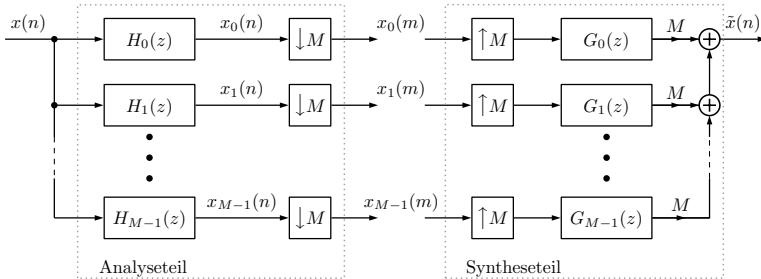


Abbildung 2.1: Flussgraph einer kritisch unterabgetasteten M-Kanal Filterbank.

Die Frequenzgänge der Analyse- und sich daraus ergebenden Synthesefilter unterteilen das Spektrum des Eingangssignales im Idealfall, durch Gewichtung mit einem konstanten Betrag der Übertragungsfunktion von 1 im Durchlassbereich und 0 im Sperrbereich, in M sich nicht überlappende Bänder. Sie entsprechen demnach denen von idealen Tief-, Band- und Hochpässen. Abbildung 2.2 beinhaltet Beispielbetragsfrequenzgänge der Subband-Filter einer 4-Kanal ($M = 4$) Filterbank, die dieser Vorgabe näherungsweise entsprechen, jedoch eine endliche Steilheit beim Übergang zwischen Durchlass- und Sperrbereich aufweisen.

Die Frequenzgänge realer Subband-Filter lassen sich grundsätzlich in drei Bereiche unterteilen, die in der Reihenfolge des Tiefpasses in Abbildung 2.2 kurz eingeführt werden sollen. Der bereits erwähnte Durchlassbereich ist der Bereich mit einem im Idealfall konstanten Betragsfrequenzgang von 1 bzw. 0 dB. Der Beginn des Übergangsbereiches wird in der Regel durch den Punkt definiert, an dem der Betragsfrequenzgang um einen definierten Wert (hier z.B. -6 dB) abgefallen ist und ab dem er weiter kontinuierlich fällt. Am Beginn des Sperrbereiches beträgt der Betragsfrequenzgang im Idealfall 0 bzw. $-\infty$ dB. Dieser beginnt im Falle einer Filterbank gemäß Definition bei der Mittenfrequenz des benachbarten Filters [Kei06]. Die Sperrdämpfung beschreibt, wie stark die Amplitude in diesem Bereich abgesenkt wird.

¹ Die Multiplikation mit M wird in manchen Darstellungen auch weggelassen, da sie in der Synthesefilterung mit den Filtern $G_i(z)$ berücksichtigt werden kann.

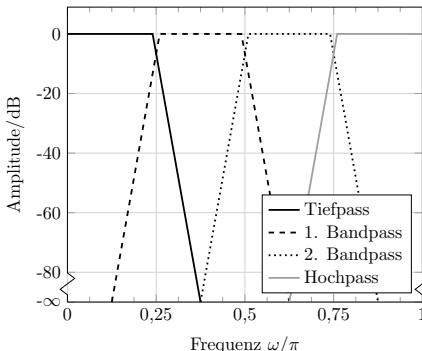


Abbildung 2.2: Beispelfrequenzgänge der Subband-Filter einer 4-Kanal Filterbank.

Durch die Abweichung der Subband-Frequenzgänge von den Idealverläufen entsteht bei der kritischen Unterabtastung Aliasing. Die durch den Sperrbereich entstehenden Anteile werden dabei in der Regel relativ stark gedämpft. Für eine perfekte Rekonstruktion durch die gesamte Filterbank muss aber, besonders auch für den Übergangsbereich, durch die so genannte *Aliasingkompensation* ein Auslöschen der Aliasingkomponenten bei der Synthesefilterung der Subbänder stattfinden.

Eine wichtige geforderte Eigenschaft einer Analyse/Synthesefilterbank ist damit die perfekte² bzw. nahezu perfekte Rekonstruktion. Hierbei wird bewertet, wie sehr das Rekonstruktionssignal $\tilde{x}(n)$ und das Eingangssignal der Filterbank $x(n)$ übereinstimmen. In [NBS92] wurde dafür die Größe des SNR_r über

$$\text{SNR}_r = 10 \log_{10} \left(\frac{\text{Signalenergie}}{\text{Rekonstruktionsfehlerenergie}} \right) = 10 \log_{10} \left(\frac{\sum_n x^2(n)}{\sum_n (x(n) - \tilde{x}(n + \tau))^2} \right) \quad (2.1)$$

eingeführt, die für definierte Signale wie *Einheitsrampensignale* und gleichverteilte Zufallssignale definierter Amplitude ausgewertet wird. Die zeitliche Verschiebung τ des Rekonstruktionssignales in der Berechnung der Rekonstruktionsfehlerenergie dient dabei der Kompensation der Gruppenlaufzeit der Filterbank.

Für den hier angestrebten Entwurf einer Filterbank für die Verwendung in einer Audiocodierung mit geringer Latenz hat diese Gruppenlaufzeit der Filterbank eine entscheidende Bedeutung, da sie die Gesamtverzögerung des Codierverfahrens maßgeblich bestimmt. Grundsätzlich wäre ein linearer Phasengang der Subband-Filter wünschenswert, um eventuelle Einflüsse auf die Subband-Signalverarbeitung auszuschließen. Für eine extrem geringe Gruppenlaufzeit sind jedoch Subband-Filter mit einem nichtlinearen Phasenverlauf unausweichlich.

²Die jedoch nur in der mathematischen Formulierung der Filterbank vorliegt und mit der, in realen Systemen bestehenden, begrenzten numerischen Genauigkeit nicht erreicht werden kann.

In der Literatur finden sich vielzählige Ansätze zur mathematischen Formulierung von Filterbankstrukturen sowie der Optimierung ihrer Subband- bzw. Prototyp-Filter. Grundsätzlich lassen sich die Verfahren durch mehrere strukturelle Eigenschaften voneinander unterscheiden. Zum einen sind sie dadurch voneinander abzugrenzen, ob die Formulierung eine Ableitung der Subband-Filter von einem gemeinsamen Prototypfilter durch Modulation vorsieht oder, zumindest in der Grundformulierung, jedes Filter für sich in seinen Koeffizienten einzeln optimiert wird. Zum anderen ist ein entscheidendes Merkmal, ob die Struktur aus Filtern endlich langer Impulsantworten (so genannten FIR-Filtern) oder unendlich langer Impulsantworten (so genannten IIR-Filtern) zusammengesetzt ist. Des Weiteren gibt es einerseits Verfahren, die in ihrer mathematischen Formulierung eine perfekte Rekonstruktion direkt vorsehen sowie diese für mit ihnen erstellte Entwürfe im Rahmen der vorliegenden Rechengenauigkeit erfüllen und andererseits Ansätze, die diese nur näherungsweise erreichen können.

Die Mehrheit der Verfahren sieht dabei eine gleichförmige Aufteilung des Eingangsspektrums und eine kritische Unterabtastung vor. Für die Codierung ist die kritische Unterabtastung obligatorisch, um die Gesamtzahl an zu verarbeitenden Signalwerten im Vergleich zum Eingangssignal nicht zu erhöhen. Zusätzlich sind viele Verfahren als reine Zweibandstrukturen formuliert, so dass höhere Bandanzahlen nur durch Kaskadierung mehrerer Filterstrukturen erreicht werden können. Selbiges gilt häufig auch für die generell aus psychoakustischen Gesichtspunkten wünschenswerte ungleichförmige spektrale Aufteilung des Eingangssignales zur angepassten Verarbeitung in den Bandbreiten des menschlichen Hörens. In beiden Fällen ergibt sich der Nachteil, dass die Gesamtverzögerung bei einer Kaskadierung und vergleichbaren Frequenz- und Rekonstruktionseigenschaften zumeist höher ausfällt als bei Verfahren, die einen direkten Entwurf mit mehr als zwei Bändern vorsehen. Daher werden diese Ansätze im weiteren Verlauf nicht betrachtet.

2.2 M-Kanal Filterbankentwurfsverfahren

Im Folgenden werden die im Zuge dieser Arbeit untersuchten Filterbankentwurfsverfahren vorgestellt. Der Inhalt dieses Abschnittes ist dabei vergleichbar zu den in [PKO12] veröffentlichten Ergebnissen, wobei dort zudem auch Untersuchungen zum Einfluss verschiedener Filterbankentwürfe auf eine frühe Version des Subband-Codecs zu finden sind.

Voraussetzung für die Auswahl an Verfahren war die in der Literatur berichtete Eignung für den Entwurf von Filterbankstrukturen mit geringer Latenz und die potentielle Möglichkeit, auch mehr als zwei Kanäle ohne eine Kaskadierung zu erhalten sowie eine erreichbare nahezu perfekte Rekonstruktion. Ziel ist es dabei, gemäß der Erfahrungen mit den Verfahren [dBGO95], [SS91] und [Kei06], Filterbänke mit Bandanzahlen zwischen vier und acht Subbändern zu erstellen.

Da die im Folgenden vorgestellten Entwurfsverfahren zum Teil recht umfangreiche mathematische Herleitungen haben und es im Rahmen dieser Arbeit somit nicht möglich ist, diese vollständig wiederzugeben sowie um unnötige Wiederholungen zu vermeiden, erfolgt die Beschreibung der Ansätze jeweils nur so weit, dass ein Verständnis der Grundkonzepte, Eigenschaften und Entwurfsparameter möglich ist. Um das Nachvollziehen der in Abschnitt 2.4 beschriebenen Methode zur globalen Optimierung der Entwurfsparameter eines der Entwurfsalgorithmen etwas zu erleichtern, wird das im nächsten Abschnitt 2.2.1 vorgestellte Verfahren dabei jedoch ein wenig ausführlicher dargestellt. Für die genauen mathematischen Herleitungen und weiterführende Literatur wird an dieser Stelle in jedem der Fälle auf die angegebenen Referenzen verwiesen.

2.2.1 Entwurfsverfahren nach Nayebi

Nayebi et al. stellen in diversen Veröffentlichungen [NBS92, NBS94] einen nach Aussage der Autoren zur damaligen Zeit neuartigen Ansatz für die Analyse und das direkte Design einer großen Klasse von allgemeinen Analyse-/Synthese-Systemen, die auf M -Kanal Filterbänken basieren, vor.

Das Verfahren beruht auf einer Zeitbereichsbeschreibung der gesamten Filterbank in einem System aus M Matrixgleichungen³

$$\mathbf{A}\mathbf{s}_i = \mathbf{b}_i \quad i = 1, 2, \dots, M. \quad (2.2)$$

In diesem System enthalten die Matrizen \mathbf{A} und \mathbf{S} (aus der \mathbf{s}_i entstammt) die Koeffizienten der Analyse- und Synthesefilter in einer Anordnung, die bei ihrer Multiplikation einer verallgemeinerten Faltung für ein Multiband Analyse/Synthese-System entspricht. Demzufolge enthält die Matrix \mathbf{B} (aus der \mathbf{b}_i entstammt) bei perfekter Rekonstruktion der Filterbank zum größten Teil Nullen bis auf eine, entsprechend der gewünschten Gruppenlaufzeit der Filterbank, verschobene und mit $1/M$ besetzte Antidiagonalmatrix. Dies wird dadurch begründet, dass diese Art einer Faltung der Subband Analyse- und Synthesefilter in M verschobenen Dirac-Impulsen resultiert. Der eigentliche Algorithmus zur Optimierung eines Filterbankentwurfes gemäß vorgegebener Parameter wie Bandanzahl M , Anzahl der Koeffizienten der Subband-Filter N und Gruppenlaufzeit der Filterbank τ ist in Abbildung 2.3 dargestellt. Eingang der Optimierung sind die in der Matrix \mathbf{A}_0 zusammengefassten Koeffizienten von Prototypentwürfen für die Analysefilter, die für Entwürfe mit geringer Latenz nach Möglichkeit nahezu minimalphasig sind. Mit der sich aus der gewünschten Gruppenlaufzeit ergebenden Matrix \mathbf{B} lässt sich das Gleichungssystem in 2.2 z.B. im Sinne eines *Least Squares-Problems* hinsichtlich der Matrix \mathbf{S} bzw. \mathbf{S}_{LS} lösen.

³In [NBS92] wird in einer allgemeinen Herleitung ein System aus R Matrixgleichungen beschrieben, wobei R den Unter- und Überabtastungsfaktoren der Filterbank entspricht. Da in dieser Arbeit nur kritisch unterabgetastete Filterbänke betrachtet werden gilt $R = M$.

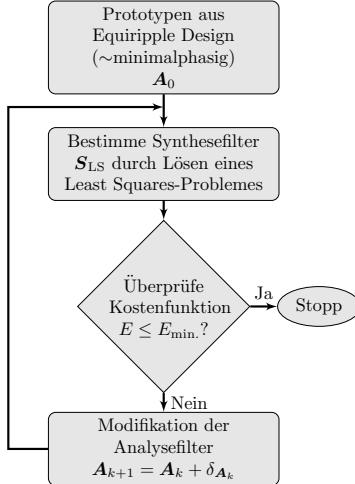


Abbildung 2.3: Flussgraph der Filterbank-Optimierung nach Nayebi.

Die Kostenfunktion für die Optimierung in der original Notation von Nayebi et al. lautet

$$E = w_1 \epsilon_r + \sum_{j=2}^J w_j \epsilon_j \quad (2.3)$$

und ist die Summe gewichteter Einzelerme, die die Rekonstruktionsqualität der gesamten Filterbank (vorderer Teil) und die Frequenzcharakteristik der Subband-Filter (hinterer Teil) erfassen.

Der Rekonstruktionsfehler ϵ_r kann über

$$\epsilon_r = \sum_{i=1}^M \|\mathbf{e}_i\|_2 \quad (2.4)$$

berechnet werden, wobei

$$\mathbf{e}_i = \mathbf{A} \mathbf{s}_{LS_i} - \mathbf{b}_i \quad i = 1, 2, \dots, M \quad (2.5)$$

der Fehler der Least Squares-Lösung ist.

Die Summanden der Kostenfunktion zur Erfassung der Frequenzcharakteristiken der Subband-Filter besteht aus verschiedenen Termen ϵ_j , wie z.B. für das Stoppband

$$\epsilon_{sb} = \int_{\omega \in \text{Stoppband}} |H_m(\omega)|^2 d\omega \quad (2.6)$$

oder die maximale Welligkeit im Durchlassbereich (Passband)

$$\epsilon_{pb} = \max\{|1 - |H_m(\omega)||\} \quad \omega \in \text{Passband.} \quad (2.7)$$

Hierbei bezeichnet $H_m(\omega)$ jeweils die diskrete Fourier Transformierte der Impulsantwort $h_m(n)$ des m-ten Analysefilters.

Die Optimierungsschleife basiert dann auf einer Auswertung der Kostenfunktion E bzgl. des Erreichens bzw. Unterschreitens einer Grenze E_{min} für den Abbruch der Iteration. Bei Nichterreichen des Abbruchkriteriums wird die Matrix A_k des aktuellen Zeitschrittes gemäß des verwendeten Algorithmus zur Minimierung der Kostenfunktion um die Schrittweite δ_{A_k} modifiziert.

In der Originalversion des Nayebi-Algorithmus in [NBS92] wird die Berechnung der Richtung und Weite eines Gradientenschrittes mit der Methode *konjugierter Gradienten* vorgeschlagen. Die Formulierung des Optimierungsproblems erlaubt grundsätzlich natürlich aber auch andere Verfahren zur nichtlinearen Optimierung ohne Nebenbedingung. In der in dieser Arbeit erfolgten Implementierung wurden daher die in dem Softwarepaket *Matlab* verfügbaren Routinen *fminunc* für eine unbeschränkte multivariate Optimierung verwendet, die für die vorliegende Problemstellung unter Zuhilfenahme des *Line Search*-Algorithmus arbeiten.

Des Weiteren werden, wie ursprünglich von Nayebi et al. vorgeschlagen, für die Subband-Filter Symmetriebedingungen eingeführt, die die Impulsantworten und damit Übertragungsfunktionen der oberen $\lfloor M/2 \rfloor$ Bänder aus den unteren $\lfloor M/2 \rfloor$ durch Spiegelung mit $h_{M-i+1}(n) = (-1)^n h_i(n)$ ableiten. Die Prototypfilter A_0 bei Optimierung von Filterbänken mit geringer Gruppenlaufzeit stammen aus einem *Equiripple Entwurf* für minimalphasige Filter oder werden als Näherung durch Verschiebung der Impulsantwort eines Entwurfs für ein linearphasiges Filter generiert.

Mit der aus dieser Arbeit hervorgegangenen Implementierung konnte eine relativ schnelle und sichere Konvergenz der Optimierung erzielt werden. Wie schon von Nayebi et al. in [NBS92] angedeutet, hängt der Ausgang der Optimierung dabei jedoch verhältnismäßig stark von den eingestellten Gewichten der Kostenfunktion (vgl. Gleichung 2.3) ab. Der einzige Hinweis, der allerdings in [NBS92] für die Auslegung der Gewichtsfaktoren angegeben wird, ist

$$w_1 >> w_j \quad j = 2, \dots, J. \quad (2.8)$$

Dies bedeutet, dass der Rekonstruktionsfehler deutlich stärker gewichtet wird als die Abweichungen der Subband-Frequenzgänge. Unter Einhaltung dieser Bedingung konvergiert das Verfahren zuverlässig und generiert Subband-Frequenzgänge mit annehmbaren Verläufen. Ein manuelles unabhängiges Einstellen der einzelnen Gewichte führt aber in den meisten Fällen zu deutlich besseren Sperrdämpfungen bei vergleichbarer Rekonstruktionsqualität.

Zusätzlich zu dem Einfluss der Gewichtsfaktoren wurde eine Abhängigkeit des Optimierungsergebnisses von den zur Initialisierung verwendeten Prototypfiltern beobachtet. Hierbei ist es für Entwürfe mit geringer Gruppenlaufzeit besonders wichtig, Prototypen zu verwenden, deren Schwerpunkte der Impulsantwort bereits in der

Nähe, des im finalen Entwurf angestrebten und sich aus der eingestellten Latenz ergebenden, liegen.

2.2.2 Entwurfsverfahren nach Keiler/Nguyen

In der Dissertation [Kei06] wird nach einer allgemeinen Einleitung zu möglichen Strukturen von Analyse-/Synthesefilterbänken sowie Verzerrungsfunktionen gleichförmiger Filterbänke ein Verfahren zum Entwurf cosinus-modulierter Filterbänke mit nahezu perfekter Rekonstruktion vorgestellt. Dieses wird anschließend auf Basis von Methoden aus [NH96] so modifiziert, dass der Entwurf von Filterbänken mit reduzierter und nahezu frei einstellbarer Latenzzeit ermöglicht wird.

Ausgangspunkt der Beschreibungen bilden die Ansätze zu cosinus-modulierten Filterbänken aus [Vai93]. Hierbei wird auf Basis einer komplex modulierten Filterbank mit $2M$ Bändern durch gewichtete Überlagerung unter Zuhilfenahme des so genannten *Pseudo-QMF*-Prinzips (PQMF) eine Aufteilung des Eingangsspektrums in M Bänder erreicht. Keiler beschreibt darauf aufbauend die Abhängigkeiten der Analyse- und Synthesefilterkoeffizienten von einem gemeinsamen Tiefpass-Prototyp sowie dessen Polyphasenzerlegung für eine effiziente Durchführung der Filteroperationen von cosinus-modulierten Filterbänken in einer Polyphasenstruktur⁴.

Weitergehend wird der Entwurf eines linearphasigen Tiefpass-Prototypen über das Entwurfsverfahren nach Creusere und Mitra [CM95] beschrieben. Dieses wird durch eine Modifikation zur automatischen Anpassung des innerhalb des Optimierungsverfahrens einzustellenden Verhältnisses der Gewichtsfaktoren von Durchlass- und Sperrbereich erweitert. Mit dieser ist es möglich, eine leichte Verbesserung der Amplitudenverzerrungen der resultierenden Filterbank zu erreichen.

Der Entwurf von Filterbänken mit besonders geringer Latenz wird über eine Umformulierung gemäß [Ngu95, NH96] durch Einführung einer Polyphasenrealisierung mit Nutzung eines nicht-linearphasigen Synthese-Prototyps erreicht. Es wird eine Methode zur numerischen Optimierung dieses Prototyps inklusive der benötigten Kostenfunktion und Methode zur Berechnung der partiellen Ableitungen gemäß der *Quadratic Constrained Least Squares*-Methode (QCLS) beschrieben. In dieser wird die Energie im Sperrbereich des Prototyps minimiert und über die Einführung der namensgebenden quadratischen Nebenbedingung die perfekte Rekonstruktion näherungsweise sichergestellt.

Abbildung 2.4 fasst den gesamten Ablauf zum Entwurf einer Filterbank in einem Flussgraphen zusammen. Am Eingang der Optimierung sind die Parameter der Filterbank gemäß den Einschränkungen bei der Formulierung des Verfahrens einzustellen, wonach die Länge des Prototypfilters N auf ein Vielfaches des Doppelten

⁴In dieser Arbeit wird der Aspekt einer Polyphasenrealisierung nur indirekt bei der Bewertung der Entwurfsverfahren aufgegriffen und es erfolgt keine Implementierung in Polyphasenstruktur, da für die Evaluation der Filterbankeigenschaften eine Realisierung in der Struktur nach Abb. 2.1 ausreichend ist.

2M der Bandanzahl M beschrkt wird. Der dabei auch anzugebende Parameter s bestimmt die Gruppenlaufzeit der Filterbank, die sich daraus ber

$$\tau(s) = 2M(s+1) - 1 \quad \text{mit} \quad s = 0, \dots, \frac{N}{M} - 1 \quad (2.9)$$

berechnet.

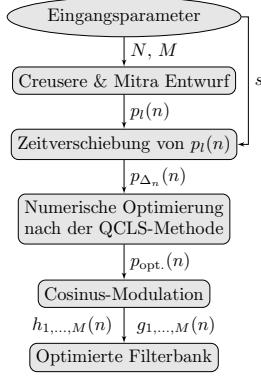


Abbildung 2.4: Flussgraph der Filterbank-Optimierung nach Keiler/Nguyen.

Gemäß der Eingangsparameter wird nun zunächst ein linearphasiges Prototypfilter $p_l(n)$ über das Optimierungsverfahren nach Creusere und Mitra entworfen. Dieses basiert auf dem Parks-McClellan-Algorithmus zum „Equiripple-Filterentwurf“ (siehe [Kei06] S. 77 ff.). Hiermit wird in einem iterativen Verfahren, über eine Variation der Durchlassfrequenz, der Frequenzgang eines Prototypfilters für eine cosinus-modulierte Filterbank optimiert. Die dabei ausgewertete Kostenfunktion erfasst die auftretenden Amplitudenverzerrungen bei Überlagerung zweier modulierter Prototypfilter gemäß des Pseudo-QMF-Prinzips.

Der aus dieser Optimierung hervorgehende linearphasige Prototyp wird gemäß des eingestellten Verzögerungsparameters s in seiner Impulsantwort zeitlich verschoben, wobei die Kausalität des Filters durch entsprechendes Abschneiden der Impulsantwort sichergestellt wird. Die verschobene Impulsantwort $p_{\Delta_n}(n)$ ist dann Startpunkt der eigentlichen numerischen Optimierung. Diese erfolgt mittels der bereits beschriebenen QCLS-Methode als gradientenbasierte Optimierung mit einer Nebenbedingung, die die näherungsweise perfekte Rekonstruktion sicherstellt.

Mit dem optimierten Prototyp $p_{\text{opt.}}(n)$ lassen sich dann über entsprechende Modulationsgleichungen die Impulsantworten der Analyse- und Synthesefilter berechnen. Für die Anwendung in einer Polyphasenstruktur kann alternativ eine Umrechnung in die Polyphasenkoeffizienten erfolgen.

Die Algorithmen von Keiler wurden im Rahmen der Studienarbeit [Kra12] sowie der Untersuchungen zu [PKO12] implementiert und analysiert. Hierbei lag der Fokus auf dem Entwurf von Filterbänken mit besonders geringer Latenz sowie dem Ausloten

der Grenzen des Verfahrens. Es wurden Entwürfe mit unterschiedlichen Bandanzahlen und Verzögerungen generiert sowie diverse „Parameter Sweeps“ durchgeführt. Grundsätzlich zeichnet sich das Optimierungsverfahren durch seinen moderaten Berechnungsaufwand und die guten Konvergenzeigenschaften aus. Es ließen sich für verschiedenste gerade Bandanzahlen Entwürfe mit Verzögerungen unter einer Millisekunde erzielen, die zumeist gute Frequenzverläufe und Rekonstruktionseigenschaften aufwiesen. Prinzipiell erlaubt die Formulierung auch ungerade Bandanzahlen, jedoch wurden hierfür keine vergleichbar guten Ergebnisse erreicht. Bei besonders geringen Latenzen traten zusätzlich verstärkt Welligkeiten im Durchlassbereich der Tief- und Hochpässe auf, wie die Ergebnisse in Abschnitt 2.3 zeigen werden.

2.2.3 Entwurfsverfahren nach Schuller

Eine Alternative zu den Verfahren von Nayebi und Keiler/Nguyen sind die in der Dissertation [Sch97] und verschiedenen Veröffentlichungen wie [SS96] und [SS95] von Schuller vorgestellten Beschreibungsformen und Algorithmen zum Entwurf von zeitvarianten⁵ Filterbänken mit niedriger Verzögerung und perfekter Rekonstruktion. Bei der Entwicklung der Methoden wurde dabei das Ziel verfolgt, Filterbänke mit einer von der Filterlänge unabhängigen Verzögerung trotz langer Impulsantworten zu entwerfen. Dies dient z.B. dem Verringern von Vorechos bei der Anwendung in der Codierung nach dem MPEG AAC Standard. Der Fokus liegt damit auf Filterbänken mit im Vergleich zum vorliegenden Anwendungsfall deutlich höheren Bandanzahlen und Verzögerungen. Trotzdem lassen sich mit dem Entwurfsverfahren durch dessen freie Formulierung auch Entwürfe mit niedriger Bandanzahl und besonders geringer Verzögerung erstellen.

Das Verfahren basiert auf einer Beschreibung einer cosinus-modulierten *M*-Kanal Filterbank über eine spezielle Formulierung in Polyphasenstruktur, die das nahezu freie Einstellen der Systemverzögerung ermöglicht. Abbildung 2.5 zeigt den grundsätzlichen Aufbau der dabei zur Anwendung kommenden Filterbankstruktur. Am Eingang der Filterbank findet im Analyseteil durch Verzögerung und Unterabtastung eine Zerlegung des Eingangssignales in seine Polyphasenkomponenten statt. Die Verarbeitung zu den Subband-Signalen erfolgt über die Analyse-Polyphasen-Matrix $\mathbf{P}_a(z)$. In der Synthesefilterbank werden zunächst mittels der Polyphasen-Matrix $\mathbf{P}_s(z)$ aus den Subband-Signalen wieder die Polyphasenkomponenten gebildet. Diese können dann durch Überabtastung und erneute Verarbeitung in einer Struktur aus Verzögerungsgliedern sowie eine Summation zum Berechnen des Rekonstruktionssignales $\tilde{x}(n)$ genutzt werden.

⁵ Neben einer Entwurfsmethode für Filterbänke mit geringer Verzögerung besteht ein wesentlicher Beitrag der Arbeiten zu [Sch97] in der Realisierung von Filterbänken mit zeitvarianter Teilbandzahl. Da der Aspekt der Zeitvarianz in dieser Arbeit nicht zur Anwendung kommt, wird dieser im Folgenden nicht weiter betrachtet.

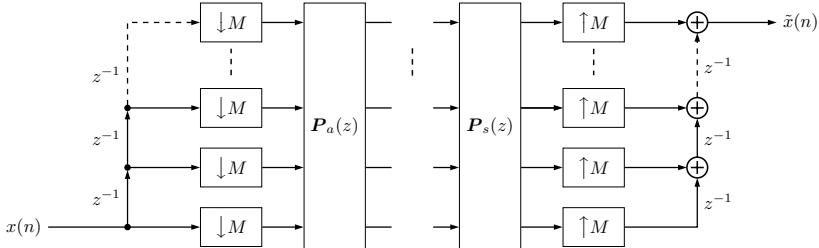


Abbildung 2.5: Flussgraph der M -Kanal Filterbank in Polyphasenstruktur für den Entwurf nach Schuller (gemäß [Sch97]).

Schuller zeigt in [Sch97], wie sich die Analyse- und Synthese-Polyphasen-Matrizen in das Produkt aus mehreren Teilmatrizen zerlegen lassen, die als *Transformations-, Verschiebungs- und Filtermatrizen* bezeichnet werden. Die Filtermatrizen können durch weitere Zerlegung in das Produkt aus so genannten *Minimal- und Maximal-Verzögerungs-Matrizen* überführt werden, wobei über letztere ein Einstellen der Verzögerung möglich wird. Damit ergibt sich unter Einhaltung gewisser Nebenbedingungen eine von der Filterlänge unabhängige und damit nahezu frei einstellbare Verzögerung der gesamten Filterbank, die im Minimum $M - 1$ betragen kann.

Der eigentliche Entwurf von Filterbänken mit der gewünschten, zunächst auf gerade Zahlen beschränkten, Bandanzahl und Filterlänge wird durch eine numerische Optimierung mittels einer an die Problemstellung angepassten Version der Methode der konjugierten Gradienten durchgeführt. Abbildung 2.6 fasst das dafür von Schuller vorgeschlagene Vorgehen in einem Flussgraphen grob zusammen. Der dabei einge-führte Vektor \boldsymbol{x} enthält die Koeffizienten, aus denen die Filtermatrizen der Analyse- und Synthesefilterbank resultieren.

In einem ersten Schritt wird eine Initialisierung des Koeffizientenvektors mit Zufallszahlen durchgeführt. Aus diesem werden die Polyphasen-Matrizen und damit die Frequenzgänge der Subband-Filter ermittelt. Über eine Kostenfunktion, die die gewichteten Abweichungen von einem vorgegebenen Wunschkontrollgang erfasst, wird dann für eine Approximation eines so genannten *Newton-Schrittes* ein Satz von Richtungsvektoren berechnet, der gemäß der Methode konjugierter Gradienten eine unabhängige Minimierung in möglichst wenigen Schritten ermöglichen soll. Die Optimierung als Gradientenabstieg innerhalb eines Iterationsschrittes erfolgt also, indem für jeden der Richtungsvektoren, deren Anzahl der an Koeffizienten des Vektors \boldsymbol{x} entspricht, eine eindimensionale Minimierung durchgeführt wird. Die Kostenfunktion wird dabei auch dazu genutzt, eine Skalierung der Schrittweite durchzuführen, falls ihr initialer Wert nicht zu einer Verringerung der Kosten geführt hat. Dieses Vorgehen wird in einer Vielzahl von Iterationen solange wiederholt, bis eine vorgegebene Schwelle für die Veränderung $\Delta\boldsymbol{x}$ des Koeffizientenvektors im Vergleich zur vorherigen Iteration unterschritten wird.

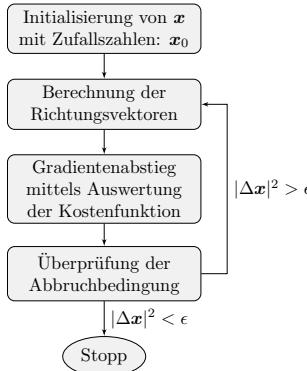


Abbildung 2.6: Flussgraph der Filterbank-Optimierung nach Schuller.

Das Optimierungsverfahren von Schuller wurde ebenfalls im Rahmen von [Kra12] und den Untersuchungen zu [PKO12] implementiert und evaluiert. Hierbei lag der Fokus, wie auch schon für die zuvor vorgestellten Entwurfsverfahren, auf der Optimierung von Filterbänken mit besonders geringer Verzögerung. Grundsätzlich ließen sich mit dem Verfahren Entwürfe mit weniger als einer Millisekunde Gesamtverzögerung und annehmbaren Frequenzverläufen generieren. Diese zeigten die, gemäß der Formulierung, zu erwartenden sehr guten Rekonstruktionseigenschaften.

Problematisch stellte sich die, in der vorliegenden Implementierung, langsame „Konvergenz“ des Verfahrens heraus. Hierbei werden zwar schon nach wenigen Iterationen erkennbare Frequenzverläufe der Subband-Filter erreicht. Für das Erstellen von Entwürfen mit glatten Verläufen im Durchlassbereich und ausreichend hohen Sperrdämpfungen mussten jedoch Berechnungsdauern von mehreren Stunden auf einem herkömmlichen Desktop-PC und ohne Parallelisierung in Kauf genommen werden. Dabei wurde eine zusätzliche starke Abhängigkeit des Optimierungsergebnisses von dem, für die Kostenfunktion zu definierenden, Wunschkontinuiergang und dem verwendeten Startvektor festgestellt. Aufgrund dieser Einschränkung waren größer angelegte Parametersweeps nicht möglich. Zusätzlich zeigte das Verfahren eine leichte Tendenz zu Welligkeiten im Durchlassbereich des Tief- und Hochpasses für Entwürfe mit besonders geringer Latenz, die jedoch deutlich geringer ausfallen wie für das Entwurfsverfahren nach Keiler/Nguyen.

2.3 Vergleich verschiedener Filterbankentwürfe

Mit den, in den vorherigen Abschnitten vorgestellten, Entwurfsverfahren wurden im Rahmen von [PO11], [Kra12] und [PKO12] verschiedene Filterbankkonfigurationen optimiert und in ihren Eigenschaften analysiert. Ziel war es dabei, innerhalb der

Einschränkungen des jeweiligen Verfahrens, Filterbänke mit vier bis acht Bändern und einer Verzögerung von weniger als einer Millisekunde zu erhalten. Im Folgenden werden die im Rahmen von [PKO12] jeweils als am erfolgversprechendsten ausgewählten Entwürfe dargestellt und miteinander verglichen.

Abbildung 2.7 zeigt die Frequenzgänge der Analysefilter verschiedener Entwurfsergebnisse. Zur besseren Übersicht wurden jeweils nicht alle Bänder dargestellt, da sie vergleichbare Verläufe aufweisen und sich z.B. für das Nayebi-Verfahren durch eingeführte Symmetrieverbedingungen aus den dargestellten Frequenzverläufen ergeben. Teilabbildung (a) beinhaltet die Frequenzgänge einer mit dem Nayebi-Verfahren optimierten 5-Kanal Filterbank mit 55 Filterkoeffizienten pro Band und 24 Samples Gesamtverzögerung. In Teilabbildung (b) sind die Frequenzgänge einer mit dem Keiler-Verfahren erstellten 6-Kanal Filterbank mit 96 Filterkoeffizienten pro Band und 35 Samples Gesamtverzögerung dargestellt. Die Teilabbildung (c) enthält die Frequenzgänge einer mit dem Schuller-Verfahren generierten 4-Kanal Filterbank mit 80 Filterkoeffizienten pro Band und 39 Samples Gesamtverzögerung. Die jeweiligen Synthesefilter weisen vergleichbare Verläufe auf und werden daher nicht gesondert dargestellt.

Die Ergebnisse veranschaulichen die bereits in den vorherigen Abschnitten ange deutete Tendenz der Verfahren nach Keiler/Nguyen und Schuller zu Welligkeiten im Durchlassbereich der Tiefpässe⁶ für Entwürfe mit besonders geringer Latenz. Bei den Ergebnissen nach Keiler/Nguyen beträgt die Überhöhung im Maximum zwar mehrere dB, stellt aber bei entsprechender Auslegung der Subband-Signalverarbeitung kein größeres Problem dar. Dafür ergibt sich eine über den gesamten Sperrbereich betrachtet relativ hohe Dämpfung. Beim Nayebi-Entwurf fällt diese besonders für den ersten Sperrbereich des ersten Bandpasses etwas geringer aus. Die Frequenzver läufe der mit dem Verfahren nach Schuller optimierten Filterbank weisen eine relativ konstante, jedoch nicht ganz so hoch ausfallende Dämpfung im Sperrbereich und nur eine leichte Überhöhung im Durchlassbereich des Tiefpasses auf.

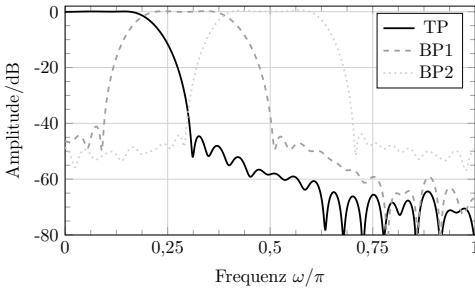
Zum Vergleich der mit den verschiedenen Verfahren entworfenen Filterbänke wurden neben einer visuellen Inspektion der Frequenzgänge in Anlehnung an [Kei06] auch Messwerte ermittelt, die die Frequenz- und Rekonstruktionseigenschaften erfassen. Eine wichtige Kenngröße ist dabei die minimale Dämpfung der Filter im Sperrbereich, also das Maximum des Betragsfrequenzgangs in Dezibel. Dieses lässt sich über

$$H_{S,\min.} = \max_{0 \leq i \leq M-1} \left\{ \max_{\omega_{S,i}} \left\{ 20 \log \left(|H_i(e^{j\omega})| \right) \right\} \right\} \quad (2.10)$$

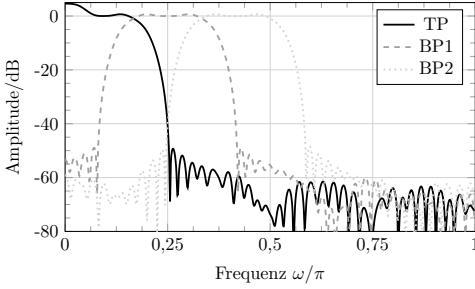
in einer Formel festhalten. Die als Sperrbereich $\omega_{S,i}$ mit $i = 0, \dots, M-1$ definierten Abschnitte des Frequenzgangs sind hierbei über

$$\frac{3\pi}{2M} \leq \omega_{S,0} \leq \pi, \quad (2.11)$$

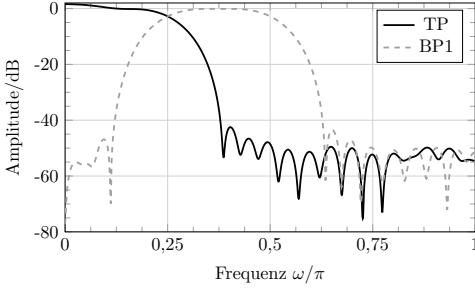
⁶Im Hochpass ergibt sich eine vergleichbare Absenkung des Amplitudenverlaufes und in den Synthesefiltern ein entsprechend zu den Analysefiltern inverses Verhalten.



(a) Nayebi 5-Kanal Filterbank mit $N = 55$ und $\tau = 24$.



(b) Keiler/Nguyen 6-Kanal Filterbank $N = 96$ und $\tau = 35$.



(c) Schuller 4-Kanal Filterbank $N = 80$ und $\tau = 39$.

Abbildung 2.7: Vergleich der Analysefilter-Betragsfrequenzgänge von Filterbänken mit geringer Latenz, die mit verschiedenen Entwurfsverfahren optimiert wurden.

$$0 \leq \omega_{S,i} \leq \frac{(2i-1)\pi}{2M} \quad \text{und} \quad \frac{(2i+3)\pi}{2M} \leq \omega_{S,i} \leq \pi, \quad (2.12)$$

und

$$0 \leq \omega_{S,M-1} \leq \frac{(2M-3)\pi}{2M} \quad (2.13)$$

beschreibbar und damit jeweils die Mittenfrequenzen der benachbarten Bänder. Die mittlere Dämpfung im Sperrbereich wird als Gesamtmittelwert aus einzelnen Mittelwerten über

$$H_{S,\text{mtl.}} = 20 \log \left(\frac{1}{M} \sum_{i=0}^{M-1} H_{i,S,\text{mtl.}} \right) \quad (2.14)$$

in Dezibel dargestellt. Dabei ergeben sich die Mittelwerte der einzelnen Filter mit

$$H_{0,S,\text{mtl.}} = \frac{1}{\pi - \frac{3\pi}{2M}} \int_{\frac{3\pi}{2M}}^{\pi} |H_0(e^{j\omega})| d\omega, \quad (2.15)$$

$$H_{i,S,\text{mtl.}} = \frac{1}{\frac{(2i-1)\pi}{2M}} \int_0^{\frac{(2i-1)\pi}{2M}} |H_i(e^{j\omega})| d\omega + \frac{1}{\pi - \frac{(2i+3)\pi}{2M}} \int_{\frac{(2i+3)\pi}{2M}}^{\pi} |H_i(e^{j\omega})| d\omega \quad (2.16)$$

und

$$H_{M-1,S,\text{mtl.}} = \frac{1}{\frac{(2M-3)\pi}{2M}} \int_0^{\frac{(2M-3)\pi}{2M}} |H_{M-1}(e^{j\omega})| d\omega \quad (2.17)$$

aus einer Integration über den entsprechenden Frequenzbereich und einer Normierung auf dessen Breite. Für die Synthesefilter $G_i(z)$ gelten die Berechnungsformeln 2.14 und 2.10 mit den entsprechenden Bandgrenzen äquivalent.

Eine weitere wichtige Messgröße erfasst die Aliasingverzerrungen der Filterbank. Zu deren Berechnung wird durch Überlagerung modulierter Teilträger mittels⁷

$$F_{\text{Alias.,max.}} = \max_{0 \leq \omega \leq \pi} \left\{ 20 \log \left(\sqrt{\sum_{l=1}^{M-1} \left| \sum_{i=0}^{M-1} G_i(e^{j\omega}) H_i(e^{j(\omega - \frac{2\pi}{M} l)}) \right|^2} \right) \right\} \quad (2.18)$$

überprüft, wie sehr die möglicherweise nicht perfekte Aliasingkompensation im Maximum über den gesamten Frequenzbereich vom Idealwert 0 bzw. $-\infty$ abweicht.

Zuletzt lassen sich die Amplitudenverzerrungen als Abweichung der Gesamtübertragungsfunktion vom Idealwert 1 über

$$F_{\text{Ampl.,max.}} = \max_{0 \leq \omega \leq \pi} \left\{ 20 \log \left(\left| 1 - \left| \sum_{i=0}^{M-1} H_i(e^{j\omega}) G_i(e^{j\omega}) \right| \right| \right) \right\} \quad (2.19)$$

ermitteln und ebenfalls in Dezibel ausdrücken.

⁷Abweichend von der in [Kei06] verwendeten Notation wird hier der für Amplitudengrößen übliche Vorfaktor 20 verwendet.

Zusammen mit dem bereits in Abschnitt 2.1 definierten SNR_r lässt sich so ein relativ vollständiges Bild der Eigenschaften eines Filterbankentwurfes erstellen. Tabelle 2.1 fasst alle in dieser Arbeit verwendeten Eigenschaften noch einmal in einer Übersicht zusammen.

Tabelle 2.1: Beschreibung der Messgrößen zur Beurteilung der Filterbankentwürfe.

Eigenschaft/ Bewertungsgröße	Erläuterung
M	Anzahl an Teilbändern der Filterbank
N	Anzahl der Subband-Filterkoeffizienten
τ	Gruppenlaufzeit der Filterbank in Samples
$H_{S,\text{mtl.}}$	Mittlere Sperrdämpfung der Analysefilter in dB
$H_{S,\text{min.}}$	Minimale Sperrdämpfung der Analysefilter in dB
$G_{S,\text{mtl.}}$	Mittlere Sperrdämpfung der Synthesefilter in dB
$G_{S,\text{min.}}$	Minimale Sperrdämpfung der Synthesefilter in dB
$F_{\text{Alias.,max.}}$	Maximale Aliasingverzerrung der Filterbank in dB
$F_{\text{Ampl.,max.}}$	Maximale Amplitudenverzerrung der Filterbank in dB
SNR_r	Gesamtrekonstruktionsfehler der Filterbank in dB

Tabelle 2.2 beinhaltet die Ergebnisse der Messgrößen für die in 2.7 dargestellten Filterbänke. Die bei den Entwürfen nach Keiler/Nguyen und Schuller bewusst zugelassene höhere Filterlänge begründet sich dabei durch die Möglichkeit einer Implementierung als effiziente Polyphasenstruktur. Die Ergebnisse verdeutlichen den über ihre Formulierung als Verfahren mit näherungsweise perfekter bzw. perfekter Rekonstruktion bestehenden Unterschied zum Verfahren nach Nayebi und sind im Endeffekt lediglich durch die „Double-Precision“-Rechengenauigkeit begrenzt. Des Weiteren wird die bereits aus den Frequenzgängen ersichtliche etwas höhere Sperrdämpfung des Entwurfes nach Keiler/Nguyen deutlich. Die Ergebnisse der Aliasing- und Amplitudenverzerrungen spiegeln sich insgesamt auch im SNR_r wider.

Um den Einfluss verschiedener Filterbankentwürfe auf die Subband-ADPCM-Codierung abzuschätzen, wurden im Zuge der Untersuchungen zu [PKO12] zusätzlich Evaluierungen der resultierenden Audioqualität in einer frühen Version des Codecs gemacht. Die unter anderem auch in ihren Tendenzen durch einen Hörversuch bestätigten Ergebnisse ließen folgende Schlüsse zu. Zum einen kann davon ausgegangen werden, dass die Rekonstruktionsqualität des Nayebi-Entwurfes trotz des schlechteren Abschneidens bei den Messwerten ausreichend hoch ist und Einflüsse auf die Audioqualität, die durch die nicht verlustfreie Codierung der Subband-Signale entstehen, überwiegen. Zum anderen deutet sich an, dass für gewisse Signalklassen eine weitere Verbesserung des Coderverfahrens durch Filterbankentwürfe mit einer erhöhten Sperrdämpfung erzielt werden können. In Abwägung der Latenz und Flachheit des Durchlassbereiches sowie unter Inkaufnahme des unter Umständen erhöhten Rechenaufwandes fiel die Entscheidung zugunsten des Nayebi 5-Band Ent-

wurfes aus. Auf Basis der gewonnenen Erkenntnisse wurde zusätzlich über die im nächsten Abschnitt 2.4 präsentierte Methode zur globalen Optimierung von Entwurfsparametern eine Verbesserung der Sperrdämpfung erreicht, die deren Differenz zum Keiler/Nguyen-Entwurf kompensiert. Auf einen Versuch der globalen Optimierung für das Keiler/Nguyen-Verfahren oder das Schuller-Verfahren wurde mangels nötiger Freiheitsgrade bzw. aufgrund des hohen Berechnungsaufwandes verzichtet.

Tabelle 2.2: Vergleich der Eigenschaften verschiedener Filterbankentwürfe über geeignete Bewertungsgrößen.

Eigenschaft/ Bewertungsgröße	Entwurfsverfahren		
	Nayebi	Keiler/Nguyen	Schuller
M	5	6	4
N	55	96	80
τ	24	35	39
$H_{S,\text{mtl.}}$	-51,4	-60,1	-52,2
$H_{S,\text{min.}}$	-33,2	-40,5	-35,6
$G_{S,\text{mtl.}}$	-52,8	-60,1	-52,0
$G_{S,\text{min.}}$	-33,2	-40,3	-35,3
$F_{\text{Alias.,max.}}$	-107,1	-274,3	-296,3
$F_{\text{Ampl.,max.}}$	-117,2	-274,7	-301,0
SNR_r	119,9	279,0	303,6

2.4 Globale Optimierung von Entwurfsparametern

Wie bereits in Abschnitt 2.2.1 erläutert, bedarf das Entwurfsverfahren nach Nayebi et al. der manuellen Einstellung diverser Gewichtsfaktoren der Kostenfunktion. Diese Parameter bestimmen den Anteil des Rekonstruktionsfehlers sowie der summierten Abweichungen der Subband-Frequenzgänge von den Verläufen idealer Filter zum Gesamtkostenwert. Mit einem manuellen Einstellen der Parameter können, wie im vorherigen Abschnitt gezeigt, in der Regel gute Ergebnisse erzielt werden. Der Einfluss der Parameter auf das Endergebnis hat sich jedoch als wenig intuitiv herausgestellt und damit ist die manuelle Anpassung der Gewichtsfaktoren für höhere Bandanzahlen sehr aufwändig. Zusätzlich lässt sich ohne Weiteres keine Aussage treffen, ob durch eine andere Gewichtung bessere Ergebnisse erzielt werden könnten.

Daher wurde in den Untersuchungen, die zu [PO15b] geführt haben, evaluiert, ob es mittels Techniken der globalen Optimierung möglich ist, den Prozess des manuellen

Einstellens von Gewichtsfaktoren durch einen vollkommen automatisierten Prozess zu ersetzen. Abbildung 2.8 zeigt einen Flussgraphen der dafür eingeführten globalen Optimierung der Nayebi-Parameter.

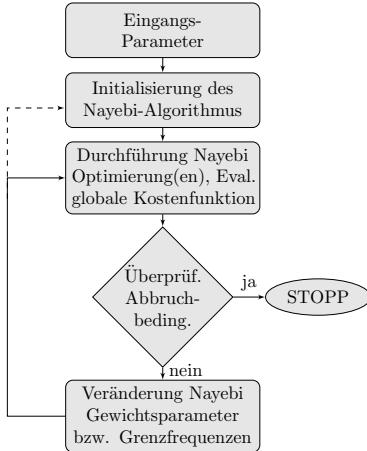


Abbildung 2.8: Flussgraph der globalen Optimierung der Nayebi-Parameter.

Eingangsparameter der Optimierung sind demnach nur noch die Bandanzahl, die zur Verfügung stehende Anzahl an Filterkoeffizienten, die gewünschte Gruppenlaufzeit der Filterbank sowie optionale zu verwendende Symmetriebedingungen. Hiermit erfolgt die Initialisierung des Nayebi-Algorithmus durch den Entwurf der Prototypfilter und damit das Erstellen der Matrizen \mathbf{A}_0 sowie \mathbf{B} . In der Iterationsschleife werden nun für jede Iteration verschiedene Kombinationen der Gewichtsparameter, je nach Algorithmus für die globale Optimierung, eine oder mehrere Nayebi-Optimierung(en) durchgeführt und die globale Kostenfunktion

$$E_{\text{global}} = \frac{1}{\text{SNR}_r [\text{dB}]} + \sum_{j=2}^J \epsilon_j \quad (2.20)$$

evaluiert. Diese erfasst im ersten Summanden den Rekonstruktionsfehler der Filterbank über das Reziproke des SNR_r in dB für ein Einheitsrampsignal in einer nichtlinearen Form. Die anderen Summanden bilden, ähnlich wie auch schon die Kostenfunktion des Nayebi-Algorithmus, die Abweichungen der Subband-Filter von den Idealverläufen ab. Die Summanden sind damit in der gleichen Größenordnung, so dass es keiner weiteren Gewichtung bedarf.

Für das Lösen des eigentlichen Optimierungsproblems wurden im Rahmen dieser Arbeit verschiedene Verfahren zum Minimieren globaler Kostenfunktionen evaluiert. Da jede Iteration der Optimierungsschleife mindestens einen Aufruf der Kostenfunktion und damit Durchführung einer Nayebi-Optimierung bedeutet, die selbst mehr

als eine Minute dauern kann und aufgrund des guten Konvergenzverhaltens wurden die besten Ergebnisse mit einem *Pattern-Search*-Ansatz [HJ61] erzielt. Bei diesem Verfahren zur direkten Suche im Parameterraum unstetiger und nicht differenzierbarer Kostenfunktionen wird dieser gemäß des namensgebenden Musters (engl. Pattern) systematisch abgesucht, wobei der Suchradius im Laufe der Iterationen verkleinert wird. Je nach Variante des Algorithmus wird dabei zusätzlich eine zum Teil zufällige Variation der Suchrichtungen vorgenommen oder die Suche innerhalb einer Iteration abgebrochen, sobald eine Minimierung der Kostenfunktion erfolgt ist. Das Vorgehen nach dem Pattern-Search-Algorithmus ermöglicht somit eine vergleichsweise robuste Konvergenz und erlaubte dadurch für die vorliegende Problemstellung ein besseres Abwägen zwischen der Qualität des Optimierungsergebnisses und der Laufzeit der Optimierung.

Eine mögliche Erweiterung des globalen Optimierungsverfahrens betrifft die Grenzen bei der Integration der Frequenzverlaufsfehler (Gleichung 2.6 und 2.7), also den Zuweisungen für $\omega \in$ Passband und $\omega \in$ Stopband. In [NBS92] werden hierfür keine näheren Angaben gemacht. Im Zuge der Entwicklung des globalen Optimierungsverfahrens hat es sich jedoch als vorteilhaft erwiesen, in ihnen einen weiteren Freiheitsgrad der Kostenfunktion zu nutzen. Demnach wurde eine kaskadierte Optimierung der Gewichtsparameter und Bandgrenzen eingeführt. Hiermit ist gemeint, dass in einem ersten Schritt die Gewichte der Nayebi-Kostenfunktion (Gleichung 2.3) global optimiert werden, während die Bandgrenzen aus Gleichung 2.6 und 2.7 festgehalten bleiben. In einem zweiten Schritt werden dann mit den ermittelten Gewichten die Bandgrenzen optimiert.

Wie im Flussgraphen in Abbildung 2.8 durch die gestrichelte Linie angedeutet, erlaubt das Verfahren zur globalen Optimierung auch einen Einbezug der Initialisierung des Nayebi-Algorithmus. Dies bezieht sich auf den Entwurf der, für den Start des Verfahrens nötigen, Prototypfilter, da sich, wie bereits in Unterabschnitt 2.2.2 angedeutet, deren Verläufe als für den Ausgang der Optimierung mitentscheidend herausgestellt haben.

Abbildung 2.9 beinhaltet einen Vergleich der Ergebnisse verschiedener Nayebi-Optimierungen für die bereits in Abschnitt 2.3 betrachtete Filterbank mit fünf Bändern, 55 Filterkoeffizienten pro Band und einer Gesamtlatenz von 24 Samples. Teilabbildung (a) zeigt Ergebnisse, die unter Einhaltung der von Nayebi et al. in [NBS92] angegebenen Bedingungen für die Einstellung der Gewichtsfaktoren erzielt wurden. In Teilabbildung (b) sind Frequenzverläufe eines Entwurfs dargestellt, bei dem durch manuelle Optimierung der Nayebi-Gewichtsparameter der Versuch einer Erhöhung der Sperrdämpfung vorgenommen wurde. In (c) sind Ergebnisse einer globalen Optimierung dieser Gewichtsparameter enthalten. Die Teilabbildung (d) verdeutlicht den Gewinn an Sperrdämpfung von bis zu 10 dB, der durch die vorgestellte kaskadierte Optimierung ohne einen signifikanten Verlust an Rekonstruktionsqualität erzielt wurde.

Weitere durchgeführte Optimierungsdurchläufe mit anderen Versionen der globalen Kostenfunktion ergaben Entwürfe mit höherer Sperrdämpfung zulasten von Welligkeiten im Durchlassbereich und einem Verlust an Rekonstruktionsqualität oder

dessen Verbesserung mit vergleichbaren Frequenzgängen zu Abbildung 2.9 (c). Das Verfahren zur globalen Optimierung wurde zudem auch dazu genutzt, Filterbänke mit anderen Bandanzahlen zu optimieren. Im Rahmen der Grenzen des Nayebi-Entwurfsverfahrens war es dabei möglich, vergleichbare Ergebnisse zu erzielen.

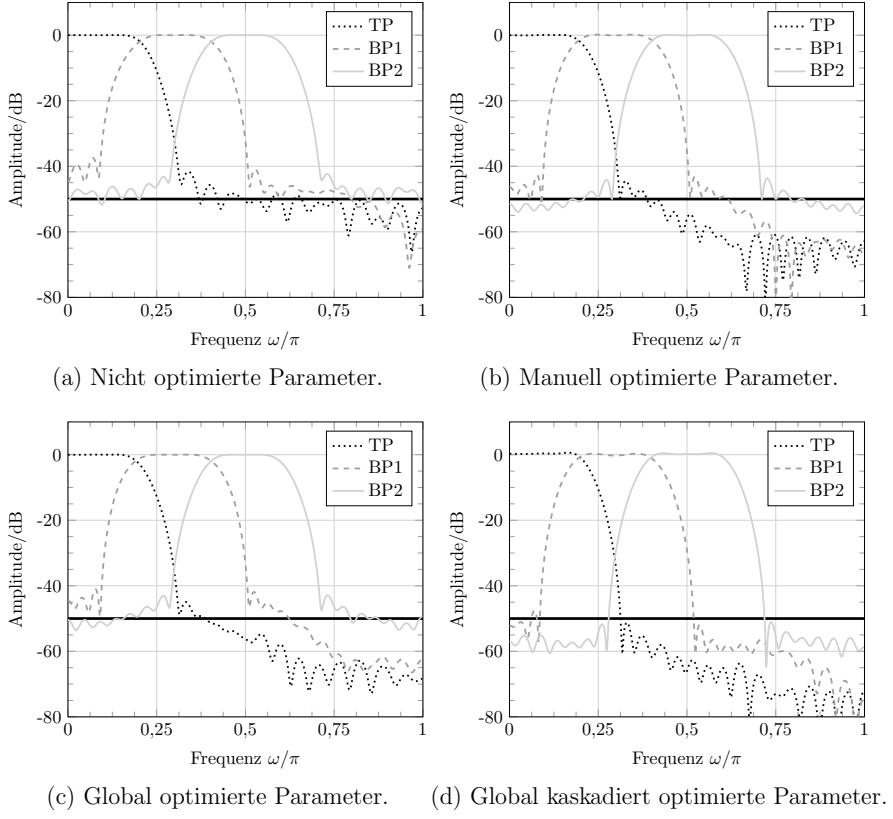


Abbildung 2.9: Vergleich der Betragsfrequenzgänge von Filterbankentwürfen nach dem Nayebi-Entwurfsverfahren für verschiedene Arten der Gewichtsparameteroptimierung.

Kapitel

Fehlerrobuste ADPCM-Codierung

3

Dieses Kapitel enthält mit der Algorithmik zur fehlerrobusten ADPCM-Codierung der Subband-Signale (Blöcke $AdPQ_i$ in Abbildung 1.3) den zweiten großen Baustein des zu entwickelnden Verfahrens zur fehlerrobusten Audiocodierung mit geringer Latenz. Hierbei werden, aufbauend auf bereits erprobten Verfahren zur adaptiven Prädiktion und Quantisierung, Maßnahmen zur Verbesserung der Audioqualität sowie zum Erreichen einer Robustheit gegenüber Kanalübertragungsfehlern vorgestellt und analysiert.

Ziel der folgenden Darstellungen ist es, das dabei gewählte Vorgehen zu motivieren und die Eignung der entwickelten Methoden für den vorliegenden Anwendungsfäll zu demonstrieren. Dazu werden nach einer allgemeinen Einleitung in die Grundlagen zur ADPCM die Algorithmen zur fehlerrobusten adaptiven Prädiktion und Quantisierung dargestellt. Des Weiteren erfolgt die Beschreibung einer Methode zur dynamischen Bit-Allokation und Erweiterung der Prädiktionsstruktur um einen statischen Anteil. Ausblickhaft werden zudem Untersuchungen zu einem Konzept zur Vektorquantisierung präsentiert, das in Kombination mit einer fehlerrobusten Breitband-ADPCM eine Verbesserung der Audioqualität bei Breitband-Codierung erlaubt und in weiterführenden Untersuchungen auf eine Subband-Codierung angewendet werden könnte.

Zusätzlich wird eine alternative Maßnahme zur Verbesserung der Audioqualität durch adaptive Vor- und Nachfilterung vorgestellt, die einen vielversprechenden Ansatz darstellt und deren Eignung demonstriert wird. Sie ist aber aufgrund ihres frühen Stadiums der Entwicklung und der geringen Praxisrelevanz nicht in die final evaluierte Codec-Version eingeflossen.

3.1 Grundlagen zur ADPCM

Das Grundprinzip der ADPCM basiert auf der Eigenschaft natürlicher Audiosignale zumindest abschnittsweise stationär zu sein. Dies macht es möglich, über spezielle Filterstrukturen im Rahmen einer so genannten *linearen Prädiktion*, auf Basis der Vergangenheit eines Signales, eine Vorhersage über dessen Verlauf für zukünftige

Zeitschritte zu machen. Da dies auf gleiche Weise sowie auf Basis der gleichen Signale im Encoder und Decoder erfolgen kann, ist es somit nur nötig, den Fehler der Prädiktion zwischen diesen zu übertragen. Aufgrund dessen geringerer Signalamplitude kann der Prädiktionsfehler einer adaptiven Quantisierung unterzogen werden, die mit einer geringeren Wortbreite im Vergleich zum Eingangssignal erfolgt. Dies resultiert in einer Reduktion der zu übertragenden Datenrate.

Abbildung 3.1 zeigt den Flussgraphen einer Signalübertragungsstrecke¹, die auf einem ADPCM-basierten Encoder und Decoder basiert. Am Eingang des Encoders wird von dem zu codierenden Signal² $x(n)$ die Prädiktion $p(n)$ subtrahiert. Der resultierende Prädiktionsfehler $e(n)$ wird im Block $Q_a(\cdot)$ einer adaptiven skalaren Quantisierung unterzogen, was einer Abbildung des Prädiktionsfehlersignals auf eine Folge von Quantisierungsindizes $\hat{q}(n)$ bedeutet, die an den Decoder übertragen werden.

Um auf dem Kanal nur die Quantisierungsindizes $\hat{q}(n)$ übertragen und auf zusätzliche Informationen wie Prädiktionsfilterkoeffizienten oder Normalisierungsfaktoren verzichten zu können, erfolgen jegliche Anpassungen rückwärtsadaptiv. Demnach beinhaltet der Encoder auch einen Decoderteil, so dass alle Adaptionen und Berechnungen³ auf Basis des rekonstruierten Prädiktionsfehlers $\tilde{e}(n)$ und des rekonstruierten Signales $\tilde{x}(n)$ erfolgen, die auch im Decoder vorliegen. Im Fall des Prädiktors wird dieses Vorgehen als *Rückwärtsprädiktion* bezeichnet.

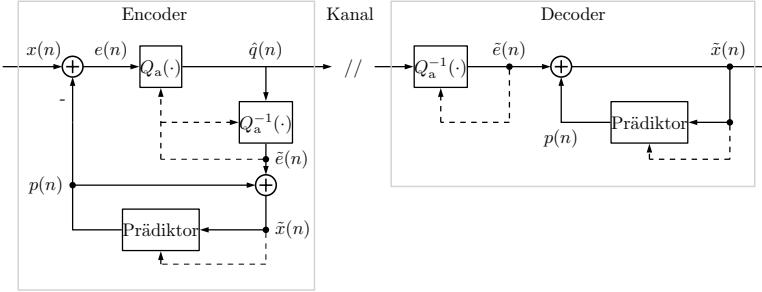


Abbildung 3.1: Flussgraph eines ADPCM-basierten Encoders und Decoders.

Abbildung 3.2 zeigt einen detaillierteren Flussgraph des ADPCM-Encoders mit der adaptiven Quantisierung durch Prädiktionsfehlernormalisierung mittels einer Hüllkurvenschätzung. Die Blöcke des adaptiven Quantisierungsoperators $Q_a(\cdot)$ und der Abbildung auf den entsprechenden Repräsentativwert und anschließender Denorma-

¹Die Darstellung erfolgt hier für einen allgemeinen Zeitindex n , auch wenn im Subband-Codec die ADPCM-Codierung der unterabgetasteten Subband-Signale mit dem Zeitindex m erfolgt.

²Das Eingangssignal wird dabei als mit 16 oder 24 Bit ausreichend fein quantisiert für die Ver nachlässigung von initialen Quantisierungseinflüssen angenommen.

³Gestrichelte Linien in Flussgraphen signalisieren im Folgenden stets Signale, die eine steuernde Funktion haben oder Berechnungsgrundlage für, in Funktionsblöcken verwendete, Adaptionsalgorithmen sind.

lisierung über $Q_a^{-1}(\cdot)$ sind dabei durch den Block zur Berechnung des Normalisierungsfaktors $v(n)$ über eine Hüllkurvenschätzung sowie den skalaren Quantisierungsoperator $Q(\cdot)$ und die Abbildung auf den entsprechenden Repräsentativwert $Q^{-1}(\cdot)$ ersetzt. Die adaptive Quantisierung besteht also in der skalaren Quantisierung des normalisierten Prädiktionsfehlers $e_{\text{norm}}(n)$, also der Abbildung auf eine begrenzte Anzahl an Repräsentativwerten $\tilde{e}_{\text{norm}}(n)$.

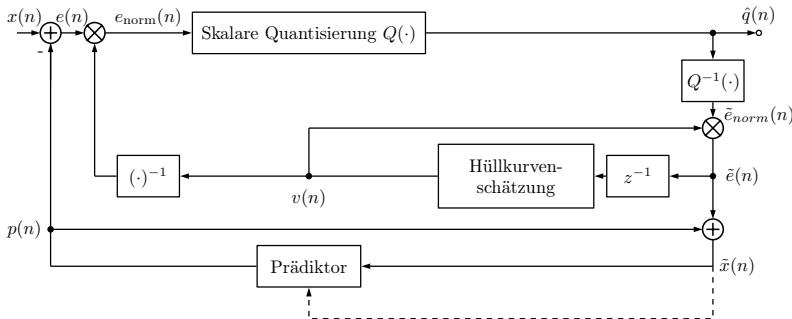


Abbildung 3.2: Detailierter Flussgraph des ADPCM-Encoders mit adaptiver Quantisierung durch Prädiktionsfehlernormalisierung über eine Hüllkurvenschätzung.

3.1.1 Lineare Prädiktion

Ein wesentlicher Teil des Codiergewinnes bzw. der Audioqualität der ADPCM hängt von der Güte der Prädiktion ab. Gelingt es mit dem Prädiktionssignal das Eingangssignal bis auf einen geringen Restfehler „vorherzusagen“⁴, kann die adaptive Quantisierung des Prädiktionsfehlers mit einer deutlich verringerten Auflösung erfolgen, ohne dass bei der Rekonstruktion im Decoder eine signifikante Abweichung zum ursprünglichen Signal entsteht.

Quantitativ lässt sich die Güte der Prädiktion in der Größe des so genannten *Prädiktionsgewinnes* über

$$G_p = \frac{\sum_n x^2(n)}{\sum_n e^2(n)} \quad (3.1)$$

als Verhältnis der Energien von Eingangssignal und Prädiktionsfehler berechnen. Um die zeitliche Veränderung des Prädiktionsgewinnes bei wechselnden Signaleigenschaften besser erfassen zu können, wird dieser häufig auch in nicht überlappenden kurzen Blöcken berechnet und aufgetragen.

Die einfachste Form der Realisierung eines linearen Prädiktors ist die in Direkt-

⁴Dies ist für ausreichend stationäre Signale, mit Unterschieden der Amplituden von Eingangssignal und Prädiktionsfehler im Bereich mehrerer Zehnerpotenzen, der Fall.

struktur. Abbildung 3.3 zeigt den Flussgraphen⁵ eines Vorwärtsprädiktionsfilters⁶ in Direktstruktur (a) und das dazu komplementäre Filter, welches als *Synthesefilter* wird (b). Hierbei ist das *Prädiktionsfilter* $P(z)$ lediglich ein FIR-Filter mit vorgeschalteter Einheitsverzögerung. Die Subtraktion des, über dieses Filter berechneten, Signales vom Eingangssignal ergibt den Prädiktionsfehler. Dies entspricht der Filterung des Eingangssignales $x(n)$ mit dem so genannten *Prädiktionsfehlerfilter* $F(z)$. Der sich daraus ergebende Prädiktionsfehler $e(n)$ wird im Synthesefilter $S(z)$, das aus einer Rückkopplung mittels des Prädiktionsfilters $P(z)$ besteht, derart verarbeitet, dass eine Rekonstruktion des Eingangssignales resultiert.

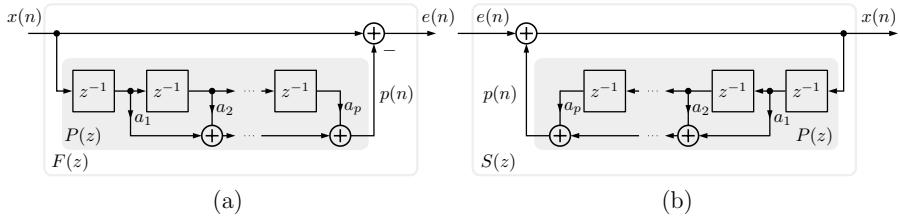


Abbildung 3.3: (a) Vorwärtsprädiktionsfehlerfilter mit Prädiktionsfilter in Direktstruktur (b) dazu komplementäres Synthesefilter.

Die Berechnung der Prädiktion erfolgt also als gewichtete Summation vergangener Signalwerte über

$$p(n) = \sum_{i=1}^p a_i x(n-i), \quad (3.2)$$

was der Modellierung des Signales als *Autoregressiver Prozess* entspricht. Die Anzahl p der Koeffizienten des Prädiktionsfilters und damit dessen Ordnung gleicht dabei der verwendeten Modellordnung. Für stationäre Signale kann schon mit wenigen Koeffizienten eine gute Prädiktion erfolgen, so dass sich ein hoher Prädiktionsgewinn und ein „rauschartiger“ Prädiktionsfehler ergibt. Durch die Filterung mit dem Prädiktionsfehlerfilter erfolgt also eine Dekorrelation des Eingangssignales, was der Entfernung redundanten Signalanteiles entspricht.

Während $F(z)$ ein FIR-Filter beschreibt, ergibt sich für das Synthesefilter eine IIR-Struktur. Da die Übertragungsfunktionen $F(z)$ und $S(z)$ exakt invers zueinander sein müssen, ist es nötig, bei der Bestimmung der Prädiktionsfilterkoeffizienten sicherzustellen, dass $F(z)$ ein stabil invertierbares Filter ergibt, was der Eigenschaft eines Minimalphasensystems gleichkommt.

⁵Die Darstellung erfolgt auch hier für ein allgemeines Eingangssignal $x(n)$ mit dem Zeitindex n , auch wenn bei der Verwendung in der ADPCM das Eingangssignal $\tilde{x}(n)$ ist.

⁶Als Vorwärtsprädiktion wird ein Vorgehen bezeichnet, bei dem die Prädiktion direkt aus dem Eingangssignal des Encoders berechnet wird.

3.1.2 Das Lattice-Filter

Die Berechnung des Prädiktionsfehlersignales und die Adaption des Prädiktors ist grundsätzlich mit der in Abbildung 3.3 dargestellten Struktur möglich. Dennoch hat es sich, gerade wegen der bereits erwähnten Notwendigkeit zum Sicherstellen der Stabilität des Synthesefilters, in der Praxis als effizienter erwiesen, dieses auf Basis einer so genannten *Kreuzgliedstruktur* (engl. Lattice-Ladder Structure⁷ [Wer07] (S. 68 ff.)) durchzuführen. Abbildung 3.4 zeigt den allgemeinen Aufbau eines solchen Lattice-Filters der Ordnung p .

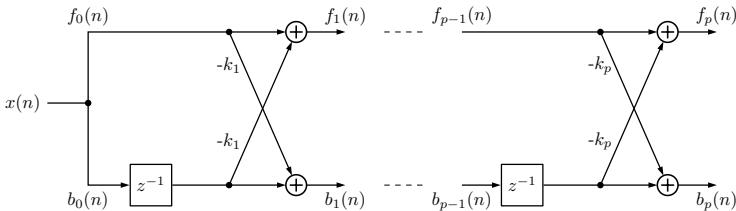


Abbildung 3.4: Lattice-Prädiktionsfilter der Ordnung p .

Das wiederum allgemein als $x(n)$ bezeichnete Eingangssignal wird dabei in zwei parallelen Zweigen verarbeitet. Diese sind über die p namensgebenden sich kreuzenden Äste miteinander verbunden, in denen die Multiplikation mit den *Lattice-Koeffizienten* k_m erfolgt. Der gesamte Aufbau besteht demnach aus p gleich aufgebaute Stufen. Der untere Zweig enthält dabei pro Stufe jeweils eine Einheitsverzögerung. Am Ausgang jeder Stufe ergibt sich der so genannte *Vorwärts-* und *Rückwärtsprädictionsfehler* $f_m(n)$ und $b_m(n)$ (engl.: forward, backward) für $m = 1, \dots, p$. Die Namensgebung des Vorwärts- und Rückwärtszweiges folgt aus den zueinander zeitlich inversen Impulsantworten der beiden sich ergebenden Prädiktionsfehlerfilter. Allgemein lässt sich damit für jede Stufe mit dem Stufenindex m der Vorwärts- und Rückwärtsprädictionsfehler über die Berechnungsvorschrift

$$f_m(n) = f_{m-1}(n) - k_m b_{m-1}(n-1) \quad (3.3)$$

$$b_m(n) = b_{m-1}(n-1) - k_m f_{m-1}(n) \quad (3.4)$$

mit den zunächst als konstant angenommenen Lattice-Koeffizienten k_m ableiten, wobei $f_0(n) = b_0(n) = x(n)$ gilt. Die Koeffizienten a_i und k_m von Lattice- und Direktstruktur lassen sich dabei über entsprechende Rekursionsformeln jeweils auseinander berechnen. Die Prädiktion bestimmt sich in der Lattice-Struktur über

$$p(n) = \sum_{m=1}^p k_m b_{m-1}(n-1). \quad (3.5)$$

⁷Im Folgenden wird der Einfachheit halber nur noch der Begriff *Lattice-Filter* verwendet.

Damit ist es möglich, das Prädiktionsfilter $P(z)$ aus Abbildung 3.3 durch ein Filter in Lattice-Struktur zu ersetzen.

Ein wesentlicher Vorteil der Berechnung und Adaption des Prädiktors in einer Lattice-Struktur resultiert aus der Tatsache, dass sich zeigen lässt, dass über die einfache Bedingung $|k_m| < 1$ für $m = 1, \dots, p$ die Stabilität des Synthesefilters sichergestellt ist [Mak77]. Des Weiteren stehen diverse Algorithmen zu Lattice-Prädiktoradaption zur Verfügung, die direkt oder indirekt eine Adaption der Lattice-Filterkoeffizienten auf Basis des Eingangssignales ermöglichen und sich für den vorliegenden Anwendungsfall der fehlerrobusten Prädiktion adaptieren lassen.

3.2 Fehlerrobuste adaptive Prädiktion

In den bisherigen Darstellungen zur linearen Prädiktion wurden die Direktstruktur- bzw. Lattice-Koeffizienten als zeitlich konstant angenommen. Um auf wechselnde spektrale Inhalte des zu prädizierenden Signales reagieren zu können und damit jeweils den maximal möglichen Prädiktionsgewinn zu erzielen, ist es jedoch notwendig, die Koeffizienten des Prädiktors zeitlich zu adaptieren. Dabei gilt es auf Veränderungen schnell reagieren und dennoch eine hohe Güte der Prädiktion für stationäre Signale gewährleisten zu können sowie den Aufwand für die Neuberechnung der Koeffizienten in Grenzen zu halten. Zusätzlich ist es, aufgrund der möglichen Einkopplung durch Fehler in dem zur Adaption genutzten Signal, wichtig sicherzustellen, dass deren Auswirkungen über der Zeit aus den Algorithmen zur Prädiktoradaption ausgekoppelt werden.

Daher werden im Folgenden Algorithmen zur Lattice-Prädiktoradaption kurz vorgestellt, diese in ihrer Eignung für den vorliegenden Anwendungsfall und ihrem Berechnungsaufwand verglichen und eine Methode zur fehlerrobusten Rückwärtsprädiktion aufgegriffen und diskutiert. Mit der Adaption auf das Fehlersignal wird ein weiteres Verfahren zur Prädiktoradaption präsentiert, das für den vorliegenden Anwendungsfall geeignet ist und eine algorithmisch sichergestellte Fehlerrobustheit ermöglicht.

3.2.1 Algorithmen zur Lattice-Prädiktoradaption

Der GAL-Algorithmus

Ein unter anderem bereits in [Hol09] erfolgreich erprobter Algorithmus ist die direkte gradientenbasierte Anpassung der Lattice-Koeffizienten auf Basis der Energien des Vorwärts- und Rückwärtsprädiktionsfehlers der einzelnen Lattice-Stufen. Bei der Formulierung im so genannten *Gradient Adaptive Lattice*-Algorithmus (GAL) wird demnach über die Ableitung der Summe der Fehlerenergien einer Lattice-Stufe nach dem jeweiligen Lattice-Koeffizient die Richtung des Gradientenschrittes berechnet. Hierdurch ergibt sich für die m -te Stufe der Lattice-Koeffizient des nächsten Zeit-

schrittes zu

$$k_m(n+1) = k_m(n) + \frac{\hat{\mu}}{\sigma_m^2(n) + \sigma_{\min}^2} \cdot (f_m(n) \cdot b_{m-1}(n-1) + b_m(n) \cdot f_{m-1}(n)), \quad (3.6)$$

wobei sich eine Normalisierung der Grundschriftweite $\hat{\mu}$ in Gleichung 3.6 mit

$$\sigma_m^2(n) = (1 - \hat{\mu})\sigma_m^2(n-1) + \hat{\mu}(f_{m-1}^2(n) + b_{m-1}^2(n-1)), \quad (3.7)$$

als vorteilhaft erwiesen hat. Diese entspricht einer Entkopplung der GAL-Schrittweite der Lattice-Stufen von den geschätzten Energien $\sigma_m^2(n)$ ihrer Eingangssignale $f_{m-1}(n)$ und $b_{m-1}(n-1)$. Zum Vermeiden einer Division durch Null bei fehlendem Eingangssignal wird zusätzlich eine kleine Konstante σ_{\min}^2 eingeführt. Des Weiteren wird eine Begrenzung der Koeffizienten zum Einhalten der Bedingung $|k_m| < 1$ vorgenommen. Grundsätzlich konvergiert der GAL-Algorithmus für stationäre Signale bis auf einen geringen Restfehler zur optimalen Lösung und liefert dadurch einen sehr hohen Prädiktionsgewinn. Er lässt sich außerdem mit einfachen Mitteln, wie in Abschnitt 3.2.2 gezeigt wird, hinsichtlich der Robustheit gegenüber Kanalübertragungsfehlern modifizieren. Jedoch wird sich im Laufe dieses Kapitels zeigen, dass trotz dieser Maßnahmen die bitexakte Synchronisierung zwischen Encoder und Decoder nicht sichergestellt ist und dass der GAL-Algorithmus in einem hohen Berechnungsaufwand resultiert. Daher wurden im Rahmen dieser Arbeit alternative Methoden und Strukturen zur Prädiktoradaption untersucht.

Der Burg-Algorithmus

Eine alternative Methode zur Adaption der Lattice-Koeffizienten ist der so genannte *Burg-Algorithmus*⁸ [Bur75], der die optimalen Lattice-Koeffizienten im Sinne der Minimierung der Fehlerenergien des Vorwärts- und Rückwärtszweiges für einen Block des Eingangssignales der Länge L sukzessive berechnet. Die Möglichkeit zu diesem Vorgehen resultiert aus der Entkopplung der Lattice-Stufen untereinander, so dass mit den Ergebnissen vorheriger Stufen der optimale Koeffizient einer darauf folgenden Stufe berechnet werden kann.

Mit der Initialisierung $f_0(n) = b_0(n) = x(n)$ für $n = 0, \dots, L-1$ berechnen sich die Koeffizienten für die Stufen $m = 1, \dots, p$ demnach über

$$k_m = \frac{2 \sum_{n=m}^{L-1} [f_{m-1}(n)b_{m-1}(n-1)]}{\sum_{n=m}^{L-1} [f_{m-1}^2(n) + b_{m-1}^2(n-1)]}, \quad (3.8)$$

wobei die Signale von Vorwärts- und Rückwärtsfehler mit jedem berechneten Koef-

⁸Detaillierte Herleitungen und weitere Diskussionen sind unter [Kei06] (S. 140 ff.) und den dort angegebenen Referenzen zu finden und werden zum Vermeiden unnötiger Wiederholungen an dieser Stelle nicht näher betrachtet.

fizienten der Stufe m gemäß

$$f_m(n) = f_{m-1}(n) - k_m b_{m-1}(n-1) \quad \text{für } n = m+1, \dots, L-1 \quad (3.9)$$

$$b_m(n) = b_{m-1}(n-1) - k_m f_{m-1}(n) \quad \text{für } n = m, \dots, L-1 \quad (3.10)$$

aktualisiert werden. Die Begrenzung des Laufindexes n auf Werte ab m bzw. $m+1$ erfolgt, da Werte außerhalb des betrachteten Blockes zu Null angenommen werden. Zum Vermeiden der Nulldivision wird vorab zudem der Nennerterm auf das Vorliegen eines Nullwertes überprüft und in diesem Fall der entsprechende Koeffizient auf Null gesetzt. Die im GAL-Algorithmus nötige Überprüfung der Koeffizienten kann, wie sich zeigen lässt, entfallen, da die Bedingung $|k_m| < 1$ durch die Formulierung des Burg-Algorithmus stets erfüllt ist.

Die Berechnung der Lattice-Koeffizienten nach dem Burg-Algorithmus kann, wie in [Kei06] (S. 162 ff.) dargestellt, auf mehrere Abtasttakte der ADPCM verteilt erfolgen. Entgegen dem Vorgehen bei den Berechnungen nach dem GAL-Algorithmus wird dabei jedoch nicht für jeden Abtasttakt des Eingangssignales und damit Adaptionsschritt der ADPCM ein vollständiger Satz an Lattice-Koeffizienten berechnet, der für die Berechnung der Prädiktion und Aktualisierung der Lattice-Zustände genutzt werden kann. Dies macht es nötig, die Lattice-Koeffizienten abschnittsweise konstant zu halten. Zudem ist es genau genommen, wie in [Kei06] (S. 149 ff.) diskutiert, erforderlich, beim Übergang zwischen zwei Koeffizientensätzen eine Aktualisierung der Lattice-Zustände durchzuführen. Die Verluste bei nicht erfolgender Aktualisierung sind für natürliche Audiosignale jedoch zu vernachlässigen. In der Praxis lassen sich zudem Algorithmen zum Erzeugen stetiger Koeffizientenverläufe zwischen aufeinanderfolgenden Blöcken finden.

Eine in der Literatur häufig diskutierte Eigenschaft des Burg-Algorithmus ist die Tendenz zu so genanntem *Line Splitting* unter ungünstigen SNRs und Phasenlagen des Eingangssignalblockes sowie eines im Vergleich zur Blocklänge hohen Prädiktordes (vgl. [Ale16] (S. 519 ff.)). Hierbei treten in dem geschätzten Spektrum des Eingangssignales für sinusartige Anteile anstelle einer Spalte zwei dicht benachbarte Maxima auf. Im Kontext der Audiocodierung und in der Anwendung für die Prädiktoradaption innerhalb der ADPCM konnte dieser Effekt zwar beobachtet, jedoch kein signifikanter Einfluss auf das Codierergebnis festgestellt werden.

Die Levinson-Durbin Rekursion

Eine weitere blockbasierte Methode zur Berechnung der Prädiktionskoeffizienten ist die so genannte *Levinson-Durbin Rekursion* [Lev47, Dur60]. Diese dient der Lösung des als *Yule-Walker-Gleichungen* bezeichneten Gleichungssystems

$$\sum_{i=1}^p a_i r_{xx}(k-i) = r_{xx}(k) \quad \text{für } k = 1, \dots, p, \quad (3.11)$$

das sich aus den partiellen Ableitungen zur Minimierung der Prädiktionsfehlerenergie eines linearen Prädiktors gemäß Gleichung 3.2 nach den Direktstruktur-Koeffizienten

ergibt. Hierbei ist

$$r_{xx}(k) = \frac{1}{L} \sum_{n=k}^{L-1} u(n)u(n-k) \quad (3.12)$$

die *Autokorrelationsfolge* (AKF) des, außerhalb des betrachteten Blockes zu Null angenommenen, Eingangssignales $u(n)$ des Algorithmus.

Der rekursive Algorithmus nach Levinson und Durbin zur Lösung dieses Gleichungssystems beinhaltet auch die Berechnung der Lattice-Koeffizienten. Zur Initialisierung wird der Eingangssignalblock der Länge L zunächst zum Vermeiden von Randeffekten mit einem geeigneten Fenster $w(n)$ zum gefensterten Signal

$$u(n) = w(n) \cdot x(n) \text{ für } n = 0, \dots, L-1 \quad (3.13)$$

verarbeitet. Hiermit wird gemäß Gleichung 3.12 für $k = 0, \dots, p$ die AKF bestimmt. Die eigentliche Koeffizientenberechnung⁹ erfolgt dann mit $\sigma_0 = r_{xx}(0)$ für $m = 1, \dots, p$ über

$$k_m = \frac{r_{xx}(m) - \sum_{i=1}^{m-1} a_i^{(m-1)} r_{xx}(m-i)}{\sigma_{m-1}}, \quad (3.14)$$

wobei mit jeder Iteration $a_m^{(m)} = k_m$ gesetzt und jeweils für $i = 1, \dots, m-1$ über

$$a_i^{(m)} = a_i^{(m-1)} - k_m a_{m-i}^{(m-1)} \quad (3.15)$$

und

$$\sigma_m = (1 - k_m^2) \sigma_{m-1} \quad (3.16)$$

ein Satz an Direktstruktur-Koeffizienten des Prädiktors m -ter Ordnung und die Leistung des aus ihm resultierenden Prädiktionsfehlers bestimmt wird. Dies erfolgt unter Ausnutzung der Unabhängigkeit einer Lattice-Stufe von den vor ihr befindlichen, woraus resultiert, dass eine Erhöhung des Prädiktorgrades nicht die Neuberechnung der Lattice-Koeffizienten notwendig macht.

Genau wie beim Burg-Algorithmus sind bei der Verwendung der Levinson-Durbin Rekursion in der ADPCM im Normalfall die Koeffizienten für eine Blocklänge konstant zu halten. Die Berechnung lässt sich dabei auf vergleichbare Weise inklusive der Bestimmung der AKF auf mehrere Takte der ADPCM-Codierung verteilen. Die Aktualisierung der Lattice-Zustände kann ebenso in der Praxis entfallen und es lassen sich wiederum Methoden zum Erzeugen stetiger Koeffizientenverläufe ermitteln. In der Literatur finden sich Hinweise, dass es bei der Spektralschätzung mit der Levinson-Durbin Rekursion zu einer numerischen Instabilität kommen kann und es werden eine Metrik zum Erfassen und Maßnahmen zum Umgang mit dieser [PKVA95] vorgestellt. Im Kontext der Prädiktoradaption sowie bei der, im übernächsten Abschnitt 3.2.3 präsentierten, Adaption auf das Fehlersignal wurde dieser Effekt jedoch bislang nicht beobachtet bzw. detektiert.

⁹Der hochgestellte Index (m) der a_i bezeichnet hierbei keine Potenzen, sondern markiert, dass es sich um die Koeffizienten eines Prädiktors m -ter Ordnung handelt.

Aufwandsvergleich der Verfahren

Zur Bewertung und Analyse der drei vorgestellten Verfahren zur Prädiktoradaption in Bezug auf den Berechnungsaufwand im vorliegenden Anwendungsfall wurde dieser, anhand von deren Basisformulierungen, abgeschätzt. Tabelle 3.1 enthält demnach Ergebnisse der bei der Berechnung nötigen Operationen. Hierbei wurden für eine praxisnahe Blocklänge $L = 96$ und die Prädiktorordnung von $p = 32$ die Anzahlen an Additionen, Multiplikation und Divisionen ermittelt. Potenzierungen mit dem Exponent zwei, z.B. zur Berechnung der Fehlerenergien, werden hierbei nicht einzeln aufgeführt, sondern als Multiplikation des jeweiligen Wertes mit sich selbst gerechnet. Um den Vergleich mit den blockbasierten Verfahren zu ermöglichen, wurde für den GAL-Algorithmus der Aufwand zum Berechnen der Koeffizienten von p Abtasttakten gegenübergestellt. Natürlich ist dieser Vergleich nur bedingt fair, da der GAL-Algorithmus hierbei für jeden Abtasttakt alle Koeffizienten berechnet. Dennoch reflektiert das hier gewählte Vorgehen die, bei der Anwendung für die gradientenbasierte Prädiktoradaption bei gleicher Prädiktorordnung, benötigten Berechnungsressourcen.

Tabelle 3.1: Vergleich des Berechnungsaufwandes der Lattice-Koeffizientenadoptionsalgorithmen für eine Blocklänge von $L = 96$ und Prädiktorordnung von $p = 32$.

Verfahren	Add.	Mult.	Div.
Burg-Algorithmus	10144	12720	32
Levinson-Durbin R.	3728	3792	65
GAL-Algorithmus	5120	7168	1024

Der GAL-Algorithmus sticht demnach aufgrund der erhöhten Zahl an benötigten Divisionen hervor. Diese kann bei den zum Teil verwendeten hohen Prädiktorordnungen (vgl. [Hol09]) die Berechnungsressourcen einiger Anwendungen übersteigen. Die Zahl der nötigen Divisionen kann im Fall der Levinson-Durbin Rekursion sogar auf einen zum Burg-Algorithmus vergleichbaren Wert reduziert werden, wenn die Division in der Berechnung der AKF durch eine Multiplikation mit einem vorberechneten Wert ersetzt wird. Damit lässt sich die Prädiktoradaption mit der Levinson-Durbin Rekursion am effizientesten durchführen. Gegebenenfalls lassen sich darüber hinaus weitere geringe Reduktionen durch zusätzliche Vorberechnungen im Rahmen systemspezifischer Implementierungen erzielen.

In [Kei06] (S. 155/156) erfolgt zusätzlich ein Vergleich der Prädiktionsgewinne von block- und gradientenbasierter Prädiktoradaption. Hierbei ergeben sich für natürliche Audiosignale kaum signifikante Unterschiede, die den erhöhten Rechenaufwand des GAL-Algorithmus rechtfertigen würden. Da mit den im Folgenden präsentierten Modifikationen zusätzlich eine weitreichende Abhängigkeit von der Parametrisierung und diversen Rückkopplungseinflüssen entsteht, findet der Nachweis einer Eignung

der Verfahren zur blockbasierten Prädiktoradaption in dieser Arbeit erst im Rahmen der finalen Evaluation im Gesamtsystem statt.

3.2.2 Fehlerrobuste Rückwärtsprädiktion

Das Vorgehen zur rückwärtsadaptiven Rückwärtsprädiktion in der ADPCM-Struktur nach Abbildung 3.1 hat den bereits erwähnten Vorteil, dass keine Zusatzinformationen zwischen Encoder und Decoder übertragen werden müssen. Allerdings bringt es das Problem mit sich, dass bei Störungen auf dem Kanal durch eine fehlerhafte Rekonstruktion des normalisierten Prädiktionsfehlersignales ein Fehler in den Adaptionsroutinen des Decoders entsteht. Dieser Fehler verlässt das System ohne weitere Maßnahmen allerdings nicht wieder und kann zu einer vollständigen Divergenz zwischen Encoder und Decoder führen. Der Einfluss des über das Rekonstruktionssignal $\tilde{x}(n)$ zurück gekoppelten Fehlers wiegt dabei doppelt, da es gleichzeitig Eingangssignal des Prädiktors ist und zu dessen Adaption genutzt wird.

Im Fall der Lattice-Koeffizientenadaption über den GAL-Algorithmus hat es sich daher als wirksam erwiesen, die in die Rückkopplung einfließenden Signalfäde des Lattice-Filters mit so genannten *Leakage*¹⁰-Faktoren $\alpha, \beta < 1$ zu versehen (vgl. [Die83, PO11, SHZ13a]). Hierdurch soll erreicht werden, dass ein eingekoppelter Fehler durch fortwährende Multiplikation mit einem Faktor kleiner Eins bei der Rückkopplung über der Zeit aus dem Decoder entweicht. Abbildung 3.5 zeigt den Flussgraphen des, um die Leakage-Faktoren erweiterten, Lattice-Filters inklusive der Signalabzweige für die Berechnung des GAL-Algorithmus gemäß [SHZ13a].

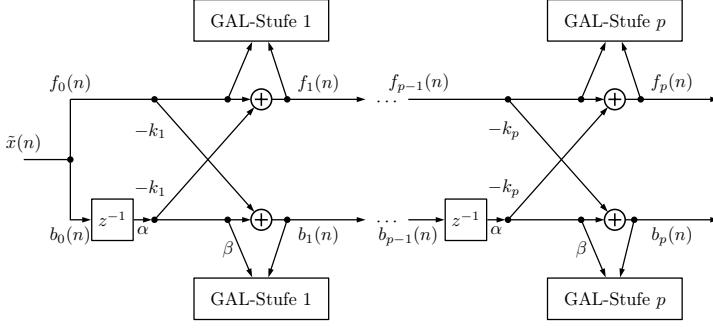


Abbildung 3.5: Lattice-Filter mit Leakage-Faktoren α und β (gemäß [SHZ13a]).

Hierbei werden die Signale des unteren Rückwärtsprädiktionsfehlerzweiges nach der Einheitsverzögerung mit α multipliziert. Zusätzlich werden die so gewichteten Signale bei der Verwendung in den einzelnen GAL-Stufen mit einem weiteren Faktor β

¹⁰Englisch für das Entweichen, der Schwund oder das Auslaufen.

gewichtet, so dass sich eine modifizierte Version des GAL-Algorithmus zur Berechnung der Lattice-Koeffizienten über

$$k_m(n+1) = k_m(n) + \frac{\tilde{\mu}}{\sigma_m^2(n) + \sigma_{\min}^2} \cdot (f_m(n) \cdot \alpha \beta b_{m-1}(n-1) + b_m(n) \cdot f_{m-1}(n)) \quad (3.17)$$

ergibt. Da die Prädiktionsberechnung auch auf Basis des Rückwärtsprädiktionsfehlers $b_m(n)$ erfolgt, wird diese dementsprechend zu

$$p(n) = \sum_{m=1}^p k_m \alpha b_{m-1}(n-1). \quad (3.18)$$

modifiziert.

Abbildung 3.6 zeigt die Verläufe der Differenzen der ersten fünf Lattice-Koeffizienten bei Codierung eines Sinussignales der Amplitude -3 dB mit einer Frequenz von 1 kHz und einem Breitband 3 Bit ADPCM-Codec, der über einen mit Leakage-Faktoren modifizierten GAL-Algorithmus adaptiert wird. Die Einkopplung eines Fehlers erfolgte durch Invertieren des *Most Significant Bit* (MSB), also des höchstwertigsten Bits eines übertragenen Quantisierungsindexes nach ca. einer Sekunde vor der Decodierung im Decoder.

Es ist zu erkennen, wie durch die eingeführte Abweichung im rekonstruierten Prädiktionsfehler des Decoders und damit auch in seinem Rekonstruktionssignal eine Differenz zu den Lattice-Koeffizienten des Encoders entsteht. Diese wird durch die verwendeten Leakage-Faktoren jedoch stetig kleiner und konvergiert nach kurzer Zeit wieder gegen Null. Für die Differenz zwischen den Rekonstruktionssignalen von Encoder und Decoder ergibt sich ein ähnliches Verhalten, wobei jedoch eine Oszillation um den Nullwert stattfindet. Trotz des weiterhin anliegenden Eingangssignales bleiben die Differenzen der Signale von Encoder und Decoder bei diesen geringen Amplituden solange kein erneuter Fehler in das System einkoppelt.

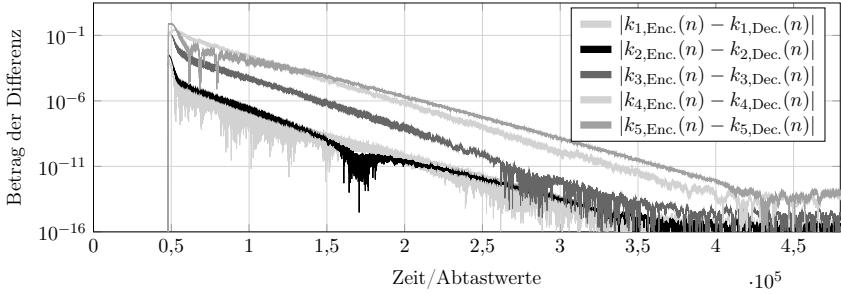


Abbildung 3.6: Betrag der Differenz der Lattice-Koeffizientenverläufe zwischen Encoder und Decoder einer Breitband-ADPCM bei Adaption mit dem fehlerrobusten GAL-Algorithmus und einem Bitfehler im MSB nach ca. 1 Sekunde.

Mit dem Einführen der Leakage-Faktoren lässt sich somit zwar nur eine langsame, aber dennoch sichere Stabilisierung des Decoders nach auftretenden Übertragungsfehlern erreichen. So werden in [SHZ13a] Fehlerwerte für die Differenz der Rekonstruktionssignale zwischen Encoder und Decoder bei sinusartiger Anregung von 10^{-4} nach 1,2 s und 10^{-12} nach weiteren 6,5 s angegeben. Allerdings hat die Modifikation des GAL-Algorithmus über die Leakage-Faktoren einen negativen Einfluss auf den Prädiktionsgewinn, was sich besonders für geringe Bitraten in einer verschlechterten Audioqualität äußert (Vgl. [SHZ13a]). Zusätzlich bleibt ohne weitere Maßnahmen unter Umständen ein Restfehler im GAL-Algorithmus. Dieser kann sich unter ungünstigen Bedingungen auch ohne das Auftreten weiterer Übertragungsfehler über der Zeit akkumulieren und damit einen negativen Einfluss auf die Audioqualität haben, ohne jedoch zu einer Divergenz von Encoder und Decoder zu führen. Im Einzelfall ist daher abzuwägen, ob dies zu akzeptieren ist oder durch Maßnahmen wie z.B. das koordinierte Zurücksetzen von Koeffizienten und Speichern des Lattice-Filters in der GAL-Adaption eine bitexakte Synchronisation¹¹ erzwungen werden muss. Diese hat jedoch in der Regel einen weiteren negativen Einfluss auf die Audioqualität oder macht eine zusätzliche Signalisierung notwendig.

Neben dem bereits diskutierten hohen Rechenaufwand des GAL-Algorithmus führen die zuvor genannten Gründe zu Untersuchungen möglicher Alternativen für eine fehlerrobuste Prädiktoradaption innerhalb dieser Arbeit.

3.2.3 Blockbasierte Adaption auf das Fehlersignal

Die Adaption der Prädiktorstruktur mit dem, um Leakage-Faktoren erweiterten, GAL-Algorithmus erlaubt es, die Divergenz zwischen Encoder und Decoder zu vermeiden und das Vorgehen ist prinzipiell auch auf die Algorithmen zur blockbasierten Prädiktoradaption übertragbar. Es wäre jedoch darüber hinaus wünschenswert, die Synchronisierung der Prädiktoradaption von Encoder und Decoder nach Übertragungsfehlern algorithmisch zu garantieren.

Daher wurde im Rahmen dieser Arbeit eine Alternative zur rückwärtsadaptiven Prädiktion auf Basis des Rekonstruktionssignales untersucht, die die Synchronität inhärent sicherstellt. Hierbei wird von einer Heuristik Gebrauch gemacht, die die Eigenschaften des Prädiktionsfehlersignales realer ADPCM-Systeme bei Codierung natürlicher Audiosignale betrifft. Abbildung 3.7 zeigt einen Vergleich des Eingangs- und Prädiktionsfehlersignales einer Breitband-ADPCM-Codierung für das SQAM-Signal 54 Male Speech (German). Es ist eine deutliche Ähnlichkeit der Wellenformen bis auf eine Skalierung der Amplituden zu erkennen und Experimente zeigen, dass sich das Prädiktionsfehlersignal trotz der zusätzlich erfolgenden adaptiven Quantisierung für die Prädiktoradaption nutzen lässt.

Demzufolge kann eine Modifikation der Prädiktoradaption vorgenommen werden, die die Rückkopplung des Rekonstruktionssignales über den Adoptionsalgorithmus vermeidet. Abbildung 3.8 zeigt einen Flussgraphen der dementsprechend modifizierten

¹¹Die im Folgenden auch als „Algorithmische Stabilität“ bezeichnet wird.

ADPCM mit Adaption auf das Fehlersignal. Hierbei wird anstelle des Rekonstruktionssignales $\tilde{x}(n)$ der denormalisierte Prädiktionsfehler $\tilde{e}(n)$ als Eingangssignal der blockbasierten Prädiktoradaption verwendet.

Hierbei erfolgt sowohl die Steuerung der adaptiven Quantisierung, als auch die Berechnung der Lattice-Koeffizienten mit dem Burg-Algorithmus oder der Levinson-Durbin Rekursion auf Basis des denormalisierten Prädiktionsfehlersignales $\tilde{e}(n)$. Da sich die fehlerrobuste adaptive Quantisierung mittels Hüllkurvenschätzung, wie in Abschnitt 3.3 gezeigt, so realisieren lässt, dass die bitexakte Gleichheit der Normalisierungssignale in Encoder und Decoder nach auftretenden Übertragungsfehlern sichergestellt ist, ist die Übereinstimmung der Grundlage zur Prädiktoradaption somit gewährleistet.

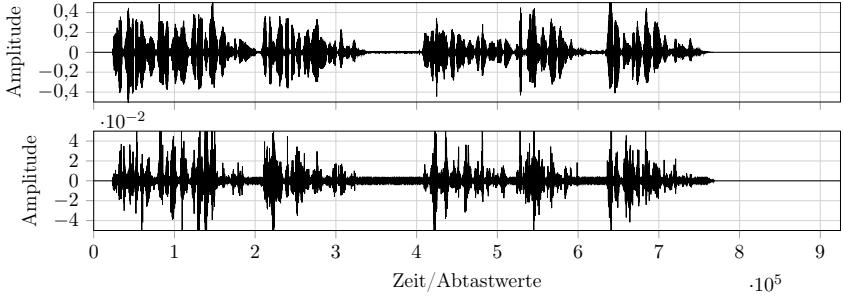


Abbildung 3.7: Vergleich des Eingangs- (oben) und Prädiktionsfehlersignales (unten) einer Breitband-ADPCM-Codierung des SQAM-Signales 54 Male Speech (German).

Sobald ein zur Koeffizientenberechnung genutzter Block des denormalisierten Prädiktionsfehlersignales $\tilde{e}(n)$ vollständig richtig rekonstruiert wurde, liegen damit auch wieder zum Encoder identische Lattice-Koeffizienten vor. Der über das Rekonstruktionssignal $\tilde{x}(n)$ zurück gekoppelte Fehler entweicht dadurch automatisch aus dem Prädiktionsfilter des Decoders. Dies ist bedingt durch die Tatsache, dass mit der Adaption auf das Fehlersignal die einzige Rückkopplung des Systems in der Berechnung der Prädiktion über Gleichung 3.5 besteht und diese durch die Eigenschaft $|k_m| < 1$ der Lattice-Koeffizienten ein natürliches Entweichen des Fehlers beinhaltet. Somit kann auch auf die zusätzliche Dämpfung mittels eines Leakage-Faktors verzichtet werden und eine Akkumulation des Restfehlers wird verhindert.

Ein vergleichbares Vorgehen wurde in [SHZ13a] für die Adaption mit dem GAL-Algorithmus gewählt, führt dort jedoch zu einer deutlich vermindernten Audioqualität. Dies liegt vermutlich an der mit 3 Bit unzureichenden Auflösung des quantisierten Prädiktionsfehlersignales bei Breitband-Codierung. Mit dem in dieser Arbeit erfolgenden Einsatz in einem Subband-Codec und den dabei verwendeten höheren Auflösungen unterer Bänder lässt sich jedoch mit der blockbasierten Adaption auf das Fehlersignal eine für den vorliegenden Anwendungsfall ausreichende Audioqualität erzielen, wie Kapitel 5 zeigen wird.

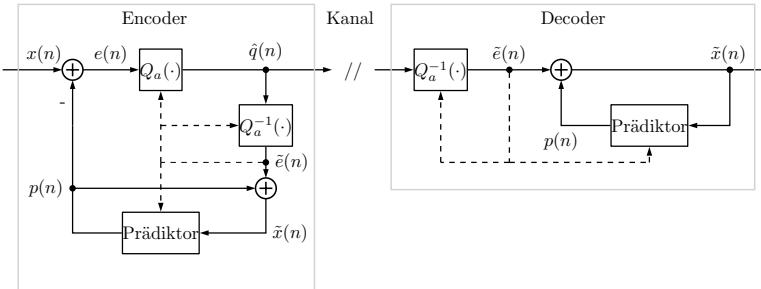


Abbildung 3.8: Adaption auf das Fehlersignal in der ADPCM.

3.2.4 Einführung eines statischen Prädiktors

Eine weitere im Rahmen dieser Arbeit zur Anwendung kommende Heuristik betrifft die Ergänzung der adaptiven Prädiktion um einen statischen Anteil. Dies hat sich im Zuge der Untersuchungen zu [PO11] als vorteilhaft erwiesen, da es sich positiv auf den Gesamtprädiktionsgewinn auswirkt und es zum Teil ein schnelleres Synchronisieren von Encoder und Decoder nach Kanalübertragungsfehlern erlaubt. Abbildung 3.9 zeigt einen um einen statischen Anteil erweiterten Prädiktor im Encoder der fehlerrobusten ADPCM.

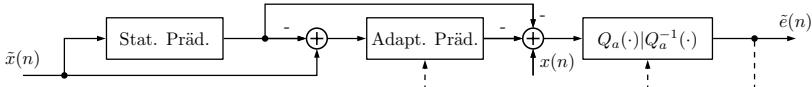


Abbildung 3.9: Um statischen Anteil ergänzter Prädiktor im Encoder der ADPCM.

Eingangssignal des adaptiven Prädiktors ist hierbei der Fehler der statischen Prädiktion. Diese wird über die Filterung mit einem statischen Direktstruktur-Filter vierter Ordnung ermittelt. In verschiedenen Experimenten hat sich u.a. die Optimierung der Filterkoeffizienten auf die Verteilungsdichte von Rosa Rauschen als zweckdienlich erwiesen. Abbildung 3.10 zeigt beispielhaft die Prädiktionsfehlerfilterübertragungsfunktion eines solchen Filters.

Im Fall der Subband-Codierung kommt der statische Prädiktor nur im Tiefpassanteil zur Anwendung. Die Koeffizienten sind optional zusätzlich Teil der globalen Parameteroptimierung (vgl. Kapitel 4).

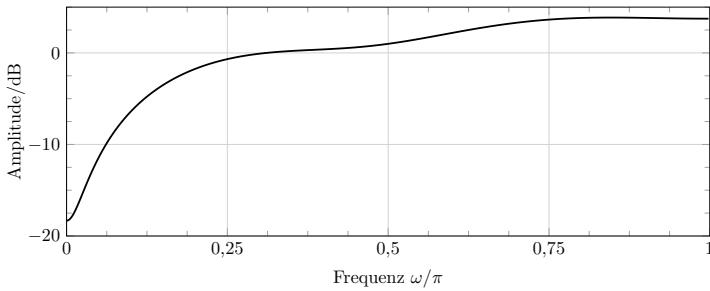


Abbildung 3.10: Betragsverlauf der Prädiktionsfehlerübertragungsfunktion eines auf Rosa Rauschen optimierten statischen Prädiktors.

3.3 Fehlerrobuste adaptive Optimalquantisierung

Mit der adaptiven Prädiktion und der Berechnung des Prädiktionsfehlers in der ADPCM erfolgt noch keine Reduktion der Datenrate. Diese kommt, wie bereits erwähnt, erst dadurch zustande, dass der Prädiktionsfehler einer adaptiven Quantisierung unterzogen wird. Diese kann aufgrund dessen geringerer Dynamik mit einer, im Verhältnis zum Eingangssignal, deutlich verminderter Auflösung geschehen. Eine Quantisierung bezeichnet den Prozess des Abbildens eines zunächst als, im Rahmen der zur Verfügung stehenden Rechengenauigkeit, kontinuierlich angenommenen Eingangssignalwertes auf einen begrenzten Vorrat an Codewörtern und die damit verbundenen Repräsentativwerte. Damit einher geht ein unweigerlich entstehender Fehler, der dem Signal unwiderruflich anhaftet.

In diesem Abschnitt wird nach einer kurzen Einführung in die Grundlagen zur Quantisierung in Unterabschnitt 3.3.1 die, in dieser Arbeit verwendete, Methode zum Entwurf der Quantisierer Codebücher vorgestellt und diskutiert. Daran anschließend wird in Unterabschnitt 3.3.2 eine Algorithmitik zur adaptiven Normalisierung des Prädiktionsfehlersignales dargestellt und diese um Maßnahmen zum Erreichen einer Robustheit gegenüber Kanalübertragungsfehlern ergänzt. Der vorletzte Unterabschnitt 3.3.3 beinhaltet eine Erweiterung des Subband-Codecs um eine dynamische Bit-Allokation. Der, in der Überschrift dieses Abschnittes, verwendete Begriff der *Optimalquantisierung* reflektiert das vorliegende Bestreben durch die genannten Maßnahmen eine nahezu optimale Anpassung an die Signaleigenschaften bei der Quantisierung der Subband-Prädiktionsfehlersignale zu erreichen.

Grundsätzlich besteht die Möglichkeit, die Quantisierung eines Signales nicht nur für einzelne Signalwerte durchzuführen, was als *skalare Quantisierung* bezeichnet wird. So können auch mehrere zeitlich aufeinanderfolgende Signalwerte zusammengefasst und einer so genannten *Vektorquantisierung* unterzogen werden. Da sich dieses Vorgehen im Grunde mit der Sample-basierten Verarbeitung und Adaption in der ADPCM widerspricht, wird im Folgenden zunächst nur die skalare Quantisierung betrachtet. In Abschnitt 3.3.4 wird aber dennoch die Integration eines Vektorquan-

tisierungsverfahrens in einen fehlerrobussten Breitband-ADPCM-Codec vorgestellt und Ergebnisse einer Evaluation des Einflusses auf die Audioqualität und das Verhalten bei Übertragungsfehlern präsentiert.

Abbildung 3.11 zeigt exemplarisch die Kennlinie eines symmetrischen gleichförmigen skalaren 3 Bit Quantisierers. Der Bereich zwischen dem Maximal- und Minimalwert $d_{\max} = 1$ und $d_{\min} = -1$ des Eingangssignales wird in K_C Segmente der Breite Δ_k unterteilt. Zwischen den Segmenten liegen die $K_C - 1$ Entscheidungsschwellen d_k . Jeder Wert, der zwischen d_k und d_{k-1} liegt, wird auf das Codewort C_k abgebildet (durch Pfeile angedeutet). Dieses entspricht einem Repräsentativwert $d_k - (d_k - d_{k-1})/2$. Bei der Abbildung auf den so ermittelten Repräsentativwert wird unausweichlich ein Fehler gemacht. Dieser erreicht sein Maximum $\Delta_k/2$ am Übergang zwischen zwei Repräsentativwerten, verringert sich jedoch in der Mitte zwischen den Entscheidungsschwellen sogar zu Null.

Die AmplitudendichteVerteilung des Quantisierungsfehlers und damit dessen Leistung ist somit direkt von der Verteilungsdichte des Eingangssignales abhängig. Das Verhältnis zwischen Signal- und Fehlerenergie, das so genannte *Signal to Noise Ratio* (SNR), verringert sich dabei für kleiner werdende Signalamplituden. Da natürliche Audiosignale, wie z.B. Sprache, und auch der zu quantisierende Prädiktionsfehler der ADPCM eine normalverteilte Amplitudendichte und somit verhältnismäßig mehr Signalwerte bei kleinen Amplituden aufweisen, stellt die gleichförmige Quantisierung nicht das Optimum hinsichtlich des erreichbaren maximalen SNRs für eine gegebene Signalfolge dar.

Weitere Aspekte, die bei der Auslegung eines Quantisierers und der damit einhergehenden Festlegung von Entscheidungsschwellen Berücksichtigung finden müssen, sind zum einen die Grenzen d_{\min} und d_{\max} , die nicht zwingend bei -1 und 1 liegen müssen und ab denen eine Übersteuerung des Quantisierers stattfindet. Zum anderen ist die Entscheidung zwischen einem symmetrischen oder unsymmetrischen Quantisierer zu treffen. Letzterer kann aus der symmetrischen Kennlinie (Abbildung 3.11) durch deren Verschiebung um $\pm \Delta_k/2$ jeweils in x - und y -Richtung abgeleitet werden und enthält damit auch ein Codewort, das auf den Repräsentativwert 0 abbildet. Durch die Wahl eines unsymmetrischen Quantisierers kann das so genannte Ruhegeräusch minimiert und das Einleiten von „Nullen“ in die nachgeschalteten ADPCM-Strukturen ohne zusätzliche Signalisierung ermöglicht werden.

Es hat sich zudem, wie schon in den Grundlagen zur ADPCM dargestellt, als vorteilhaft erwiesen, das Eingangssignal des Quantisierers vorab adaptiv zu normalisieren, um dessen Varianz weitestgehend konstant zu halten und damit den Quantisierer vollständig aussteuern zu können. Die dadurch zum Teil auftretenden Über- bzw. Untersteuerungen des Quantisierers bei instationären Signalanteilen werden dabei zugunsten der besseren mittleren Aussteuerung bewusst in Kauf genommen. Zusammen mit einer globalen Optimierung der, für die Signalnormalisierung verwendeten, Parameter (vgl. Kapitel 4) kann so ein sinnvoller Arbeitspunkt sowie eine an die Signalstatistik angepasste Lage der Entscheidungsschwellen gefunden werden.

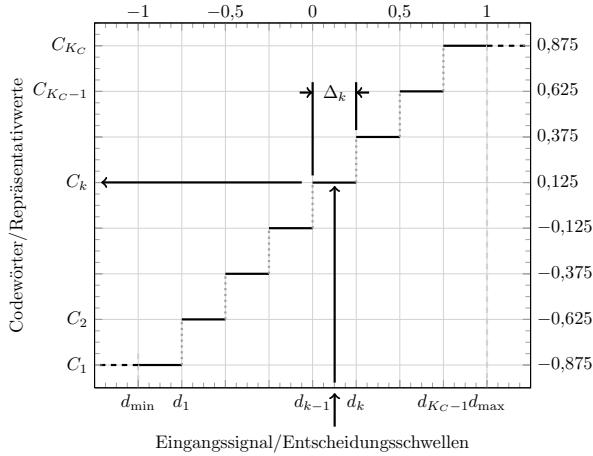


Abbildung 3.11: Kennlinie eines symmetrischen gleichförmigen Quantisierers (Repräsentativwerte: schwarze Linien, Entscheidungsschwellen: gepunktete graue Linien).

3.3.1 Erzeugung der Quantisierungskennlinien

Auch wenn mit der adaptiven Normalisierung des Eingangssignales eine gleichmäßige Aussteuerung des Quantisierers erreicht werden kann, ist es weiterhin von Vorteil die Lage der Entscheidungsschwellen und damit die Breite der Quantisierungsintervalle ungleichmäßig zu gestalten. Dies erfolgt dabei zumeist mit dem Ziel, für einen gewissen Amplitudenbereich des Eingangssignales das SNR konstant zu halten und einen Gewinn im SNR für kleine Signalamplituden zu erreichen.

Generell wäre es erstrebenswert, die Quantisiererkennlinie direkt aus der Verteilungsdichte des Prädiktionsfehlersignals für eine gegebene Signalfolge zu optimieren. Dies kann z.B. über den so genannten *Lloyds Algorithmus* [Llo82] erfolgen, der iterativ den mittleren quadratischen Fehler eines Quantisierers gegebener Wortbreite bei einer vorgegebenen Amplitudenverteilungsdichte über die näherungsweise Lösung eines nichtlinearen Gleichungssystems minimiert. Bei Benutzung des *Lloyds Algorithmus* in einer „Open Loop“-Optimierung würden allerdings nicht die, durch Rückkopplung des Rekonstruktionssignales innerhalb der ADPCM entstehenden, Einflüsse der Quantisierung erfasst und bei der Optimierung berücksichtigt. Zusätzlich ist der Lloyds Algorithmus verhältnismäßig rechenaufwändig, was eine Verwendung innerhalb der Parameteroptimierung des Codecs erschwert.

Eine andere Möglichkeit zum Erzeugen der Quantisierungskennlinien ist die in [Hol09] durchgeführte direkte Optimierung der Codebucheinträge des Quantisierers im Rahmen der globalen Parameteroptimierung. Dieses Vorgehen ist bei den in [Hol09] mit 3-4 Bit Wortbreite recht kleinen Quantisierungstabellen, die zudem symmetrisch gehalten wurden, unproblematisch, führt im Endeffekt jedoch auch nur auf eine zu er-

wartende nichtlineare Kennlinie. Zusätzlich werden für den hier vorliegenden Anwendungsfall einer Subband-Codierung durch die ungleichmäßige Verteilung der „Bits pro Band“ deutlich größere Codebücher benötigt, was den Parameterraum signifikant vergrößern würde.

Daher hat es sich im Rahmen dieser Arbeit als zweckdienlich erwiesen, den Codebuchentwurf über eine parametrierbare nichtlineare Kennlinie, die zusätzlich linear skaliert wird, zu realisieren.

Eine in der Sprachkommunikation zur Anwendung kommende und von der ITU im Standard G.711 [ITU93] festgehaltene Berechnungsvorschrift ist die so genannte *A-Law Kennlinie* [GG10]. Diese über

$$y = \begin{cases} \frac{A}{1 + \ln(A)} \cdot \operatorname{sgn}(x) \cdot |x| & \text{für } 0 \leq |x| \leq \frac{1}{A} \\ \frac{1 + \ln(A|x|)}{1 + \ln(A)} \cdot \operatorname{sgn}(x) & \text{für } \frac{1}{A} < |x| \leq 1 \end{cases} \quad (3.19)$$

definierte¹² Kennlinie setzt sich aus einem linearen Ast (oberer Teil) und einen nichtlinearen Ast (unterer Teil) zusammen und wird über den einzigen namensgebenden Parameter $A \geq 1$ parametrisiert. Abbildung 3.12 zeigt den Kennlinienverlauf für verschiedene Werte des A -Parameters.

In Abbildung 3.13 sind die entsprechenden SNR-Verläufe bei Quantisierung nach dem *A-Law* bei einer 12 Bit Quantisierung dargestellt. Es ist deutlich zu erkennen, wie sich über die Einstellung des A -Parameters die Abwägung zwischen der Größe des Bereiches konstanten SNRs sowie dem Gewinn im SNR für kleine Signalamplituden und dem maximal erzielbaren SNR für große Signalamplituden treffen lässt. Zusammen mit der in Kapitel 4 vorgestellten globalen Parameteroptimierung lassen sich damit über das Einstellen von jeweils nur zwei Parametern nahezu optimale Quantisiererkennlinien für die Subbänder erzeugen.

In der Praxis kann die Quantisierung nach dem *A-Law* auch über eine als *Compadding* bezeichnete Verzerrung des Eingangssignales mittels der *A-Kennlinie* und einer anschließenden gleichförmigen Quantisierung realisiert werden. In der vorliegenden Arbeit werden jedoch die Codebücher direkt über eine äquidistante Abtastung der *A-Law* Kennlinie für den jeweiligen A -Parameter erzeugt.

3.3.2 Rückwärtsadaptive Prädiktionsfehlernormalisierung

Die adaptive Normalisierung des Prädiktionsfehlers kann auf verschiedene Weisen realisiert werden. Zum einen kann durch so genanntes *Block Compadding* (vgl. [SHZ13b]), bei dem für einen Block definierter Länge der Maximalwert bestimmt und der gesamte Block mit diesem normalisiert wird, sogar ein vollständiges Vermeiden von Übersteuerungen erreicht werden. Da in dem vorliegenden Anwendungsfall

¹²Wobei $\operatorname{sgn}(x)$ die Signumfunktion ist.

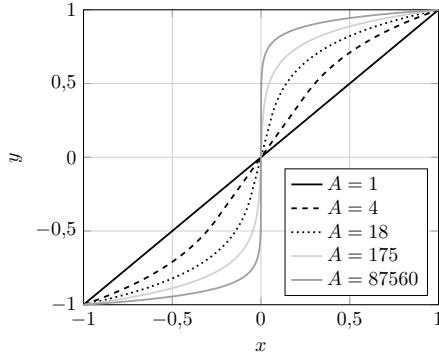


Abbildung 3.12: A-Law Kennlinien für verschiedene Werte des A-Law Parameters.

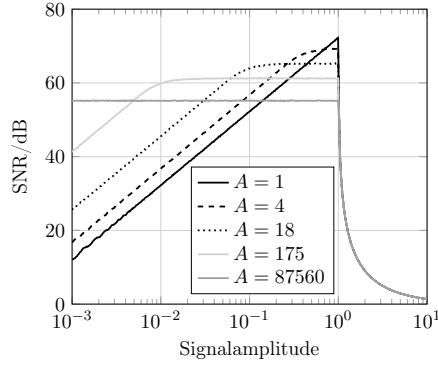


Abbildung 3.13: SNR-Verläufe in Abhängigkeit der Signalamplitude für verschiedene Werte des A-Law Parameters bei Quantisierung mit 12 Bit.

jedoch keine zusätzliche Latenz durch Blockbildung sowie eine weitere Erhöhung der Datenrate für die Übertragung des Normalisierungsfaktors akzeptiert werden kann, scheidet diese Möglichkeit aus.

Als Alternative haben sich in der Vergangenheit unter anderem auch die in [Kei06] und [Hol09] verwendeten Methoden als zielführend erwiesen, bei denen rückwärtsadaptiv die Einhüllende des Eingangssignales geschätzt und dieses damit normalisiert wird. Hierbei erfolgt in beiden Fällen die Schätzung der Einhüllenden des Prädiktionsfehlers über die Verarbeitung mit einem rekursiven Filter erster Ordnung, dessen Zeitkonstanten adaptiv geschaltet werden.

In [Kei06] wird bei der Prädiktionsfehlernormalisierung der Fokus auf das Vermeiden ungewollter Übersteuerungen des Quantisierers gelegt und es wird durch zusätzliche Erweiterungen des Verfahrens versucht, deren Auftreten weitestgehend zu

unterbinden. Im Gegensatz hierzu konnte in [Hol09] gezeigt werden, dass auch ohne diese Maßnahmen und trotz in Kauf genommener kurzzeitiger Übersteuerungen des Quantisierers eine sehr gute Audioqualität erreicht werden kann. Hierbei spielt auch die Berücksichtigung der Maximalamplituden des normalisierten Signales in der Auslegung des Quantisierers eine wichtige Rolle. Daher wird dieses Verfahren inklusive seiner Notation in der vorliegenden Arbeit weitestgehend übernommen und um Maßnahmen erweitert, die eine Robustheit gegenüber Kanalübertragungsfehlern ermöglichen.

Abbildung 3.14¹³ zeigt im rechten Teil einen Flussgraphen des Quantisierers mit rückwärtsadaptiver Normalisierung.

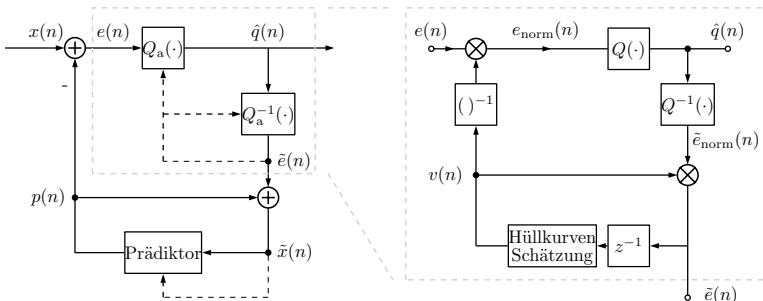


Abbildung 3.14: Aufbau des adaptiven Quantisierers mit rückwärtsadaptiver Normalisierung (rechts) zur Verwendung in der ADPCM (links).

Das Prädiktionsfehlersignal $e(n)$ wird demnach am Eingang des adaptiven Quantisierers mit dem Inversen des Normalisierungssignales $v(n)$ multipliziert, woraus das Signal des normalisierten Prädiktionsfehlers $e_{norm}(n)$ resultiert. Dieses wird durch den Block $Q(\cdot)$ einer skalaren Quantisierung unterzogen, was die Abbildung auf eine Folge von Quantisierungsindizes $\hat{q}(n)$ bedeutet. Zur rückwärtsadaptiven Berechnung des Normalisierungssignales wird aus dem rekonstruierten normalisierten Prädiktionsfehler $\tilde{e}_{norm}(n)$ und $v(n)$ der rekonstruierte Prädiktionsfehler $\tilde{e}(n)$ berechnet, der die Grundlage für die Hüllkurvenschätzung des nächsten Zeitschrittes ist. Durch dieses Vorgehen bei der Hüllkurvenschätzung ist sichergestellt, dass bei fehlerfreier Übertragung zwischen Encoder und Decoder exakt das gleiche Normalisierungssignal berechnet und damit das Prädiktionsfehlersignal rekonstruiert wird.

Die Basisformulierung der Berechnung des Normalisierungssignals $v(n)$ erfolgt als Schätzung der Momentanleistung des denormalisierten rekonstruierten Prädiktionsfehlersignals $\tilde{e}(n)$ mit einer exponentiell gewichteten Mittelwertbildung über

$$v^2(n) = \lambda(n) \cdot \tilde{e}^2(n-1) + (1 - \lambda(n)) \cdot v^2(n-1), \quad (3.20)$$

¹³Die folgenden Darstellungen enthalten wiederum einen allgemeinen Signalindex n .

wobei die Gewichtung über

$$\lambda(n) = \begin{cases} \lambda_{\text{AT}} & \text{für } \tilde{e}^2(n-1) > v^2(n-1) \\ \lambda_{\text{RT}} & \text{sonst.} \end{cases} \quad (3.21)$$

zeitlich adaptiert wird. Die Indizes AT und RT stehen dabei für die aus der *Dynamikkompression* bekannten Zeitkonstanten für den Ansprech- und Rücklauf-Fall (engl.: Attack-Time, Release-Time). Um ein ungewolltes „Wegdriften“ bei Signallücken mit Nullwerten und eine dabei sonst auftretende Division durch Null zu vermeiden, wird das Signal $v(n)$ nach unten zudem auf v_{\min} begrenzt.

Abbildung 3.15 zeigt ein Beispiel der rückwärtsadaptiven Einhüllendenschätzung für das SQAM-Signal [EBU88] 27 Castanets st rhythm (linker Kanal, ab Sekunde 0,559) für $\lambda_{\text{AT}}=0,8$ und $\lambda_{\text{RT}}=0,1$. Hierbei wurde das Signal direkt als Eingangssignal der Einhüllendenschätzung verwendet und die Wellenform wird zudem als Betragsverlauf dargestellt.

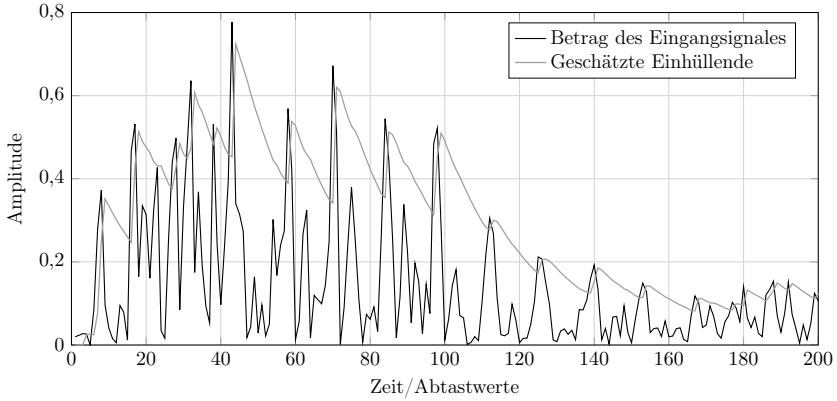


Abbildung 3.15: Einhüllendenschätzung am Beispiel des SQAM-Signales *Castanets*.

Es ist deutlich zu erkennen, wie der Verlauf des Normalisierungssignales auf steigende Signalpegel reagiert und bei fallenden Amplituden langsam gegen Null bzw. v_{\min} strebt und damit die Einhüllende des Eingangssignales näherungsweise nachgebildet wird. Allerdings kommt es bei sehr starken Anstiegen der Signalamplitude zur Unterschätzung, so dass der Pegel des Eingangssignales den der Einhüllenden deutlich übersteigt. Dieses findet zwar jeweils nur für wenige Samples statt, führt aber zu einer Amplitude des normalisierten Signales, die deutlich größer als Eins ist. Zusätzlich können die auftretenden Maximal- und Minimalamplituden, wie bereits erwähnt, in der Auslegung des Quantisierers durch Skalierung seiner Kennlinie berücksichtigt werden, so dass der Anteil und das Ausmaß an wirklicher Übersteuerung des Quantisierers gering bleibt. Die Überschätzung, die zu einer nicht vollständigen Aussteuerung des Quantisierers führt, ist weniger kritisch und wird in Abwägung

einer besseren Schätzung für stationäre Signale in Kauf genommen.

Zur effizienteren Berechnung wird in [Hol09] vorgeschlagen, durch Einsetzen von $\tilde{e}(n) = v(n) \cdot \tilde{e}_{norm}(n)$ in Gleichung 3.20 eine Umformung zu

$$v^2(n) = \underbrace{(1 - \lambda(n) + \lambda(n)\tilde{e}_{norm}^2(n-1))}_{r_k^2(n)} \cdot v^2(n-1) \quad (3.22)$$

vorzunehmen, wobei sich damit die Bedingung

$$\lambda(n) = \begin{cases} \lambda_{AT} & \text{für } \tilde{e}_{norm}^2(n-1) > 1 \\ \lambda_{RT} & \text{sonst.} \end{cases} \quad (3.23)$$

ergibt. Damit lässt sich dann $r_k(n)$ über eine *Lookup-Tabelle* in Abhängigkeit von $\tilde{e}_{norm}(n-1)$ (und damit $\hat{q}(n-1)$) berechnen und der gesamte Quantisierer wie in Abbildung 3.16 dargestellt realisieren.

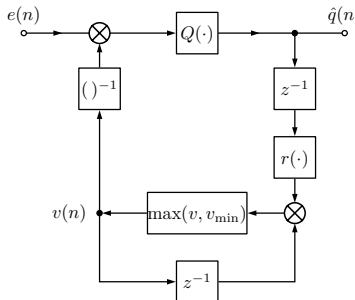


Abbildung 3.16: Modifizierter adaptiver Quantisierer gemäß [Hol09].

Zum Erreichen einer Robustheit gegenüber Fehlern bei der Übertragung der zur Adaption genutzten Quantisierungsindizes $\hat{q}(n)$ über einen realen Kanal ist es notwendig, die Rekursion in der Berechnung nach Gleichung 3.20 aufzubrechen. Dies lässt sich z.B. durch Einführen eines zusätzlichen Leakage-Faktors $\zeta < 1$ im Exponenten der Rückkopplung von $v(n)$ erreichen (vgl. auch [SHZ14]). Die modifizierte Gleichung zur Berechnung der Einhüllenden lautet demnach

$$v^2(n) = \lambda(n) \cdot \tilde{e}^2(n-1) + (1 - \lambda(n)) \cdot v^{2\zeta}(n-1), \quad (3.24)$$

wobei $\lambda(n)$ wiederum nach Gleichung 3.21 adaptiert wird.

Abbildung 3.17 zeigt simulierte Verläufe der Normalisierungssignale $v(n)$ sowie deren Differenz für einen Encoder und Decoder gemäß der ADPCM nach Abbildung 3.1. Hierbei wurde die Prädiktion abgeschaltet und eine Quantisierung mit einem symmetrischen 4 Bit Quantisierer, dessen Maximal- und Minimalamplituden an die

des normalisierten Signales angepasst wurden, durchgeführt. Eingangssignal der adaptiven Quantisierung ist ein Sinussignal der Frequenz 1 kHz mit einer Maximalamplitude von -3 dB. Die Zeitkonstanten wurden wiederum zu $\lambda_{AT}=0,8$ und $\lambda_{RT}=0,1$ gewählt. In den Verläufen der Einhüllenden ist zu sehen, dass der Pegel des Sinussignales, wie für stationäre Signale zu erwarten, bis auf kleine Schwankungen verhältnismäßig gut geschätzt wird.

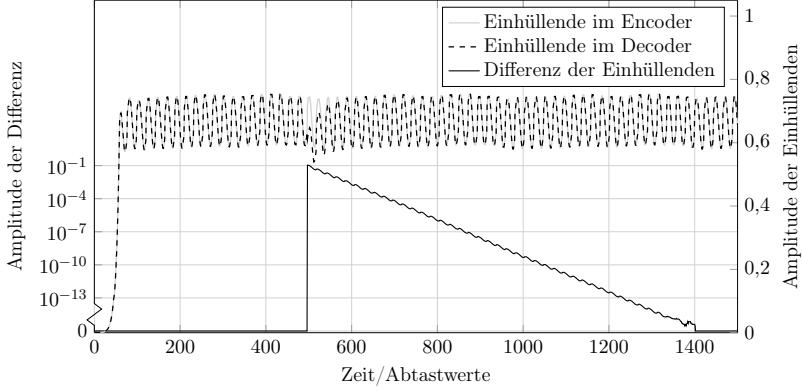


Abbildung 3.17: Simulierte Verläufe der Normalisierungssignale von Encoder und Decoder bei eingekoppeltem Fehler sowie deren Differenz.

Zum Demonstrieren des Einflusses von Übertragungsfehlern wurde der Austausch eines Amplitudenwertes im Maximum der quantisierten Wellenform durch den kleinst möglichen positiven Repräsentativwert des Quantisierers nach ca. 500 Samples vorgenommen. Durch die Abweichung in der Amplitude des Rekonstruktionssignales im Decoder weichen die Einhüllenden von Encoder und Decoder ab dem Zeitpunkt des fehlerhaften Rekonstruktionswertes voneinander ab. Die Differenz fällt jedoch kontinuierlich exponentiell über der Zeit ab und konvergiert, sofern kein erneuter Fehler auftritt, garantiert gegen Null.

Auch für die fehlerrobuste Version der adaptiven Quantisierung lässt sich eine Umformulierung für eine Nutzung von Lookup-Tabellen durchführen. Demnach müssen zum Erreichen der Fehlerrobustheit keine signifikanten Einbußen in Bezug auf den Rechenaufwand in Kauf genommen werden.

Der Einfluss von Übertragungsfehlern im quantisierten Prädiktionsfehler auf die Hüllkurvenschätzung und damit die Prädiktoradaption und Rekonstruktion des dekodierten Signales lässt sich darüber hinaus weiter verringern, indem eine so genannte Prädiktionsfehlernormalisierungsfaktorübertragung in die Codecstrukturen eingeführt wird. Hierzu wird encoderseitig in definierten zeitlichen Abständen eine quantisierte Version des Normalisierungsfaktors berechnet und zur weiteren Adaption verwendet sowie diese auch zum Decoder übertragen und dort anstelle des aus dem Prädiktionsfehler berechneten Wertes in die ADPCM-Struktur eingeleitet. Dies

führt dazu, dass die Differenz der Hüllkurven zwischen Encoder und Decoder ab dem Zeitpunkt dieses Einleitens schlagartig auf Null absinkt, was auch das Einschwingen der Prädiktoradaption beschleunigt und die Auswirkungen des Fehlers auf das Rekonstruktionssignal begrenzt. Das Erkennen von Fehlern in den übertragenen Hüllkurvenwerten kann dabei über einen geeigneten Fehlerschutz sowie eine zusätzliche Detektions-Algorithmitk realisiert werden [SHZ14].

3.3.3 Einführung einer dynamischen Bit-Allokation

Die rückwärtsadaptive Normalisierung des Prädiktionsfehlersignales im Zuge der adaptiven Quantisierung erlaubt es zwar, eine sehr gute mittlere Aussteuerung der Subband-Quantisierer zu erreichen, dennoch ist es für ein verbessertes Codierergebnis von Vorteil, zusätzlich eine so genannte *dynamische Bit-Allokation* (dBA) einzuführen. Hierbei wird den Quantisierern der Blöcke zur adaptiven Prädiktion und Quantisierung $AdPQ_i$ auf Basis eines Allokationsalgorithmus dynamisch der Anteil an einer im Vorfeld festgelegten Gesamtbitzahl zugewiesen. Dies erfolgt mit dem Ziel, eine Art indirekte adaptive Rauschformung zu erreichen.

Abbildung 3.18 zeigt einen Flussgraphen der Subband-ADPCM, der um eine dynamische Bit-Allokation erweitert wurde. In einem hier mit „ALLOK. & ZUW.“ (Allokation und Zuweisung) bezeichneten Block werden dabei auf Basis von Informationen, die aus den Subband-Signalen gewonnen werden, die Allokationssignale ermittelt, mit denen die Subband-Quantisierer gesteuert werden. Das Umschalten zwischen verschiedenen Allokationsergebnissen geschieht dabei nicht im Takt der Subband-Signale, sondern die ermittelte Allokation wird jeweils für eine auf den Takt des Eingangssignales bezogene Blockgröße J konstant gehalten, um den Aufwand für die Signalisierung gering zu halten.

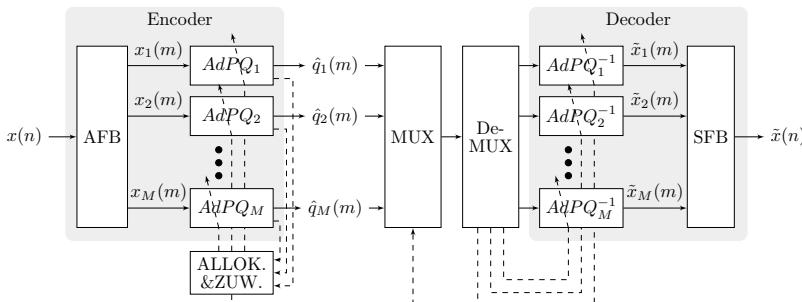


Abbildung 3.18: Flussgraph der ADPCM-basierten Subband-Codierung mit Erweiterung um eine dynamische Bit-Allokation.

Die Nutzung einer dynamischen Bit-Allokation macht es nötig, mit dem Block „MUX“ eine Art *Multiplexer* einzufügen, der aus den Subband-Quantisierungsindizes und der

Signalisierung der Allokation einen seriellen Datenstrom mit entsprechender Struktur erzeugt. Die Übertragung der Daten erfolgt damit in Rahmenstruktur, wobei die Rahmenköpfe die Allokation für den folgenden Block sowie nötige Synchronisationsinformationen beinhalten. Im Decoder erfolgt die Verteilung der Daten durch einen entsprechenden *Demultiplexer* (De-MUX) und die Steuerung der Blöcke $AdPQ_i^{-1}$ für die Abbildung der Quantisierungsindizes auf die entsprechenden Rekonstruktionswerte und deren Nutzung zur Decodierung und Adaption.

In der Literatur lassen sich Methoden zur Bit-Allokation auf Basis einfacher Metriken finden, die in Bezug auf das Allokationsergebnis vergleichbare Ergebnisse wie deutlich komplexere Methoden liefern [MLRR99]. Diese wurden im Zuge der Untersuchungen zu [PO15a] adaptiert, analysiert und modifiziert. Abbildung 3.19¹⁴ verdeutlicht das dabei beim Verteilen der Bits auf die Subband-Quantisierer erfolgende Vorgehen.

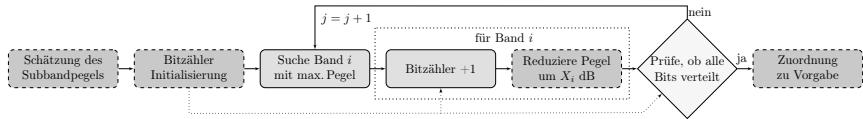


Abbildung 3.19: Flussgraph der Berechnungen zur dynamischen Bit-Allokation.

Hierbei arbeitet der Kern der Routine nach einer Art „Abstimmungsverfahren“. Am Eingang des Algorithmus wird eine Schätzung des Pegels $s_i(n)$ der Subband-Signale bzw. Subband-Prädiktionsfehlersignale durchgeführt. Hierbei wurde zum Vermeiden von unnötigen Umschaltvorgängen, ähnlich wie bei der adaptiven Hüllkurvenschätzung, eine adaptive Zeitkonstante $\alpha(n)$ eingeführt. Die Pegelschätzung erfolgt in [PO15a] zudem nicht direkt auf dem Prädiktionsfehlersignal, sondern auf Basis einer zusätzlichen Filterung der über $v(n)$ geschätzten Einhüllenden des i -ten Bandes. Dies hat im Rahmen verschiedenster Experimente mit natürlichen und synthetischen Signalen deutlich konsistenter Allokationsergebnisse ergeben. Optional kann die Pegelschätzung aber auch direkt über eine Filterung der Subband-Signale erfolgen, sofern daraus eine ausreichende Glättung resultiert. In jedem Fall ergibt sich eine Formulierung gemäß

$$s_i(n) = (1 - \alpha(n)) \cdot s_i(n-1) + \alpha(n) \cdot v_i(n), \quad (3.25)$$

wobei die zeitliche Adaption über

$$\alpha(n) = \begin{cases} \alpha_{\text{AT}} & \text{falls } v_i(n) > s_i(n-1) \\ \alpha_{\text{RT}} & \text{sonst} \end{cases} \quad (3.26)$$

¹⁴Gestrichelt umrandete Blöcke verdeutlichen dabei die im Vergleich zu [MLRR99] erfolgten Modifikationen bzw. Ergänzungen. Diese betreffen im Wesentlichen die im Folgenden beschriebene veränderte Schätzung der Subbandpegel, die Initialisierung der Subband-Bitzähler sowie die Zuordnung des Allokationsergebnisses auf eine Bitkombination aus einer vorgegebenen Menge.

realisiert wird.

Wie im zweiten Block in Abbildung 3.19 dargestellt, erfolgt vor dem Eintritt in die Iterationsschleife zunächst eine Initialisierung der Subband-Bitzähler. Diese erfolgt entweder rein statisch oder zum Teil in Abhängigkeit der geschätzten Subband-Pegel und kann Teil der in Kapitel 4 präsentierten Parameteroptimierung sein. In der eigentlichen Routine zur Verteilung einzelner Bits auf die Subbänder wird nun nach dem Band mit dem maximalen Pegel gesucht und dessen Bitzähler um Eins erhöht sowie eine Reduktion des Subband-Pegels des betroffenen Bandes um einen definierten Pegel X_i vorgenommen. Falls noch weitere Bits zu verteilen sind, beginnt eine neue Iteration solange bis alle Bits verteilt wurden.

Die Reduktion X_i des Pegels des i-ten Bandes wird in [MLRR99] statisch mit 6 dB angegeben. In den Untersuchungen zu [PO15a] wurde jedoch für die modifizierte Schätzung des Subband-Pegels festgestellt, dass eine optimierte sowie vom Iterationsindex j abhängige Verminderung des Subband-Pegels bessere Ergebnisse liefern kann. Die Erhöhung des Reduktionsfaktors mit steigender Iterationszahl hilft hierbei im Falle von sehr großen Pegelunterschieden der Subband-Signale ein ungewolltes Abschalten der Bänder mit geringem Pegel zu verhindern. Die Reduzierung X_i des Pegels in der j -ten Iteration lässt sich demnach über

$$X_i(\phi, \rho, \kappa, j) = 20 \cdot \log_{10} \left(\frac{1}{\phi \cdot (2\rho)^{\kappa \cdot j}} \right) \quad (3.27)$$

mit

$$\phi > 1, \rho > 0,5, \kappa \geq 0 \quad (3.28)$$

parametrisieren und damit einer globalen Optimierung (vgl. Kapitel 4) zugänglich machen. Hierbei kann über ein Einstellen von $\kappa = 0$ und $\phi = 2$ auch die 6 dB Reduktion aus [MLRR99] erreicht werden. Eine derart freie Formulierung ist zwar nicht zwingend nötig, erlaubt es jedoch weitere Freiheitsgrade bei der Optimierung der dBA auszunutzen und damit eine bessere Anpassung an unterschiedliche Signalklassen und ihre Eigenschaften zu erreichen.

Abbildung 3.20 zeigt exemplarisch die zeitlichen Verläufe der Allokationsergebnisse für Ausschnitte verschiedener SQAM-Signale wie Kastagnetten, Triangel und Sprache zusammen mit deren auf die Abtastfrequenz der Subbandsignale unterabgetasteten Wellenformen des Eingangssignales des Subband-Codecs. Die dBA ist hierbei so eingestellt, dass eine zeitweise Abschaltung der oberen Bänder zugelassen wird. Es ist deutlich zu sehen, wie sich die Allokationsverläufe für verschiedene Signalklassen unterscheiden und wie die dynamische Bit-Allokation auf wechselnde Signaleigenschaften über eine veränderte Bitzuweisung reagiert.

Eine weitere Modifikation der dBA betrifft die optionale Begrenzung der Menge möglicher Bit-Allokationen. Sie basiert auf der Beobachtung, dass nicht alle theoretisch möglichen Bit-Allokationskombinationen für natürliche Audiosignale mit der vorgestellten Methodik auftreten. Demnach ermöglicht die Begrenzung auf eine definierte Anzahl an erlaubten Allokationskombinationen unnötigen Signalisierungsaufwand zu vermeiden, ohne das Codierergebnis wesentlich zu beeinflussen. Abbildung

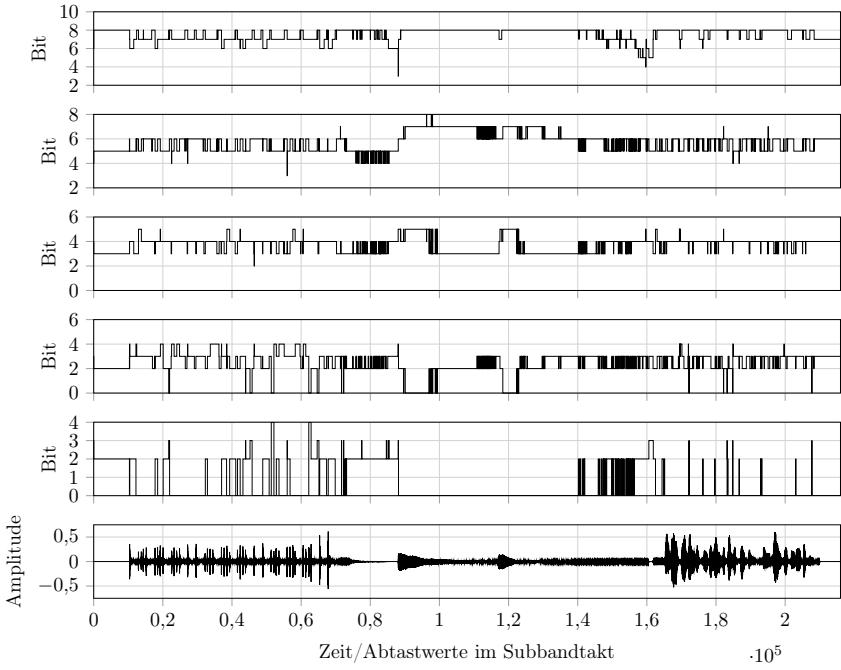


Abbildung 3.20: Beispiel der Subband-Allokationsverläufe für Ausschnitte von SQAM-Signalen verschiedener Signalklassen (Bänder absteigend mit aufsteigender Bandzahl sortiert) im Vergleich zur Wellenform des Eingangssignales.

3.19 enthält daher zusätzlich einen nachgeschalteten Block zur Zuweisung des Allokationsergebnisses zu dem Ähnlichsten aus einer vorgegebenen Menge an erlaubten Bitkombination. Dieses Vorgehen reflektiert implizit auch die höhere Bedeutung einer Übereinstimmung der Bit-Zahlen unterer Bänder.

Die Suche nach einer möglichst einfachen Algorithmik zum Abbilden eines Allokationsergebnisses auf das erlaubte Bitmuster führte zu einer Formulierung über die Darstellung des Bit-Allokationsvektors \mathbf{b} als Dezimalzahl $D_{\text{alloc.}}$ über

$$D_{\text{alloc.}}(\mathbf{b}) = \sum_{i=1}^M \mathbf{b}(i) \cdot 10^{(M-i)} \quad (3.29)$$

und die Suche innerhalb der ebenfalls umgerechneten Kandidatenmenge nach demjenigen mit dem geringsten Betrag der Differenz zum ursprünglichen Allokationsergebnis.

Das Generieren der Menge am häufigsten genutzter Bitmuster kann dabei Teil der in Kapitel 4 vorgestellten Optimierung der Codecparameter sein. Hierbei werden für

einen gegebenen Parametervektor und das bei der Optimierung zu codierende Testset die Allokationssignale der Subbänder ermittelt und hieraus die Häufigkeitsverteilung der Allokationsmuster erstellt. Die mit der Einführung der dBA benötigten Codebücher eines Subbandes mit verschiedenen Wortbreiten werden dabei aus einem, gemäß dem Vorgehen in Abschnitt 3.3.1 erzeugten, Basiscodebuch durch Unterabtastung abgeleitet.

Mit der Beschränkung der Menge möglicher Allokationsmuster lässt sich, wie in [PO15a] gezeigt, die Datenrate für die Signalisierung der Bit-Allokation um bis zu ca. 75 % reduzieren ohne signifikante Einbußen bei der Audioqualität hinnehmen zu müssen. Je nach Einstellung der, bei der Übertragung verwendeten, Rahmenlänge kann diese Reduktion damit eine relevante Größenordnung erreichen. Da das Erzeugen der Menge erlaubter Bit-Allokationen Teil der globalen Parameteroptimierung ist, lässt sich für eine vorgegebene Rahmengröße und Datenrate zur Signalisierung sowie ein gegebenes Testset somit ein optimaler Satz vorgegebener Allokationsmuster bestimmen.

Tabelle 3.2 enthält die Ergebnisse einer Evaluation der Verbesserungen der Audioqualität einer früheren Version des Subband-Codecs mit dem PEAQ-Verfahren (vgl. Abschnitt 4.1) bei Einführung einer dynamischen Bit-Allokation und einer globalen Parameteroptimierung (vgl. Kapitel 4) aus [PO15a].

In den evaluierten Fällen enthält der Codec folgende Konfigurationen:

- Fall 1: Keine Bit-Allokation, statische Bit-Verteilung {6,5,3,3,3}, manuell optimierte Parameter
- Fall 2: Bit-Allokation, manuell optimierte Parameter, keine Begrenzung der Menge erlaubter Allokationen
- Fall 3: Bit-Allokation, global optimierte Parameter, Begrenzung der Menge erlaubter Allokationen.

Die Ergebnisse verdeutlichen, dass durch Einführen einer dynamischen Bit-Allokation eine deutliche Verbesserung der PEAQ-ODG Werte für kritische Signale und damit deren Audioqualität erreicht werden kann. Zusätzlich lassen sich weitere Gewinne durch die globale Optimierung der Parameter erzielen. Die dabei eingeführte Begrenzung der Menge erlaubter Allokationen resultiert bei keinem der Signale in einer signifikanten Verschlechterung des Ergebnisses.

Natürlich resultiert ein Großteil des Gewinnes in der Audioqualität aus der Tatsache, dass bei Nutzung der dynamischen Bit-Allokation den unteren Bändern zeitweise eine höhere Zahl an Bits zugeordnet wird. Da im Fall der statischen Bit-Verteilung eine dauerhafte Abschaltung eines Bandes nicht in Frage kommt, um den Anspruch eines Vollband-Audiocodecs gerecht zu werden, erfolgte in der präsentierten Evaluation die Auswahl einer statischen Bit-Verteilung, die eine ausreichende Rekonstruktionsqualität in den oberen Bändern sicherstellt.

Tabelle 3.2: Ergebnisse einer Evaluation der objektiven Audioqualität verschiedener Signalklassen für unterschiedliche Codecversionen (aus [PO15a]).

Signalname	PEAQ-ODG		
	Fall 1	Fall 2	Fall 3
Gesang/S. Vega	-0,66	-0,69	-0,46
Sprache männl. Eng.	-0,79	-0,74	-0,56
Sprache weibl. Eng.	-0,76	-0,63	-0,50
Trompete	-1,16	-0,61	-0,43
Orchester	-0,65	-0,71	-0,54
Big Band	-0,42	-0,45	-0,40
Cembalo	-0,46	-0,40	-0,32
Kastagnetten	-2,08	-1,90	-0,93
Stimmpfeife	-0,86	-0,47	-0,44
Dudelsack	-0,64	-0,51	-0,46
Glockenspiel	-3,25	-1,91	-0,82
Gezupfte Saiten	-0,49	-0,49	-0,51
Mittelwert	-1,02	-0,79	-0,53

3.3.4 Vektorquantisierung

Die adaptive skalare Quantisierung der Subband-Prädiktionsfehlersignale mit nicht-linearen Quantisierererkennlinien und einer dynamischen Zuweisung von Quantisiererauflösungen auf die Subbänder ermöglicht bereits eine sehr gute Audioqualität der Subband-ADPCM. Dennoch ist es denkbar, durch Einsatz von so genannter *Vektorquantisierung* zusätzliche Gewinne in der Rekonstruktionsqualität der ADPCM zu erzielen und damit auch die Audioqualität des Codierungsverfahrens weiter zu erhöhen.

Abbildung 3.21 zeigt einen Vergleich der segmentellen SNRs von gleichförmiger und nach dem *A-Law* kompanderter Quantisierung mit dem einer *Sphärisch Logarithmischen Quantisierung* (SLQ) (vgl. Matschkal [Mat08]) für das normalisierte SQAM-Signal 39 und eine resultierende Codewortlänge von $R = 8$ Bits/Sample sowie einem *A*-Parameter von $A = 2229,7$ und einer Dimensionalität der SLQ von $D= 6$. Es ist zu erkennen, dass vergleichbar zur nichtlinearen Quantisierung nach dem *A-Law* das SNR über einen weiten Aussteuerungsbereich konstant ist und dabei in der Größenordnung des Maximums der gleichförmigen Quantisierung liegt.

Die SLQ basiert auf einer Darstellung eines Eingangssignalvektors \boldsymbol{x} der Länge D in sphärischen Koordinaten und der Suche nach einer Quantisierungszeile auf einer D -dimensionalen Sphäre. Die damit vorliegenden $D-1$ Winkel und der Radius werden dabei getrennt quantisiert, wobei die Quantisierung des Radius zusätzlich gemäß des *A-Laws* nichtlinear erfolgt. Abbildung 3.22 zeigt die Verteilung der Quantisie-

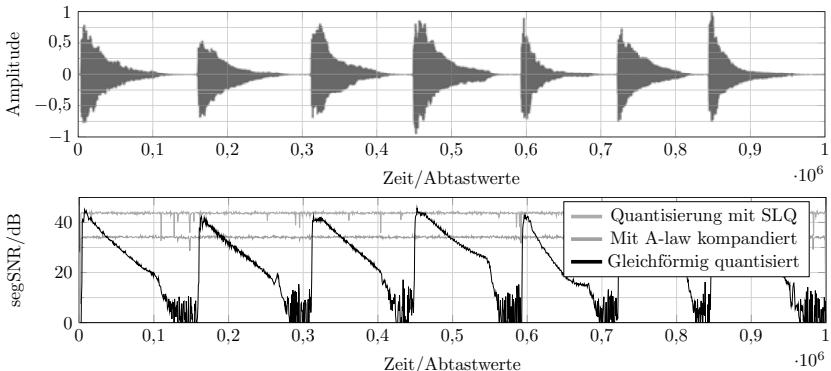


Abbildung 3.21: Vergleich segmenteller SNRs verschiedener Quantisierungsverfahren mit einer resultierenden Codewortlänge $R = 8$ Bits/Sample für einen Teil (Min. 01:02-01:23) des normalisierten SQAM-Signales 39.

rungszellen und entsprechenden Rekonstruktionsvektoren auf der Oberfläche einer dreidimensionalen Sphäre (a) sowie die logarithmische Quantisierung des Radius zum Erhalten näherungsweise gleicher Kantenlängen (b) für $D= 3$ und eine resultierende Codewortlänge von $R= 4$ Bits/Sample.

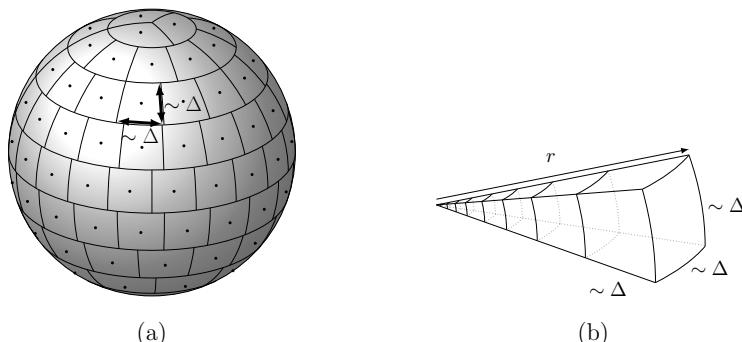


Abbildung 3.22: (a) Verteilung der Quantisierungszellen und Rekonstruktionsvektoren (\bullet) auf der Oberfläche einer dreidimensionalen Sphäre (b) Logarithmische Quantisierung des Radius r für näherungsweise gleiche Kantenlängen Δ (beide gemäß [MH10]).

Während das Grundkonzept der SLQ intuitiv und verhältnismäßig einfach ist, basieren die Herleitungen zur optimalen Verteilung der Quantisierungszellen und die Algorithmen zur Indexsuche auf einer umfangreichen mathematischen Formulierung.

An dieser Stelle wird daher nur das Grundkonzept erläutert und die nötige Terminologie sowie die Parameter eingeführt, um das gewählte Vorgehen zu motivieren und zu beschreiben. Für weitere Details und Formeln sei auf [Mat08] verwiesen.

In [Mat08] und [MH10] finden die Untersuchungen zur SLQ bereits im Kontext der Audiocodierung mit geringer Latenz statt und es wird eine Algorithmik zur Anwendung in Kombination mit ADPCM präsentiert. Allerdings liegt der Fokus hierbei auf einem Nachweis der grundsätzlichen Eignung des Verfahrens und es wird auf aufwändigere Maßnahmen zur Verbesserung der Audioqualität verzichtet. Daher verstehen die Autoren von [MH10] die Untersuchungen mehr als eine Art Machbarkeitsstudie. Zusätzlich erfolgt die Evaluation des Codierergebnisses lediglich mit Hilfe von objektiven Messungen durch z.B. das segmentelle SNR.

Ziel der Arbeiten zu [Lam15], die unter anderem auch in [PLO16] resultierten, war es demnach, in einem ersten Schritt, unabhängig vom Aufwand bei der Berechnung, den Einfluss auf die Audioqualität einer ADPCM bei Nutzung einer Vektorquantisierung abzuschätzen. Dazu wurde ermittelt, welche Gewinne in der PEAQ-ODG durch Kombination von SLQ und Breitband-ADPCM mit statischer Rauschformung, die dem Stand der Technik entspricht, erreicht werden können.

Als Referenz wurde dazu das in [SHZ13a, SHZ14] dokumentierte und evaluierte Verfahren zur fehlerrobusten Breitband-ADPCM mit einer statischen Rauschformung durch Rückkopplung des Codierfehlers $e_{ADPCM}(n) = \tilde{x}(n) - x(n)$ verwendet. In diesem Codec erfolgt die Prädiktoradaption nach dem fehlerrobusten GAL-Algorithmus und die adaptive Quantisierung durch eine Version der fehlerrobusten Hüllkurvenschätzung (vgl. Abschnitt 3.2 und 3.3).

Vektorquantisierung und Breitband-ADPCM

Auf den ersten Blick scheint eine Kombination von Vektorquantisierung und breitbandiger ADPCM mit ihrer adaptiven Verarbeitung im Takt des Eingangssignales unmöglich. Dies liegt vor allem daran, dass die Aktualisierung des Prädiktors auf Basis des Rekonstruktionssignals $\tilde{x}(n)$ erfolgt und dieses aus dem quantisierten Prädiktionsfehler berechnet wird, der im Falle der Vektorquantisierung in der Regel jedoch erst vorliegt, wenn ein ganzer Block bzw. Vektor an Signalwerten quantisiert wurde. Auf die SLQ bezogen bedeutet dies, dass die Entscheidung für eine Quantisierungszelle vom Radius abhängt, der aber erst mit dem Vorliegen aller Signalwerte berechnet werden kann.

Um dieses Problem zu lösen, hat Matschkal einen Algorithmus vorgeschlagen, der iterativ den quantisierten Radius ermittelt und eine sukzessive Indexsuche beinhaltet, so dass schon vorab Rekonstruktionswerte für die Adaption der ADPCM verwendet werden können. Algorithmus 1 fasst das dabei durchzuführende Vorgehen zur Bestimmung des SLQ-Indexes im Encoder zusammen. Der Eingangssignalvektor x wird hier aus dem normalisierten Prädiktionsfehler $e_{norm}(n)$ gebildet, da die adaptive Signalnormalisierung auch in der SLQ-ADPCM zum Einsatz kommt.

Am Anfang des Algorithmus wird in einem initialen Schritt eine Schätzung des Radius r durchgeführt. Hierzu wird der Zustand der ADPCM aus dem vorherigen

Algorithmus 1 : SLQ-ADPCM-Encoder

Schätze r durch Verarbeitung des Vektors \mathbf{x} ohne Quantisierung

Für alle \tilde{r} in $[r + \Delta, r - \Delta]$ {

Quantisiere Radius $\tilde{r} \rightarrow \hat{r}$

Verarbeite durch SLQ-ADPCM $\rightarrow \hat{\mathbf{x}}$

Berechne euklidische Distanz $d^2(\mathbf{x}, \hat{\mathbf{x}})$

}

Suche nach \tilde{r} mit dem geringsten $d^2(\mathbf{x}, \hat{\mathbf{x}})$

Berechne SLQ-Index N_{SLQ} aus N_r und N_φ und übertrage diesen

Zeitschritt verwendet und die Codierung für den aktuellen Block der Länge D bei ausgeschalteter Quantisierung durchgeführt. Das Ergebnis dieser Schätzung wird für die Suche nach einem optimalen Radius innerhalb eines vorgegebenen Suchbereiches benutzt. Innerhalb der Iteration über den Suchbereich wird der Radius quantisiert, mit ihm, durch Verarbeitung in der SLQ-ADPCM, über die sukzessive Winkelquantisierung der quantisierte Signalvektor $\hat{\mathbf{x}}$ bestimmt und jeweils der euklidische Abstand zum Originalvektor ermittelt. Der Block mit dem geringsten Abstand wird dann final benutzt, um den Radiusindex N_r sowie den Winkelindex N_φ zu berechnen und den daraus resultierenden SLQ-Index N_{SLQ} für das Übertragen an den Decoder zu erstellen.

Natürlich führt das wiederholte Berechnen der SLQ-ADPCM zu einem linear mit der Iterationszahl η^{15} steigenden Rechenaufwand, erlaubt aber zugleich eine voneinander unabhängige starke Parallelisierung. Aufgrund der Eigenschaft der Breitband-ADPCM das Signal nicht zu verzögern, resultiert die einzige Latenz der SLQ-ADPCM aus dem Zusammenfassen von D Signalwerten zum blockweisen Verarbeiten. In einem echten System mit einer sequenziellen Übertragung der Daten des SLQ-Indexes und einer begrenzten Rechenkapazität ist aber mit einer zusätzlichen Verzögerung von $D-2D$ Abtasttakten zu rechnen. Mit der üblichen maximalen Dimensionalität von $D=12$ liegt die Latenz der SLQ-ADPCM jedoch immer noch deutlich unter 1 ms. Die Integration der SLQ in die Breitband-ADPCM führt zu zusätzlichen Parametern. Diese und die Parameter der adaptiven Quantisierung und Prädiktion wurden im Rahmen der Untersuchungen zu [PLO16] mit dem Optimierungsansatz gemäß dem Vorgehen in Kapitel 4 optimiert.

Die optimierten Parameter sind demnach (mit den Leakage Faktoren $\alpha=0,98$, $\beta=0,91$ des GAL-Algorithmus):

- D , η und A der SLQ-ADPCM,
- λ_{AT} , λ_{RT} , v_{\min} und $\zeta < 1$ der adaptiven Hüllkurvenschätzung,
- $\hat{\mu}$ und σ_{\min}^2 des GAL-Algorithmus,

sowie ein konstanter Faktor, der zur Skalierung des normalisierten Prädiktionsfehlersignales am Eingang der SLQ genutzt wird, um den Kompromiss zwischen Über-

¹⁵Wobei $10 < \eta < 100$ üblich ist.

und Untersteuerung zu finden.

Für den Vergleich mit den Ergebnissen aus [SHZ14] wurden die dort ausgewählten Signale des SQAM-Testsets in der angegebenen Konfiguration verwendet und mit dem PEAQ-Algorithmus (vgl. Abschnitt 4.1) evaluiert. Abbildung 3.23 beinhaltet eine Auflistung der erzielten Ergebnisse für 3 Bits/Sample.

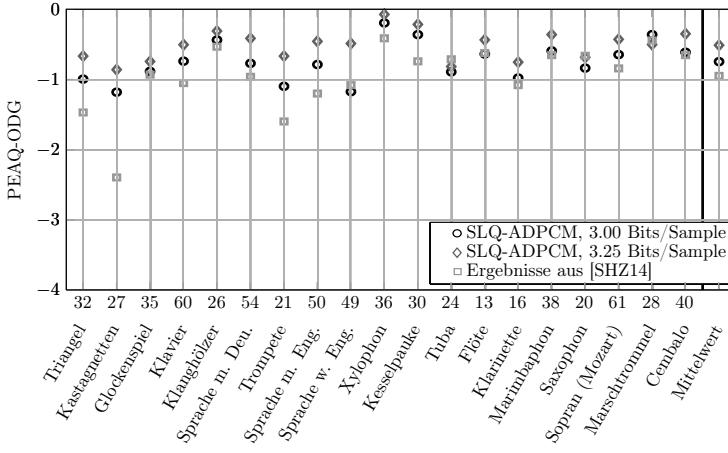


Abbildung 3.23: Vergleich der SLQ-ADPCM zu den Ergebnissen aus [SHZ14].

Es ist zu erkennen, dass für viele der kritischen Signale wie Triangel, Kastagnetten und Trompete Verbesserungen in der ODG von bis zu einem Punkt erreicht werden. Im Mittel liegt die Verbesserung trotz minimaler Verschlechterung für einzelne Signale bei 0,2 in der ODG. Zusätzlich wurde für die synthetischen Signale 1-7 des SQAM-Testsets eine wesentliche Verbesserung mit zum Teil mehr als einem Punkt erzielt. Ein weiterer Vorteil der Breitband-SLQ-ADPCM ist, dass sie rationale Anzahlen in der Bitrate erlaubt. In der Abbildung 3.23 sind demnach auch Ergebnisse für 3,25 Bits/Sample aufgenommen. Diese verdeutlichen, dass mit nur einer geringen Erhöhung der Bitrate die Audioqualität der SLQ-ADPCM in Bereiche fällt, die für die hier anvisierten Anwendungsfälle ausreichend erscheint.

Verbessertes Fehlerverhalten

Ein bekanntes Problem von Vektorquantisierungsverfahren im Allgemeinen ist, dass durch auftretende Einzelbitfehler bei der Übertragung der Quantisierungsindizes jeweils immer ein gesamter Block an Signalwerten gestört wird. Daher resultieren Übertragungsfehler ohne weitere Maßnahmen auch bei der SLQ-ADPCM in starken Störungen des Ausgangssignales, auch wenn der Codec zu keiner Zeit vollständig destabilisiert wird und der eingekoppelte Fehler nach Störungen wieder abklingt. Zur Verbesserung des Fehlerverhaltens ohne aufwändige Kanalcodierung wurde daher ein einfaches, aber wirkungsvolles Vorgehen gewählt. Hierbei erfolgt die Reduktion

der Länge der SLQ-Indizes auf $D \cdot R - 1$ und das Anhängen eines *Paritätsbits*. Bei hierdurch erkannten Einzelbitfehlern werden nun Nullen in den Decoder anstelle des schadhaften Blockes eingeleitet.

Abbildung 3.24 zeigt den Vergleich der Audioqualität bei einer Fehlerrate von 10^{-4} zwischen SLQ-ADPCM und den Ergebnissen aus [SHZ14]. Für die SLQ-ADPCM wurden hierzu die Ergebnisse von 1000 Realisierungen eines zufälligen Bitfehlermusters gemittelt. Die Balken zwischen den Einzelergebnissen verdeutlichen eine Verbesserung oder Verschlechterung der SLQ-ADPCM gegenüber der Referenz. Mit der eingeführten Modifikation kann die SLQ-ADPCM demnach als mindestens ebenbürtig zu den Ergebnissen aus [SHZ14] betrachtet werden. Natürlich lässt sich das Fehlerverhalten, z.B. durch aktive Maßnahmen wie die in [SHZ14] vorgestellte Normalisierungsfaktübertragung, weiter verbessern.

Mit den vorgestellten Untersuchungen konnte gezeigt werden, dass sich durch Ersetzen der skalaren Quantisierung durch Vektorquantisierung in einer Breitband-ADPCM deutliche Verbesserungen in der Audioqualität erreichen lassen, ohne die Verzögerung auf einen kritischen Wert zu vergrößern. Demnach wäre es erstrebenswert, in einem weiteren Schritt, diese Gewinne auch auf den Subband-Codec zu übertragen. Hierbei kann die Anwendung der Vektorquantisierung auf zwei verschiedene Weisen erfolgen. Zum einen kann der Eingangsvektor der SLQ aus den Subband-Signalen eines Abtastaktes gebildet werden. Zum anderen ist es denkbar, in jedem Band aufeinanderfolgende Signalwerte zu einem Vektor zusammen zu fassen und einzeln in der SLQ zu verarbeiten. Im ersten Fall ergibt sich der Vorteil, dass die Radiusberechnung direkt erfolgen kann, so dass die Algorithmik zu dessen Schätzung und damit ein wesentlicher Teil des Rechenaufwandes entfallen kann. Problematisch bei der Anwendung in der vorliegenden 5-Kanal Struktur ist jedoch, dass aufgrund der damit vorliegenden geringen Dimensionalität die Gewinne der SLQ-Quantisierung nicht so hoch ausfallen. Bei sequenzieller Verarbeitung der Subband-Signale mit der SLQ-ADPCM ergibt sich aufgrund der Blockbildung der unterabgetasteten Subband-Signale eine zusätzliche Latenz der Verarbeitung von mindestens $M \cdot D$ Samples, die für die hier anvisierten Anwendungsfälle unter Umständen zu hoch ist. Leider konnten beide hier diskutierten Fälle der Subband-SLQ-ADPCM in Bezug auf ihren Einfluss auf die Audioqualität nicht abschließend untersucht werden, da zumindest für letzteren Fall der Aufwand für die Optimierung der Codec-Parameter den Rahmen der Untersuchungen zu [PLO16] sprengte.

3.4 Maßnahme zur Verbesserung der Subband-Codierung

Die in Abschnitt 3.3.3 beschriebene dynamische Bit-Allokation lässt sich mit verhältnismäßig wenig Aufwand für die Berechnung und verringerten Signalisierungsaufwand in den Subband-Codec integrieren und darin optimieren. Es sind jedoch auch weitaus aufwändiger Maßnahmen zur Erhöhung der Audioqualität oder wei-

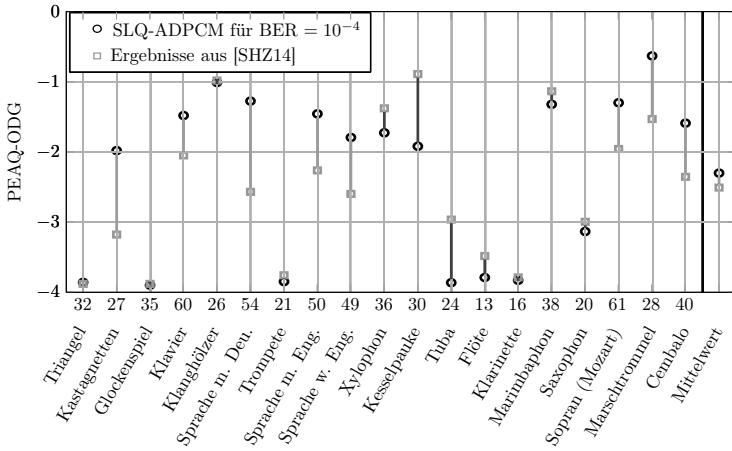


Abbildung 3.24: Vergleich der SLQ-ADPCM zu den Ergebnissen aus [SHZ14] für eine Fehlerrate von 10^{-4} .

teren Reduktion der Datenrate denkbar, deren Praxisrelevanz für den vorliegenden Anwendungsfall nach derzeitigem Stand jedoch nur begrenzt vorhanden ist.

Dennoch wurde im Rahmen der vorliegenden Arbeit eine aussichtsreiche Methoden zur Verbesserung der fehlerrobusten ADPCM-Codierung untersucht und in Bezug auf ihren Einfluss auf das Codierergebnis evaluiert. Hierbei war es von Interesse, ob sich ein Verfahren zur psychoakustischen Rauschformung auf den Subband-Fall anwenden lässt und damit eine Alternative zur dynamischen Bit-Allokation ist, die ohne zusätzlichen Signalisierungsaufwand auskommt.

3.4.1 Teilbandbasierte Vor- und Nachfilterung

Eine Möglichkeit, die wahrgenommene Audioqualität des Subband-Codecs zu verbessern, besteht darin, das Nutzsignal vor und nach der Codierung zusätzlich zu filtern. Die Grundidee der adaptiven Vor- und Nachfilterung, welche auf Arbeiten wie [SYHE02] zurückgeht, ist es, das Ausgangssignal eines verlustbehafteten Audiocodecs derart zu filtern, dass die Leistungsdichte des, als spektral gleichverteilt angenommenen, Codierfehlers in Richtung oder unterhalb der Maskierungsschwelle verschoben und dadurch dessen Hörbarkeit deutlich reduziert wird. Dies bedeutet, dass die Übertragungsfunktion des Nachfilters annähernd der Maskierungsschwelle entspricht und das Vorfilter hierzu exakt invers sein sollte, um den Frequenzinhalt des Nutzsignals nicht zu ändern.

Die Vor- und Nachfilterung mit exakt invertierbaren minimalphasigen Filtern kann so realisiert werden, dass sie keine zusätzliche Verzögerungen verursacht. In [SYHE02] wird jedoch zur Adaption der Filterkoeffizienten ein blockbasiertes psychoakustisches

Modell benutzt, so dass eine unvermeidliche Verzögerung von 4-6 ms resultiert. Aus diesem Grund präsentierte Holters in [Hol09] verschiedene Ansätze zur Ergänzung des bereits erwähnten ADPCM-basierten Breitband-Codecs um eine verzögerungsfreie adaptive Vor- und Nachfilterung. Am erfolgreichsten und praxisrelevantesten ist dabei die teilbandbasierte Vor- und Nachfilterung. Dieser Ansatz wurde auch bereits in [Kei06] für die Anwendung in einem Subband-Codec beschrieben, dort jedoch nicht final in den Codec integriert und demnach auch nicht darin evaluiert. Das Verfahren basiert auf einer groben Schätzung der Maskierungsschwelle und einer Rauschformung durch eine Subband-Dynamikkompression und erlaubt eine zum Teil deutliche Verbesserung der Audioqualität für kritische Signale. Da die Ergebnisse aus [Hol09] sehr vielversprechend sind, wurde der Ansatz für die Nutzung in der in dieser Arbeit betrachteten Subband-Codierung modifiziert und optimiert. Abbildung 3.25 zeigt einen Flussgraphen des um die adaptive Vorfilterung erweiterten Subband-Codecs. Der Basiscodec (graue Boxen) wurde dabei um die funktionalen Blöcke der Vorfilterkaskade und der Gain-Schätzung erweitert, die im Folgenden beschrieben werden. Das Nachfilter im Decoder hat die gleiche Struktur, enthält aber die Filter der Nachfilterkaskade in umgekehrter Reihenfolge, um die perfekte Inversion der Übertragungsfunktion zu gewährleisten. Anstelle eines blockbasierten psychoakustischen Modells werden die benötigten und hier als *Gain* bezeichneten Verstärkungen oder Pegelreduktionen der aus K Bändern bestehenden Vorfilterkaskade direkt aus den rekonstruierten Subband-Signalen $\tilde{x}_i(m)$ geschätzt. Dies hat den erwähnten zusätzlichen Vorteil, dass im Gegenteil zur dynamischen Bit-Allokation keine Seiteninformationen in Form von Filterkoeffizienten oder Gain-Werten übertragen werden müssen.

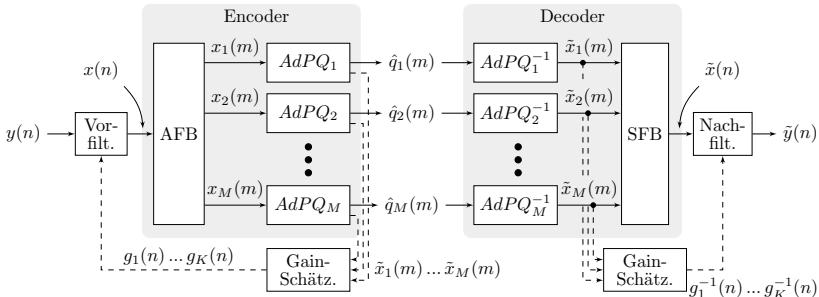


Abbildung 3.25: Flussgraph der ADPCM-basierten Subband-Codierung mit Erweiterung um eine adaptive Vor- und Nachfilterung.

An dieser Stelle wird die Struktur und Formulierung der von Holters abgeleiteten Vor- und Nachfilterung nur kurz zusammengefasst, um die zur Verwendung im Subband-Codec erfolgten Modifikationen und das Vorgehen bei der Evaluation darstellen zu können. Für die Herleitungen der benutzten Shelving-Filter aus zeitkontinuierlichen Prototypen und weitere Diskussionen über die Adaption der Filterko-

effizienten sei auf [Hol09] und die dort angegebenen Referenzen sowie auf [PWO15] verwiesen.

Abbildung 3.26 verdeutlicht den detaillierten Aufbau der Vorfilterkaskade und Gain-Schätzung.

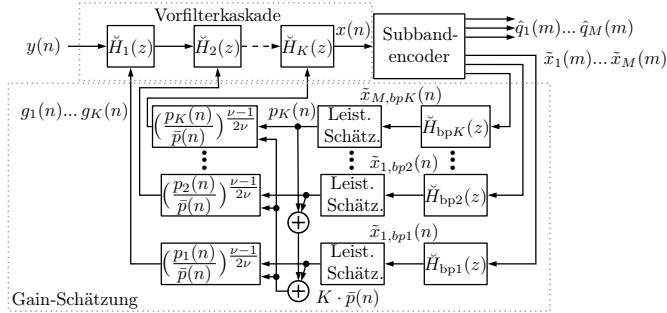


Abbildung 3.26: Flussgraph der Leistungsschätzung zur Subband-Codierung mit adaptiver Vorfilterung im Encoder (angelehnt an [Hol09]).

Der erste funktionale Block ist dabei das eigentliche Vorfilter, das aus einer Kaskade von K so genannten *Shelving-Filtern* besteht. Der Frequenzgang des k -ten Filters mit dem gewünschten und zunächst als statisch angenommenen Gain $g_k(n)$ wird dabei aus einem „Prototyp“ erster Ordnung gemäß

$$\check{H}_{k,\text{prot.}}(z) = \frac{G_k \sqrt{g_k} + 1 + (G_k \sqrt{g_k} - 1) \cdot z^{-1}}{\frac{G_k}{\sqrt{g_k}} + 1 + \left(\frac{G_k}{\sqrt{g_k}} - 1\right) \cdot z^{-1}} \quad (3.30)$$

gewonnen. Der Koeffizient G_k wurde beim Ableiten der Filter $\check{H}_{k,\text{prot.}}(z)$ aus einem zeitkontinuierlichen Prototyp durch bilineare Transformation eingeführt und erlaubt ein Einstellen der Grenzfrequenz. Er kann über

$$G_k = \tan\left(\frac{\omega_{w,k}}{2}\right) \quad (3.31)$$

berechnet werden, wobei die Breite ω_w des k -ten Bandes aus der gewünschten oberen und unteren Grenzfrequenz über $\omega_{w,k} = \omega_{u,k} - \omega_{l,k}$ ermittelt wird.

Die „Prototypen“ $\check{H}_{k,\text{prot.}}(z)$ werden frequenzverschoben, indem die Einheitsverzögerungen durch Alpässen gemäß

$$A_k(z) = z^{-1} \frac{\cos(\omega_{c,k}) - z^{-1}}{1 - \cos(\omega_{c,k}) z^{-1}} \quad (3.32)$$

ersetzt werden. Hierbei ist $\omega_{c,k}$ die gewünschte Mittenfrequenz und berechnet sich

über

$$\omega_{c,k} = 2 \arctan \left(\sqrt{\tan\left(\frac{\omega_{l,k}}{2}\right) \cdot \tan\left(\frac{\omega_{u,k}}{2}\right)} \right). \quad (3.33)$$

Aus diesem Vorgehen resultiert ein Satz an Shelving-Filtern zweiter Ordnung mit den Übertragungsfunktionen

$$|\check{H}_k(e^{j\omega})| = \sqrt{\frac{(\cos(\omega_{c,k}) - \cos(\omega))^2 + (G_k \sin(\omega))^2 g_k}{(\cos(\omega_{c,k}) - \cos(\omega))^2 + \frac{(G_k \sin(\omega))^2}{g_k}}}, \quad (3.34)$$

die in einer modifizierten Direktstruktur implementiert werden können. Diese reduziert den Berechnungsaufwand bei der, aus der zeitlichen Adaption der Gains $g_k(n)$ hervorgehenden, Neuberechnung der Filterkoeffizienten.

Die oberen und unteren Grenzfrequenzen $\omega_{u,k}$ und $\omega_{l,k}$ und damit die Mittenfrequenz $\omega_{c,k}$ werden so gewählt, dass sie näherungsweise der Bark-Skala [Zwi61] mit einer unteren Grenzfrequenz von 0 Hz entsprechen, was in $K= 25$ Bändern resultiert.

Abbildung 3.27 zeigt ein Beispiel für die sich so ergebenden Shelving-Filter Übertragungsfunktionen $\check{H}_k(e^{j\Omega})$ (unten) und der daraus resultierenden Vorfilterübertragungsfunktion $\check{H}_{\text{pre}}(e^{j\Omega})$ (oben) für den Fall, dass alle g_k zu $\sqrt{10}$ gesetzt werden mit Ausnahme von g_9 , welches den Wert $0,1 \cdot \sqrt{10}$ erhält. Dies führt zu dem Einbruch der Gesamtübertragungsfunktion bei 1 kHz.

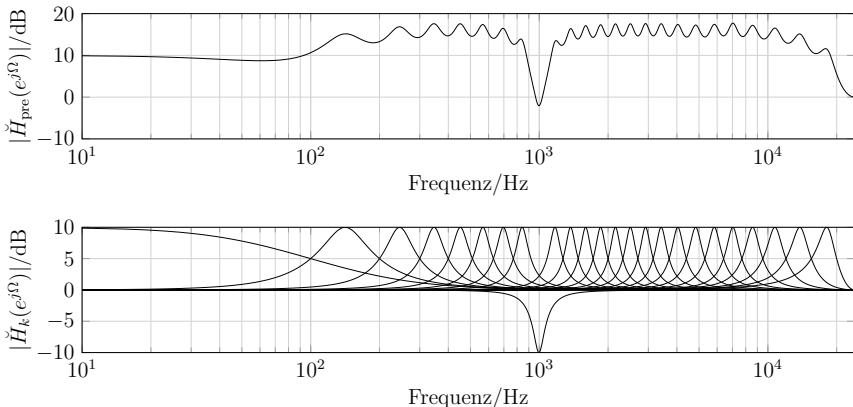


Abbildung 3.27: Beispiel der Shelving-Filter Übertragungsfunktionen (unten) und resultierende Vorfilterübertragungsfunktion (oben), wobei $g_k = \sqrt{10}$ für $k = 1, \dots, K$ bis auf $g_9 = 0,1 \cdot \sqrt{10}$.

Aufgrund der Tatsache, dass die Shelving-Filter selbst nach der Transformation lediglich zweiter Ordnung sind, ähneln ihre Frequenzverläufe eher denen von *Peak-Filtern*. Dies ist der Grund dafür, dass selbst bei identischen Gains das resultierende Vorfilter nicht den gewünschten vollständig glatten Frequenzverlauf hat. Wie bereits

von Holters in [Hol09] diskutiert, resultieren Filter höherer Ordnung jedoch in ungewollten Phasenmodulationen im Bereich der Bandgrenzen. Die Abwägung zwischen besserer Bandtrennung und der sonst bei höheren Ordnungen entstehenden Verzerrung führte daher zu der eher geringen Ordnung. Wie sich im Folgenden zeigen wird, verhindern die nicht idealen Verläufe der Shelving-Filter es aber nicht, eine ausreichende Rauschformung zu erreichen.

Der zweite funktionale Block der adaptiven Vor- und Nachfilterung beinhaltet die Schätzung der Gains für die Dynamikkompression. Der Teil des Flussgraphen selbst kann in die einzelnen Abschnitte Bandpassfilterung, Leistungsschätzung und Abbildung auf den finalen Gain-Wert unterteilt werden. Die Schätzung der Bandleistungen $p_k(n)$ erfolgt mit der Hilfe von rekursiven Filtern erster Ordnung über

$$p_k(n) = (1 - \lambda_p)p_k(n - 1) + \lambda_p \left(\frac{\pi}{\omega_{w,k}} \tilde{x}_{j,\text{bpk}}^2(n) + p_{\min} \right), \quad (3.35)$$

wobei die kleine Konstante p_{\min} verhindert, dass die geschätzte Leistung $p_k(n)$ vollständig auf Null abfallen kann. In der in dieser Arbeit erfolgten Modifikation zur Anwendung auf den Subband-Fall sind die Eingangssignale dieser Filter die Energien der überabgetasteten und bandpassgefilterten Subband-Signale $\tilde{x}_{j,\text{bpk}}(n)$, wobei eine Normalisierung auf die jeweilige Bandbreite vorgenommen wird. Die Bandpässe $\check{H}_{\text{bpk}}(z)$, die zur Filterung der überabgetasteten Rekonstruktionssignale $\tilde{x}_i(n)$ verwendet werden, resultieren aus einer bilinearen Transformation eines Butterworth Prototypen und wurden ebenfalls über eine Allpastransformation frequenzverschoben. Die Zuordnung der Subband-Rekonstruktionssignale zu den Bandpässen erfolgt derart, dass die oberen und unteren Grenzfrequenzen der Bandpässe $\check{H}_{\text{bpk}}(z)$ zum größten Teil innerhalb der Frequenzbandbreite des Analysefilters $H_i(z)$ der Analysefilterbank liegen.

Die Summe der geschätzten Bandleistungen wird zusätzlich zum Berechnen der mittleren Bandleistung über

$$\bar{p}(n) = \frac{1}{K} \sum_{k=1}^K p_k(n) \quad (3.36)$$

genutzt. Das finale Abbilden auf die signaladaptiven Gain-Werte $g_k(n)$ zur Dynamikkompression erfolgt durch Normalisierung von $p_k(n)$ mit $\bar{p}(n)$ und die Nutzung einer Exponentialkurve

$$g_k(n) = \left(\frac{p_k(n)}{\bar{p}(n)} \right)^{\frac{\nu-1}{2\nu}}, \quad (3.37)$$

die zum Einstellen ihrer Steilheit über einen Faktor $0 < \nu \leq 1$ parametrisiert wird. Abbildung 3.28 zeigt Beispiele der Kurvenverläufe für verschiedene Werte von ν . Die Formulierung der Shelving-Filter $\check{H}_k(z)$ erlaubt die exakte Inversion der Übertragungsfunktion durch das Verwenden der inversen Verstärkungsfaktoren $g_k(n)$ für die Adaption der Filterkoeffizienten des Nachfilters.

Zur Evaluation des Einflusses der modifizierten adaptiven Vor- und Nachfilterung auf

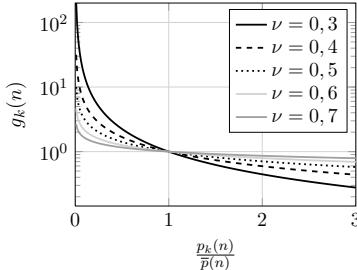


Abbildung 3.28: Beispiele der Exponentialkurvenverläufe für verschiedene Werte des Parametrisierungsfaktors ν .

die Audioqualität des Subband-Codecs wurden im Rahmen der Untersuchungen zu [PWO15] verschiedenste Experimente durchgeführt. Hierbei erfolgte eine Optimierung der Parameter des Codecs sowie der adaptiven Vor- und Nachfilterung gemäß der in Kapitel 4 präsentierten Methodik.

Die evaluierten Fälle sind dabei (jeweils mit eigenem optimierten Parametersatz):

- Fall 1: Keine Vor- und Nachfilterung
- Fall 2: Vor- und Nachfilterung (Gain-Schätzung auf Basis von $x(n)$)
- Fall 3: Vor- und Nachfilterung (Gain-Schätzung auf Basis von $\tilde{x}_i(m)$)
- Fall 4: Ergebnisse der dynamischen Bit-Allokation aus [PO15a].

Der Subband-Codec mit der in Kapitel 2 präsentierten 5-Kanal Filterbank wurde dabei mit einer statischen Verteilung der Bits auf die Subbänder von $\{6,5,3,3,3\}$ betrieben. Die Verwendung von mindestens 3 Bit in den oberen Bändern begründet sich wiederum durch den Anspruch einen Vollband-Audiocodec zu entwerfen und somit die statische Abschaltung oberer Bänder nicht zu gestatten. Insgesamt resultiert die Bitverteilung damit in einer mittleren Codewortlänge von 4 Bits/Sample. Der Vergleich zu der Gain-Schätzung auf Basis des Eingangssignales $x(n)$ des Subband-Encoders erfolgte zusätzlich, um der ursprünglichen Art der Leistungsmessung aus [Hol09] möglichst nahe zu kommen und damit eine Abschätzung des Gewinnes durch die adaptive Vor- und Nachfilterung im besten Fall zu bekommen. Da das Signal $x(n)$ jedoch im Decoder nicht vorliegt, schließt sich diese Art der Adaption allerdings in einem echten Übertragungssystem aus. Die Ergebnisse aus [PO15a] werden zusätzlich angegeben, um einen Vergleich zu dem, durch die dynamische Bit-Allokation zu erreichenden, Einfluss auf die Audioqualität zu ermöglichen.

Tabelle 3.3 enthält die PEAQ-ODG Ergebnisse (vgl. Abschnitt 4.1) der Evaluation eines Testsets, das häufig z.B. im Rahmen der Standardisierung verwendet wird, da es eine gute Mischung aus Sprach- und Musiksignalen sowie kritische Signale zum Provozieren von Codierartefakten enthält. Obwohl für manche der Signale eine kleine Verschlechterung in der ODG auftritt, zeigen die Ergebnisse der besonders kritischen Signale „Kastagnetten“, „Glockenspiel“ und „Trompete“ die deutliche Verbesserung

der Audioqualität, die durch die adaptive Vor- und Nachfilterung in Kombination mit der globalen Parameteroptimierung erzielt werden kann. Im Mittel beträgt die Verbesserung von Fall 1 zu Fall 3 ungefähr 0,15 Punkte der PEAQ-Skala. Zusätzlich ist kaum ein Unterschied zwischen den Fällen 2 und 3 zu erkennen, was zeigt, dass die vorgeschlagene Gain-Schätzung aus den rekonstruierten Subband-Signalen sehr ähnlich zu der aus dem breitbandigen Signal ist.

Tabelle 3.3: Ergebnisse einer Evaluation der objektiven Audioqualität verschiedener Signalklassen für unterschiedliche Codecversionen (aus [PWO15]).

Signalname	PEAQ-ODG			
	Fall 1	Fall 2	Fall 3	Fall 4
Gesang/S. Vega	-0,51	-0,52	-0,52	-0,46
Sprache männl. Eng.	-0,56	-0,51	-0,56	-0,56
Sprache weibl. Eng.	-0,70	-0,58	-0,57	-0,50
Trompete	-0,99	-0,25	-0,37	-0,43
Orchester	-0,46	-0,26	-0,52	-0,54
Big Band	-0,35	-0,27	-0,28	-0,40
Cembalo	-0,31	-0,46	-0,44	-0,32
Kastagnetten	-1,32	-0,72	-0,76	-0,93
Stimmpfeife	-0,48	-0,48	-0,49	-0,44
Dudelsack	-0,28	-0,23	-0,38	-0,46
Glockenspiel	-1,71	-0,76	-0,74	-0,82
Gezupfte Saiten	-0,39	-0,60	-0,56	-0,51
Mittelwert	-0,67	-0,47	-0,52	-0,53

Die Ergebnisse verdeutlichen, dass sich mit der adaptiven Vor- und Nachfilterung vergleichbare Ergebnisse zur dynamischen Bit-Allokation erzielen lassen, ohne dass eine Signalisierung zwischen Encoder und Decoder erfolgen muss. Prinzipiell ist die Algorithmik zur Adaption dabei auch ohne weitere Maßnahmen robust gegenüber Kanalübertragungsfehlern und damit in einem echten System anwendbar, da in der Gain-Schätzung im Decoder keine Rückkopplung stattfindet und die Synchronisation der Rekonstruktionssignale $\tilde{x}_i(m)$ sichergestellt ist. Dennoch ist der Aufwand zur Gain-Schätzung und Filteradaption, trotz der effizienten Realisierbarkeit der Shelving- und Bandpassfilter und deren geringer Ordnung, aufgrund der hohen Bandanzahl nicht unerheblich und steht in einem ungünstigen Verhältnis zur, im Vergleich zur dynamischen Bit-Allokation, eingesparten Datenrate für die Signalisierung. Demnach wird die adaptive Vor- und Nachfilterung im Folgenden und auch in der finalen Evaluierung des in dieser Arbeit entwickelten Audiocodecs nicht weiter betrachtet, zumal die zusätzlichen Gewinne bei der prinzipiell möglichen Kombination von Bit-Allokation und Vor- und Nachfilterung als gering eingeschätzt werden.

Kapitel

Audiocodec Optimierung

4

In diesem Kapitel wird die Suche nach geeigneten Codecparametern über ein Verfahren zur globalen Optimierung beschrieben. Dieses basiert auf einer Methode zum Abschätzen der resultierenden Audioqualität des Codecs für einen potentiellen Parametersatz und die Signale eines gegebenen Testsets sowie dem Abbilden der Einzelergebnisse auf einen Gesamtkostenwert.

Abschnitt 4.1 gibt eine kurze Einführung in Audiocodec Bewertungsgrößen und stellt mit der *Perceptual Evaluation of Audio Quality* (PEAQ) den verwendeten Algorithmus zum Ermitteln der geschätzten Audioqualität vor.

Im Abschnitt 4.2 wird das Vorgehen bei der Optimierung präsentiert. Es werden Algorithmen zur metaheuristischen und stochastischen globalen Suche betrachtet und in Bezug auf ihre Eignung für das vorliegende Minimierungsproblem erörtert.

In einem letzten Abschnitt 4.3 wird das Vorgehen bei der Anwendung des Verfahrens zur Audiocodecoptimierung beschrieben und die erzielten Ergebnisse durchgeföhrter Optimierungen exemplarisch dargestellt sowie diese abschließend diskutiert.

4.1 Audiocodec Bewertungsgrößen

Zur Bewertung der Güte von verlustbehafteten Audiocodierungsverfahren können verschiedenste Maße zur Anwendung kommen. Diese unterscheiden sich grundsätzlich in der Tatsache, ob auch die psychoakustische Wahrnehmbarkeit von Codierfehlern Berücksichtigung findet.

So kann über das aus dem Eingangs- und Ausgangssignal des Codecs berechnete SNR eine Aussage über die Abweichungen der codierten Wellenform vom Originalsignal erlangt werden. Wird das SNR durch das so genannte *segmentelle SNR* (segSNR) in Blöcken berechnet und über der Zeit aufgetragen, so lassen sich auch Erkenntnisse über signalabhängige Veränderungen des SNRs gewinnen. Während der Entwicklung von Codecbausteinen, wie z.B. beim Vergleich verschiedener Prädiktorstrukturen, stellen diese Größen damit wertvolle Bewertungsgrundlagen bereit. Für die Evaluierung und Optimierung eines gesamten Audiocodierungsverfahrens ist es allerdings von Vorteil, auch psychoakustische Effekte zu berücksichtigen und damit eine Aussage über die tatsächliche Wahrnehmbarkeit bestehender Codierungsfehler

treffen zu können.

Keiler benutzt daher in [Kei06] unter anderem auch das so genannte *Noise to Mask Ratio* (NMR). Zur Berechnung dieser Größe wird aus dem Eingangssignal über eine gefensterte *Fast Fourier Transform* (FFT) und weitere nachgeschaltete Verarbeitungsschritte blockweise die Signalleistung in den Frequenzgruppen der menschlichen Wahrnehmung berechnet und damit die Maskierungsschwelle geschätzt. Dies erfolgt unter Berücksichtigung der Ruhöhörschwelle, der Tonalität des Eingangssignales sowie der bandübergreifenden Maskierung. Mit dem aus dem Originalsignal und dem Ausgangssignal des Codecs berechneten Codierfehler wird über eine weitere FFT-basierte Verarbeitung die Fehlerleitung in Frequenzgruppen bestimmt. Indem diese mit der Maskierungsschwelle verglichen wird, kann ermittelt werden, ob die Abweichungen vom Originalsignal theoretisch hörbar werden. Auf Basis der damit erhaltenen Ergebnisse erfolgt eine Auswertung der Mittel- und Maximalwerte sowie Varianzen von den Überschreitungen der Maskierungsschwelle und der Anzahl an Blöcken mit deren Auftreten für eine gegebene Signalsequenz.

Über diese Vorgehen lässt sich eine Aussage über die Wahrnehmbarkeit von Codierfehlern treffen. Allerdings erfolgt keine Bewertung auftretender Artefakte hinsichtlich ihres Einflusses auf die global wahrgenommene Audioqualität. Außerdem muss die Betrachtung der Ergebnisse für ein gegebenes Signal immer anhand mehrerer Messgrößen durchgeführt werden.

Eine Methode, die das Abschätzen der Leistungsfähigkeit eines Codecs für ein gegebenes Eingangssignal über nur einen einzigen Messwert erlaubt, ist der bereits erwähnte PEAQ-Algorithmus. Dieses von der ITU im Standard BS.1387 [ITU01] beschriebene Verfahren zur Evaluation der wahrgenommenen Audioqualität stellt neben dem so genannten *Distortion Index* (DI)¹ diverse *Model Output Variables* (MOVs) an seinem Ausgang bereit. Die direkte Auswertung und Interpretation der MOVs ist jedoch nur für erfahrene Nutzer bzw. bei dem Nachvollziehen konkreter Unzulänglichkeiten des untersuchten Audiocodecs sinnvoll. Daher erfolgt zudem eine Abbildung auf die so genannte *Objective Difference Grade* (ODG), die das Ergebnis eines durchgeführten subjektiven Hörversuches simulieren soll.

Abbildung 4.1 (a) fasst den Aufbau² des PEAQ-Algorithmus gemäß [TTB⁺98] grob in einem Flussgraphen zusammen. In Teil (b) der Graphik ist die Abbildung der ODG-Werte auf die Beschreibungen zur Wahrnehmbarkeit des Codierfehlers aus der *Subjective Difference Grade* (SDG), also der subjektiven Wertung des Unterschiedes zu dem Referenzsignal, festgehalten [ITU01].

Am Eingang des PEAQ-Algorithmus wird das zu bewertende und das Originalsignal übergeben. Die beiden Signale müssen dabei zeitlich und in ihrem Pegel aneinander angeglichen sein. Über ein *Peripheres Gehörmodell* erfolgt eine Verarbeitung in den Frequenzgruppen menschlicher Wahrnehmung und die Berechnung von Maskierungsschwellen. Dieser Teil des PEAQ-Algorithmus basiert in den Grundzügen auf einer ähnlichen Vorverarbeitung, wie sie auch schon für die Berechnung des NMR

¹Der DI ist für sich schon ein guter globaler Qualitätsindikator, wird aber bei dem Vergleich von Codierungsverfahren mit hoher Audioqualität in der Literatur nur selten verwendet.

²In [TTB⁺98] sind weitere Details zum Aufbau und dem Vorgehen zusammengefasst.

nach Keiler beschrieben wurde. Zusätzlich werden aber auch noch diverse weitere kognitive Eigenschaften, lineare sowie nichtlineare Verzerrungen und spektrale Eigenschaften des Codierfehlersignales ermittelt.

Die Struktur des Peripheren Gehörmodells wurde von der ITU in zwei Ausführungen spezifiziert, woraus sich eine Unterscheidung in die *Basic-* und *Advanced-Version* ergibt. Erstere sieht eine FFT-basierte Verarbeitung vor. Bei letzterer kommt eine Filterbankstruktur zum Einsatz, in deren Signalverarbeitung implizit auch kognitive Effekte abgebildet werden.

Die Ausgangssignale des Gehörmodells werden in einem weiteren Block zur Berechnung der Qualitätsparameter über ein kognitives Modell verwendet. Insgesamt stehen bei der Basic-Version an dessen Ausgang 11 MOVs zur Verfügung, die in die globalen Qualitätsparameter eingehen. Die Abbildung der MOVs auf die ODG erfolgt dabei über ein *neuronales Netz*, das mit Daten aus Hörversuchen zum Codec-Vergleich für über 600 codierte Musikstücke trainiert wurde.

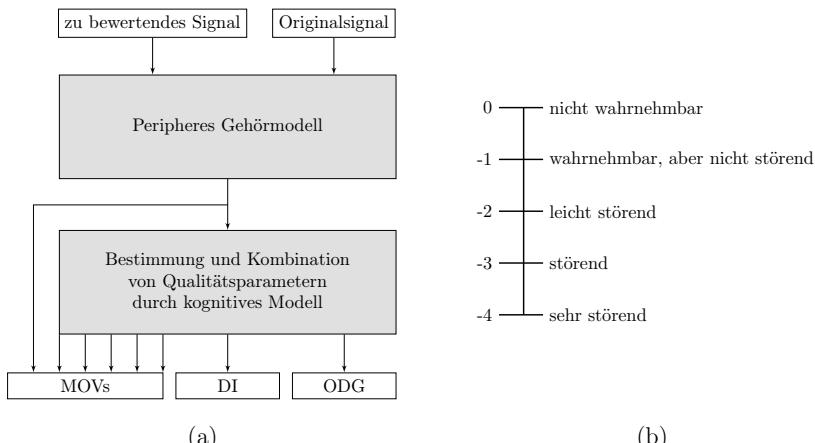


Abbildung 4.1: PEAQ-Detaile: (a) Flussgraph des übergeordneten Aufbaus des PEAQ-Algorithmus (gemäß [TTB⁺98]) (b) Zuordnung der PEAQ-ODG Werte zu Beschreibungen der Wahrnehmbarkeit des Codierfehlers aus der SDG.

Wie bereits erwähnt, wurde das PEAQ-Verfahren von der ITU im Rahmen eines Standards spezifiziert. Dieser enthält auch Bedingungen für eine Überprüfung der Konformität anhand der Ergebnisse für ein definiertes Testset. Laut [Kab03] und [HZ15] enthalten die Beschreibungen in [ITU01] jedoch diverse Inkonsistenzen und Mehrdeutigkeiten, so dass eine konforme Implementierung erschwert wird. Die genaue Einhaltung der Konformität bleibt damit den an der Entwicklung beteiligten Firmen vorbehalten.

Dennoch gibt es mehrere quelloffene Implementierungen des PEAQ-Algorithmus

[Kab03, Got03, HZ15], die für die meisten praktischen Zwecke ausreichend nah an den von der ITU vorgegebenen Referenzwerten liegen (vgl. [HZ15]). Mit dem kürzlich veröffentlichten *GstPEAQ* [HZ15] wurde zudem eine weitere Implementierung veröffentlicht, die sogar die Advanced-Version beinhaltet. *GstPEAQ* ist allerdings in beiden Versionen ebenfalls nicht standardkonform und zum Zeitpunkt des Verfassens lagen dem Autor dieser Arbeit noch keine Langzeiterfahrungen hierzu vor.

Solange die PEAQ-Implementierungen im Wesentlichen für relative Vergleiche verwendet werden, wird davon ausgegangen, dass eine nicht vollständig erreichte Konformität zum Standard nur geringe Einflüsse hat. Im Rahmen dieser Arbeit wurde, unter Abwägung vorhandener Erfahrungswerte und zur Vergleichbarkeit mit vorherigen Veröffentlichungen, an der Implementierung *PQevalAudio* aus [Kab03] festgehalten. Diese erlaubt lediglich die Nutzung der Basic-Version und wurde mit Standardparametern ausgeführt. Die aus den Quelldateien erstellte Software wurde über die von der ITU veröffentlichten Signale abgeglichen, wobei sich die Ergebnisse der systematischen Untersuchung aus [HZ15] bestätigten. Hiernach liegen die Ergebnisse des DI für die Mehrzahl der Signale sehr nah an der Referenz. Lediglich für eines der 16 Signale ergibt sich, wie auch für die anderen quelloffenen Implementierungen, eine Abweichung des DI von ca. 0,6 Punkten.

Als letzte und im Zweifel zu bevorzugende Möglichkeit zur Bewertung der Audioqualität von Audiocodierungsverfahren ist die eines subjektiven Hörversuches zu nennen. Eine Verifikation der Ergebnisse einer PEAQ-Evaluation ist für belastbare Aussagen unumgänglich. Allerdings ist der Aufwand für die Durchführung subjektiver Tests sehr hoch, was deren Durchführung für Zwischenversionen des Codecs schwierig gestaltete. Für die Entwicklung und Optimierung im Rahmen dieser Arbeit wurden daher PEAQ-Ergebnisse sowie eine Bestätigung durch informelle Tests verwendet. Ein Hörversuch wurde lediglich zur Evaluierung der im Rahmen dieser Arbeit finalen Version des Codecs durchgeführt.

4.2 Methode zur globalen Parameteroptimierung

Das in dieser Arbeit vorgestellte Codierungsverfahren beinhaltet eine Vielzahl von Parametern, die zum Teil in schwer kontrollierbarer Wechselwirkung zueinander stehen und deren Einfluss stark vom zu codierenden Signal abhängt. Dies macht ein manuelles Einstellen für eine vollständige Optimierung schwierig.

Allerdings hat sich bereits in den Arbeiten zu [HZ09] herausgestellt, dass sich mit Hilfe des PEAQ-Algorithmus und Methoden der globalen Suche ein sinnvolles Einstellen von Audiocodecparametern vornehmen lässt. Die Idee einer PEAQ-basierten Optimierung wurde daher in dieser Arbeit aufgegriffen und in einer modifizierten Form umgesetzt.

Demnach wird im Folgenden das entwickelte Optimierungsverfahren beschrieben, die qualitativen Ergebnisse der Parameteroptimierung vorgestellt sowie das dazu nötige Vorgehen und das Konvergenzverhalten diskutiert.

4.2.1 PEAQ-basierte Optimierung

Die Realisierung der PEAQ-basierten Optimierung erfolgte im Rahmen dieser Arbeit gemäß des in Abbildung 4.2 dargestellten Flussgraphens.

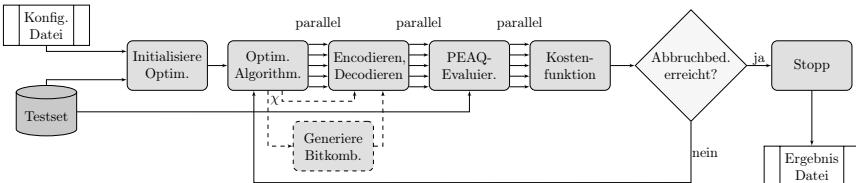


Abbildung 4.2: Flussgraph des PEAQ-basierten Optimierungsverfahrens inkl. optionaler Generierung erlaubter Bitkombinationen der dBA (gestrichelter Block).

Für ein gegebenes Testset und eine eingestellte Konfiguration des Codecs wird demnach eine Initialisierung der Optimierung vorgenommen. Über den verwendeten Optimierungsalgorithmus wird nun das parallele Encodieren und Decodieren für verschiedene Parametersätze inklusive der nötigen Latenzkompensation sowie eine nachgeschaltete PEAQ-Evaluation gesteuert.

Das Abbilden der einzelnen ODG-Werte der PEAQ-Evaluationen für einen gegebenen Parametervektor χ auf einen Kostenwert $K(\chi)$ geschieht für ein Testset mit einer Anzahl an Signalen T , wie in [Hol09] vorgeschlagen über

$$K(\chi) = \sum_{i=1}^T (\text{ODG}_i(\chi))^4, \quad (4.1)$$

wobei eine Begrenzung der ODGs auf Werte kleiner Null stattfindet³. Hierbei gehen die einzelnen Werte mit der vierten Potenz in die Summe ein, um ein schlechteres Abschneiden bei der Evaluierung der Audioqualität mit dem PEAQ-Verfahren deutlich stärker zu bestrafen. Die Optimierung endet, sobald ein Abbruchkriterium des Optimierungsalgorithmus erreicht wurde.

Mit dem soeben erläuterten Optimierungsverfahren wurden in dieser Arbeit sowie im Zuge von [PO15a] und [PWO15] diverse Optimierungen mit verschiedensten Codecvarianten, Optimierungszielen und Testsets gemacht. Daraus ergaben sich auch optionale Ergänzungen, wie der in Abbildung 4.2 dargestellte Block zum Generieren erlaubter Bitkombinationen der dBA. Die vorgenommene starke Parallelisierung des Encodierens und Decodierens sowie Evaluierens einzelner Signale erlaubt eine Optimierung auch für große Testsets bei dennoch annehmbaren Laufzeiten. Damit ist es möglich, das Vermeiden einer Überanpassung an das Testset sicherzustellen, sofern dieses eine Vielzahl unterschiedlicher Signal- bzw. Instrumentklassen enthält.

³Der PEAQ-Algorithmus sieht trotz der Definition eines Wertebereiches der ODG von 0 bis -4 keine Begrenzung des Ergebnisses einer Evaluation vor und erzeugt für identische Eingangssignale Werte von $\approx 0,2$.

4.2.2 Algorithmen zur globalen Parameteroptimierung

Zur Lösung des vorliegenden Optimierungsproblems können verschiedenste Methoden zur globalen Minimierung verwendet werden, die keine explizite Berechnung eines Gradienten benötigen. Grundsätzlich können also die meisten der Verfahren zur metaheuristischen Optimierung, wie der bereits in Abschnitt 2.4 benutzte Pattern-Search Algorithmus, Verwendung finden. Aufgrund der Größe des Suchraumes und zu erwartender starker Unstetigkeiten der Kostenfunktion gestaltet sich eine Optimierung auf Basis rein deterministischer Suchverfahren jedoch als schwierig.

In [Hol09] wird daher vorgeschlagen, mit einer modifizierten Version des *Simulated Annealing*-Ansatzes [KGJV83] ein Verfahren zur stochastischen globalen Optimierung zu verwenden und dieses um eine nachgeschaltete Verfeinerung des Optimierungsergebnisses durch die *Rosenbrock-Methode* [Ros60] zu ergänzen. Bei letzterem handelt es sich um eine Methode zur lokalen Suche, die wie der Pattern-Search Algorithmus ein systematisches Evaluieren des Parameterraumes vorsieht, jedoch an dieser Stelle nicht weiter betrachtet werden soll.

Simulated Annealing ist ein Optimierungsansatz, der an die physikalischen Prozesse beim Abkühlen metallischer Werkstoffe angelehnt ist. Hierbei beschreibt der Parametervektor gemäß der Analogie den aktuellen atomaren Zustand und dieser wird gemäß der Vorzugsrichtungen des Systems quasi zufällig variiert. Der zum jeweiligen Systemzustand gehörende Kostenwert bildet die Energie des Systems nach, das das Bestreben hat, einen energetisch niedrigen Zustand anzunehmen. In jeder Iteration wird demnach ein neuer Systemzustand erzeugt, die Energie ermittelt und über die Annahme des Zustandes entschieden. Dabei werden gelegentlich zwar auch temporäre Vergrößerungen der Energie akzeptiert, die Wahrscheinlichkeit von deren Akzeptanz wird jedoch als Verlauf der Temperatur in einem so genannten Abkühlungsschema über der Zeit verringert.

Durch dieses Vorgehen wird eine globale Suche ermöglicht, die über der Zeit eine Konvergenz in die Nähe eines lokalen oder des globalen Minimums vollzieht. Stellgrößen, die den Verlauf der Optimierung beim Simulated-Annealing wesentlich beeinflussen sind das gewählte Abkühlungsschema sowie die Methodik zum Variieren des Parametervektors. Die größte Schwierigkeit besteht in der zu treffenden Abwägung zwischen Sicherheit und Dauer der Konvergenz, die sich bei der Auslegung des Abkühlungsschemas ergibt. Es kann daher bei ungünstigen Konstellationen und Startpunkten der Optimierung dazu kommen, dass das Verfahren in einem lokalen Minimum stecken bleibt. Holters macht in [Hol09] wenig Aussagen zum generellen Konvergenzverhalten und wie stark die Abweichungen der Ergebnisse einzelner Optimierungsdurchläufe sind. Experimente mit Simulated Annealing im Rahmen dieser Arbeit zeigten, dass das Verfahren im vorliegenden Anwendungsfall grundsätzlich gute Optimierungsergebnisse hervorbringt, jedoch eine große Varianz des finalen Kostenwertes aufweist.

Daher wurde im Zuge der Untersuchungen zu [PO15a] mit dem so genannten *Genetic Algorithm* [Gol89] ein alternatives Verfahren zur stochastischen globalen Suche eingesetzt. Dieser an den Mechanismus der Natürlichen Selektion bei der Evolution

angelehnte Algorithmus hat sich schon in vielen Anwendungen als geeignet für die Optimierung von komplexen Optimierungsaufgaben mit einer hohen Zahl an Parametern und unbekannter Topologie der Kostenfunktion herausgestellt [GCM⁺13]. Das Grundprinzip des Genetic Algorithm soll im Folgenden kurz erläutert werden. Der zu optimierende Parametervektor wird dabei als Gensequenz interpretiert und der damit einhergehende Kostenwert als Fitness eines hiermit ausgestatteten Individuums bezeichnet. Abbildung 4.3 zeigt einen Flussgraphen des Vorgehens bei der eigentlichen Optimierung. Die Beschreibung des Algorithmus erfolgt unter Verwendung der üblichen Terminologie.

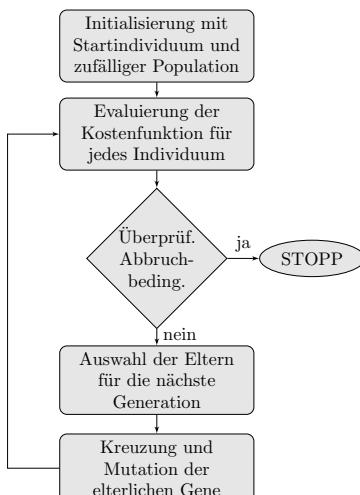


Abbildung 4.3: Flussgraph der Optimierung nach dem Genetic Algorithm.

Demnach erfolgt eine Initialisierung mit einem Startindividuum und einer Population, die innerhalb der vorgegebenen Parametergrenzen zufällig erzeugt wird. Für jedes Individuum wird nun die Kostenfunktion evaluiert und somit die Fitness bestimmt. Es wird überprüft, ob ein Abbruchkriterium, wie z.B. das Unterschreiten einer Schwelle für die Varianz der Fitnesswerte, für die aktuelle Population erreicht wurde. Anhand der Kostenwerte erfolgt nun die Auswahl einer vorgegebenen Zahl an stärksten Individuen, der so genannten Elite. Mit diesen „Überlebenden“ wird durch Kreuzung und Mutation die Population der nächsten Generation erzeugt. Dieses Vorgehen wird für eine vorgegebene Anzahl an Generationen wiederholt, solange das Abbruchkriterium nicht erreicht wird.

Mit dem Genetic Algorithm wurde in dieser Arbeit eine verhältnismäßig sichere und schnelle Konvergenz erzielt. Die Optimierungsergebnisse wiesen hierbei eine relativ geringe Varianz auf und das Verfahren bedurfte keiner manuellen Intervention. Die, z.B. für die Initialisierung der dBA nötigen, Ganzzahlwerte konnten dabei in der verwendeten Implementierung direkt mit optimiert werden. Auf eine nachgeschal-

tete lokale Suche wurde verzichtet, da davon ausgegangen wird, dass die erhaltene Lösung nah genug am globalen oder einem lokalen Optimum liegt.

4.3 Ergebnisse von globalen Optimierungen der Audiocodec Parameter

Zum Ausloten der Grenzen des entwickelten Audiocodierungsverfahrens und für das Erstellen von Testsignalen für den im folgenden Kapitel 5 dokumentierten Hörversuch wurden im Rahmen dieser Arbeit Optimierungen mit verschiedensten Zielbitraten durchgeführt. Die Ergebnisse dieser Optimierungen werden im Folgenden zusammen mit dem Vorgehen bei der Optimierung und dem Konvergenzverhalten kurz qualitativ festgehalten und diskutiert.

Der Subband-Codec wurde dabei mit der global optimierten Filterbank aus Abschnitt 2.4 betrieben. Für die blockbasierte Prädiktoradaption fiel die Wahl auf eine Adaption mittels des Burg-Algorithmus, da dieser garantiert numerisch stabil ist und der Fokus der Untersuchungen nicht auf einer Minimierung des Berechnungsaufwandes lag. Zusätzliche andere Optimierungen mit der Levinson-Durbin Rekursion zeigten allerdings vergleichbare Ergebnisse.

Die innerhalb der Optimierung eingestellten Parameter sind:

- A-Law Parameter und Skalierungsfaktoren zum Entwurf der Quantisierer-Codewörter für jedes Subband,
- Zeitkonstanten für den Ansprech- und Rücklauf-Fall und optional zu optimierende Leakage-Faktoren der Prädiktionsfehlernormalisierung für jedes Subband,
- Koeffizienten des statischen Prädiktors des untersten Subbandes,
- Initialisierungs- und Pegelreduktionsparameter der dBA,
- Pegelschätzungsparameter der dBA

sowie zwei Parameter der blockbasierten Prädiktoradaption, die für alle Bänder gleich sind. Letztere werden in Kapitel 3 nicht weiter beschrieben und haben nur geringen Einfluss auf das Optimierungsergebnis. Insgesamt können damit bis zu 37 Parameter und Koeffizienten gleichzeitig optimiert werden.

Startpunkte der Optimierungen sind Parameterwerte, die sich im Zuge der Codecentwicklung als sinnvoll herausgestellt haben sowie Ergebnisse der in [Hol09] durchgeföhrten Optimierungen. In manchen Optimierungsdurchläufen hat es sich zudem als hilfreich erwiesen, die Initialisierung bewusst mit weniger geeigneten Parametern durchzuführen, um eine vorzeitige Konvergenz in ein lokales Minimum zu verhindern und ein weiter ausgedehntes Absuchen des Parameterraumes zu veranlassen. Als obere und untere Grenzen der zu optimierenden Parameter werden Erfahrungswerte aus der Codeentwicklung bzw. „physikalisch sinnvolle“ Werte verwendet.

Die Anwendung des Genetic Algorithm erfolgt weitestgehend mit den von der genutzten Implementierung vorgegebenen Standardparametern [MW16]. Diese sehen eine Optimierung über 100 Generationen vor, solange keines der Abbruchkriterien erreicht wurde. Die Größe der Population musste jedoch aus Laufzeitgründen auf den geringen Wert 20^4 begrenzt werden. Aus jeder Generation „überleben“ jeweils zwei Individuen, aus deren Genen durch zufälliges Ziehen $[0,8 \cdot 18] = 14$ Kreuzungen für die nächste Generation berechnet werden. Die verbleibenden vier Individuen werden durch zufällige Mutation auf Basis der Elitegene erzeugt.

Abbildung 4.4 zeigt ein Beispiel für den Verlauf des besten und mittleren Kostenwertes für eine Optimierung mit einer Zielbitzahl von 3 Bit/Sample in einer logarithmischen Darstellung.

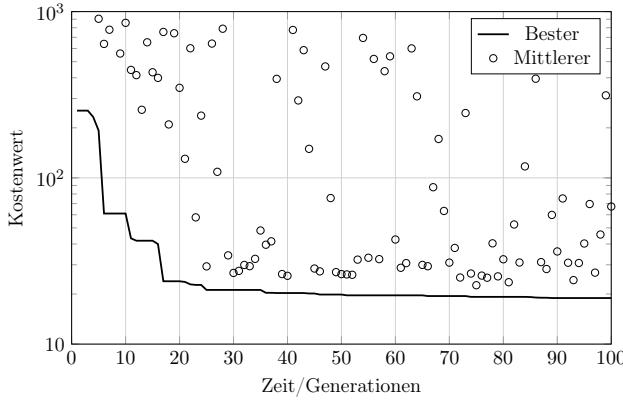


Abbildung 4.4: Beispielverlauf des minimalen und mittleren Kostenwertes bei einer Optimierung mit 3 Bit/Sample.

Hierbei ergibt sich besonders innerhalb der ersten 50 Generationen eine deutliche Verringerung des Kostenwertes. Die dargestellten mittleren Kosten der Individuen jeder Generation verdeutlichen die starke Nichtlinearität und Unstetigkeit der Kostenfunktion. Letztere resultiert aus der Tatsache, dass selbst bei fortgeschritterer Optimierung durch kleine Parameteränderungen innerhalb der Kreuzung und Mutation der Codec in einen ungünstigen Arbeitspunkt geraten kann, was aufgrund der

⁴Diese geringe Populationsgröße schränkt die initiale Suche im Parameterraum ein und erhöht das Risiko in einem lokalen Minimum stecken zu bleiben. Allerdings ist aufgrund des hohen Aufwandes zur Berechnung der Kostenfunktion für die Optimierung auf das gesamte SQAM-Testset eine deutliche Erhöhung der Populationsgröße in die sonst üblichen Wertebereiche von 50-200 für die vorliegende Problemstellung nicht ohne Weiteres praktikabel. Um den Einflüssen der geringen Populationsgröße ein wenig entgegenzuwirken, wurden im Rahmen der Untersuchungen für jede Optimierung einer Konfiguration des Codecs zudem mehrere Durchläufe durchgeführt. Eine im Nachgang zu den in dieser Arbeit dokumentierten Durchläufen durchgeführte Optimierung mit einer Populationsgröße von 200 führte nur auf geringfügig bessere Ergebnisse.

Formulierung der Kostenfunktion zu sehr hohen Kostenwerten führen kann.

Wichtige durch die Optimierung einzustellende Parameter sind die in den Entwurf der Subband-Quantisierer einfließenden A-Law Parameter und Skalierungsfaktoren. Bei ihrer Optimierung gilt es, wie bereits in Kapitel 3 diskutiert, die Abwägung zwischen ausreichend feiner Quantisierung über den gesamten Aussteuerungsbereich und zu vermeidender längerfristiger Übersteuerung zu finden. Abbildung 4.5 zeigt ein Beispiel von skalierten A-Law Verläufen, die innerhalb der Optimierung zum Erstellen von Codebucheinträgen genutzt wurden. Diese zeigen die, aufgrund der geringeren Quantisiererauflösungen zu erwartende, stärkere Krümmung der Kennlinien vom dritten Bandpass und dem Hochpass. Zusätzlich erlaubt die deutlich höhere Anzahl an Quantisierungsintervallen einen etwas größeren Aussteuerungsbereich für das Tiefpass-Band. Zum Teil ergeben sich im Rahmen der Optimierung im Zusammenspiel mit der adaptiven Prädiktionsfehlernormalisierung allerdings auch deutlich höhere Maximalamplituden der Kennlinien.

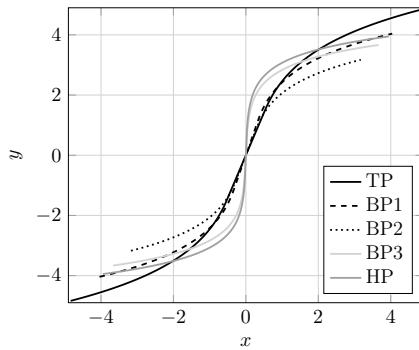


Abbildung 4.5: Beispielverläufe optimierter skalierte A-Law Kennlinien zum Generieren von Codebüchern der Subband-Quantisierer.

Die ebenso in das Verhalten der adaptiven Quantisierung einfließenden Zeitkonstanten der Hüllkurvenschätzung weisen mit Werten zwischen 0,7 und 0,9 für den Ansprech- und 0,1 sowie 0,3 für den Rücklauf-Fall nur geringe Abweichungen zu den in [Hol09] gefundenen Parametern einer Breitband-ADPCM auf. Grundsätzlich wäre besonders aufgrund der Verarbeitung der Subband-Signale im unterabgetasteten Takt eine schnellere Adaption im Ansprech-Fall wünschenswert. Das gewählte Vorgehen und die Parameter erlauben aber auch so eine sehr gute Audioqualität für transiente Signale, wie die Evaluation im folgenden Kapitel 5 zeigen wird.

Ein zusätzlicher in der Optimierung genutzter Freiheitsgrad besteht in der Optimierung der Koeffizienten des statischen Prädiktors. Abbildung 4.6 zeigt ein Beispielergebnis in einer Darstellung als Prädiktionsfehlerfilter. Hierbei bestätigt sich die grundsätzliche Eignung eines Prädiktors mit hohem Prädiktionsgewinn bei tiefen Frequenzen. Manche Optimierungen zeigen zudem eine Tendenz zu einem Einbruch

in der Mitte des Übertragungsbereiches, wie sie auch schon in [Hol09] für die Fehlerrückkopplungsfilter beobachtet wurde. Aufgrund der Verarbeitung unterabgetasteter Signale und dem Einfluss anderer Codecbausteine entziehen sich die Frequenzverläufe in dem vorliegenden Fall jedoch einer direkten psychoakustischen Interpretation.

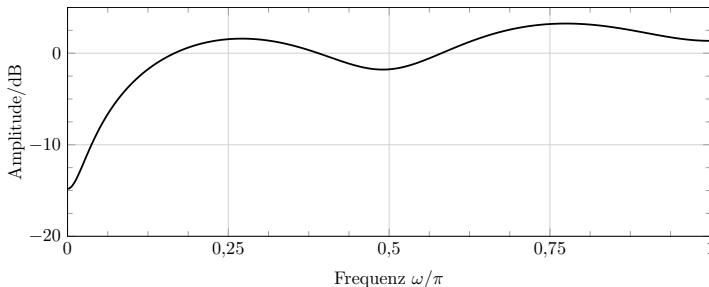


Abbildung 4.6: Beispielbetragsfrequenzgang eines optimierten statischen Prädiktors in der Darstellung als Prädiktionsfehlerfilter.

Als weitere wichtige Parameter sind die Initialisierungs- und Adoptionsparameter der dynamischen Bit-Allokation zu nennen. Die in den verschiedenen Optimierungs-durchläufen erhaltenen Ergebnisse gehen mit den bereits in Kapitel 3 angedeuteten Tendenzen einher. Demnach favorisiert die Optimierung eine Parametrisierung, die für viele der natürlichen Signale dem Tiefpass-Band acht bis zehn Bit zuweist und in einer zeitweisen Abschaltung oberer Bänder resultiert. Für die Initialisierung der Bitzähler der unteren Bänder innerhalb der dynamischen Bit-Allokation ergibt sich in der Regel ein mit steigender Bandzahl abfallender Verlauf. Die Pegelschätzungs-parameter führen zu einer schnellen Adaption bei steigenden und deutlich verlangsamten Anpassung bei fallenden Pegeln. Für die Pegelreduktion pro Iteration der dynamischen Bit-Allokation ergeben sich bei abgeschalteter Abhängigkeit von der Iterationszahl Werte zwischen 3 und 6 dB.

Kapitel

5

Evaluierung

In diesem Kapitel werden die Ergebnisse einer Evaluation des vorgestellten Audio-codierungsverfahrens in Bezug auf die mit ihm erreichbare Audioqualität und den Einfluss von Kanalübertragungsfehlern präsentiert.

Abschnitt 5.1 enthält die, über die im vorherigen Kapitel 4 beschriebene Parameter-optimierung, erzielten Ergebnisse einer PEAQ-Evaluation. Diese werden im Detail vorgestellt und diskutiert und mit den Ergebnissen anderer Verfahren mit geringer Latenz verglichen. Zusätzlich erfolgt die Darstellung der mittleren PEAQ-Bewertung der Codec-Audioqualität über der Datenrate.

Abschnitt 5.2 beinhaltet Details und Erkenntnisse eines durchgeführten Probandenhörversuches zur Bestätigung der, über den PEAQ-Algorithmus ermittelten, Tendenzen zur erreichbaren Audioqualität des entwickelten Audiocodierungsverfahrens für verschiedene Bitraten.

Im Abschnitt 5.3 werden Ergebnisse einer Abschätzung des Einflusses von Kanal-übertragungsfehlern vorgestellt und mit Ergebnissen aus der Literatur verglichen.

5.1 Objektive Ergebnisse

5.1.1 Vorgehen

Für die Optimierung und Evaluation des vorgestellten Audiocodierungsverfahrens wurden zur besseren Vergleichbarkeit und Nachvollziehbarkeit der Ergebnisse die Signale der *Sound Quality Assessment Material recordings* (SQAM) [EBU88] in der mittlerweile frei unter [EBU08] erhältlichen Version verwendet. Dieses Testset deckt ein breites Spektrum an Instrumenten-, Sprach- und Gesangsaufnahmen sowie synthetischen Signalen ab und erlaubt es somit, sich ein umfassendes Bild der Leistungsfähigkeit eines Codierverfahrens zu machen.

Da die Mehrzahl der Signale als Stereo-Aufnahme vorliegt und die Codierung keine spezielle Verarbeitung dieser vorsieht, wurde jeweils nur der linke Kanal der Signale verwendet. Dies erfolgt auch, um eine mögliche positive Beeinflussung der PEAQ-Ergebnisse, die durch Artefakte im Originalsignal, welche durch eventuelles „Downmixing“ entstehen können, auszuschließen. Die Signale wurden in voller Länge und ohne Vorverarbeitung durch Normalisierung genutzt. Da das SQAM-Testset ur-

sprünglich als mit 16 Bit linear „Pulse-Code-modulierte“ CD-Version veröffentlicht wurde, fand die Verarbeitung mit einer Auflösung von 16 Bit statt, auch wenn der Codec von seinem Dynamikumfang her ebenfalls eine Ein- und Ausgangsauflösung von 24 Bit unterstützt.

Für die Ermittlung der PEAQ-Ergebnisse wurde, wie bereits in Kapitel 4 diskutiert, die Implementierung PQevalAudio gemäß [Kab03] verwendet. Die Benutzung des PEAQ-Algorithmus macht es dabei nötig, dass die Referenz und das codierte Signal in einer Abtastrate von 48 kHz vorliegen. Die Signale des SQAM-Testsets wurden daher vor der Codierung einer Abtastratenerhöhung unterzogen, um auch in diesem Fall die Beeinflussung der PEAQ-Ergebnisse zu vermeiden.

Der Codec wurde zum Beschleunigen der Berechnungen in der Programmiersprache C implementiert und als so genanntes „MEX-file“ einer über das Softwarepaket Matlab gesteuerten globalen Optimierung und PEAQ-Evaluation zugänglich gemacht. Obwohl die Codec-Implementierung eine Berechnung in einfacher Genauigkeit erlaubt, die für Audiosignale mit 16 Bit Auflösung nur geringe Unterschiede in der Audioqualität zu doppelter Genauigkeit aufweist, wurden die Simulationen mit letzterer durchgeführt, um die Grenzen des Verfahrens auszuloten. Auf ein nachgeschaltetes „Dithering“ wurde verzichtet, da zu dessen Einfluss auf die wahrgenommene Audioqualität des Codierverfahrens wenig Erkenntnisse vorlagen und um die reine Rekonstruktionsqualität des Codecs zu erfassen. Von einer beim „maschinellen Lernen“ üblichen Aufteilung in ein Trainings- und Testset wurde bei den durchgeführten Optimierungen und Evaluierungen wie in [Hol09] abgesehen, da angenommen wird, dass dieses groß und divers genug ist, um eine Überanpassung zu verhindern.

Zum Vergleich mit Ergebnissen von Verfahren zur Codierung mit geringer Latenz aus der aktuellen Literatur wurden die Ansätze aus [SHZ13a] und [SHZ14] implementiert und evaluiert. Diese basieren auf einer Breitband-ADPCM mit einer Prädiktoraadaptation über einen fehlerrobusten GAL-Algorithmus (vgl. Abschnitt 3.2.2). Sie ergeben bei entsprechender Parametrisierung die Codec-Version aus [Hol09], was den Abgleich mit den dort präsentierten detaillierteren Ergebnissen ermöglicht. Diese konnten ebenso wie die Ergebnisse aus [SHZ14] mit der vorliegenden Implementierung im Rahmen der verfügbaren Informationen bestätigt werden. Bei den Erweiterungen zum Erreichen bzw. Verbessern der Fehlerrobustheit wurde auf die Übertragung grob quantisierter Normalisierungsfaktoren der adaptiven Quantisierung (vgl. Abschnitt 3.3.2) verzichtet, da diese die Codec-Datenrate signifikant erhöht und, wenn überhaupt, nur einen geringen bzw. sogar negativen Einfluss auf die Audioqualität im fehlerlosen Übertragungsfall hat. Dies erfolgte, um den Vergleich bzgl. der verwendeten Datenrate und damit erreichbaren Audioqualität so fair wie möglich ausfallen zu lassen. Der Codec wurde zur besseren Vergleichbarkeit bei einer Abtastrate von 48 kHz betrieben.

5.1.2 Ergebnisse

Im Folgenden werden die Ergebnisse der PEAQ-basierten Evaluation präsentiert. Die Codierung erfolgte unter Verwendung der dynamischen Bit-Allokation mit einer mittleren Bitzahl von 4 Bit/Sample und 3 Bit/Sample, was in Nettodatenraten von 144 kbit/s bzw. 192 kbit/s resultiert. Das in Abschnitt 3.3.3 vorgestellte Abbilden der Bitkombinationen auf eine vorgegebene und begrenzte Menge fand hierbei aufgrund des zu hohen Aufwandes bei der Optimierung des gesamten SQAM-Testsets keine Anwendung. Analysen der Histogramme auftretender Bitkombinationen für das SQAM-Testset und den jeweiligen Parametervektor zeigten jedoch, dass sich die, in Abschnitt 3.3.3 formulierten, Beobachtungen bestätigen, wonach für natürliche Audiosignale nicht alle Bitkombinationen auch wirklich auftreten und die Auftretenshäufigkeit nach den häufigsten 10 bis 20 stark abnimmt. Daher kann davon ausgegangen werden, dass die Bit-Allokation mit vernachlässigbarem Mehraufwand inklusive eines Fehlerschutzes ohne negativen Einfluss auf die Audioqualität übertragen werden kann. Zusätzlich wurde für das Tiefpass-Band mit jedem Rahmenkopf¹ ein mit 12 Bit logarithmisch quantisierter Prädiktionsfehlernormalisierungsfaktor im Encoder verwendet und zum Decoder übertragen, um dessen Einfluss auch schon im fehlerlosen Fall zu berücksichtigen. Dieses Vorgehen resultiert ebenso in einer zu vernachlässigenden Erhöhung der Bitzahl von in diesem Fall 0,075 Bit/Sample.

Abbildung 5.1 zeigt die Ergebnisse einer Codierung und Evaluation des SQAM-Testsets für 4 Bit/Sample. Die Verbindungen zwischen den Ergebnissen eines Signales für jedes der Verfahren verdeutlichen über gestrichelte oder durchgezogene Linien eine bessere oder schlechtere Bewertung des Subband-Codecs im Vergleich zum Referenzverfahren gemäß [SHZ14].

Es ist zu erkennen, dass die Ergebnisse für viele der Signale relativ nah aneinander und zumeist oberhalb der „-0,5 Punkte Linie“ liegen. Lediglich für einige wenige der Signale aus der Klasse der Holz- und Blechblasinstrumente (17, 23) sowie der Rhythmussektion (26, 29, 30) weist die PEAQ-ODG Bewertung deutliche Unterschiede zugunsten der Breitband-Codierung auf. Das schlechtere Abschneiden des Subband-Codecs bei transienteren Signalen liegt vermutlich an einer, aufgrund der Verarbeitung der unterabgetasteten Subband-Signale, zu langsamen Anpassung an wechselnde Signalamplituden bei der adaptiven Quantisierung. Der deutliche Ausreißer im Ergebnis des Codecs gemäß [SHZ14] für das synthetische Signal „06 Elektr. Gong 500 Hz vibrato“ ist dem Anschein nach Resultat eines nicht ausreichend hohen Prädiktionsgewinnes aufgrund der eingeführten Maßnahmen zum Erreichen der Coderebustheit. Dieses spielt jedoch eine untergeordnete Rolle für die Bewertung der Leistungsfähigkeit der Verfahren. Bei der Codierung eines der kritischsten SQAM-Signale, dem Signal „35 Glockenspiel“, erreichen beide Codierverfahren ihren schlechtesten ODG-Wert für natürliches Audio, der jedoch noch in einem Bereich für die ODG liegt, in dem die Unterschiede zum Originalsignal nur für geübte Hörer erkennbar sind.

¹Die Rahmengröße ist dabei bezogen auf das Eingangssignal über 100 Samples lang, was den entstehenden Signalisierungsaufwand im Vergleich zu den in [SHZ14] verwendeten Werten deutlich geringer ausfallen lässt.

bar werden. Im Mittel liegen die Verfahren mit Werten von ca. -0,3 (Subband-Codec) und ca. -0,2 (Breitband-Codec) deutlich in dem, in dieser Arbeit avisierten, ODG-Bereich. Sie erzielen damit eine Audioqualität, die ungefähr der bei MP3-Codierung mit konstanter und gleicher Bitrate für ein Stereosignal und Verwendung einer aktuellen Version des Lame-Codecs [Lam11] erzielbaren entspricht. Die numerischen Ergebnisse inklusive der Zuordnung der Position im Testset zu aussagekräftigen Signalnamen befindet sich im Anhang dieser Arbeit.

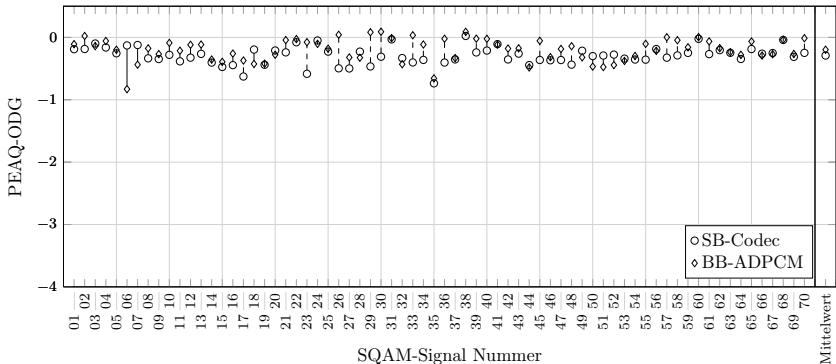


Abbildung 5.1: Ergebnisse der PEAQ-Evaluation des entwickelten Subband-Codecs (SB) für das SQAM-Testset und 4 Bit/Sample (Netto) im Vergleich zu einer Breitband-ADPCM identischer Bitrate gemäß [SHZ14].

Ein etwas anderes Bild ergibt sich für eine Codierung mit deutlich niedrigerer Bitrate. Abbildung 5.2 zeigt einen erneuten Vergleich des Subband- und Breitband-Codecs in diesem Fall jedoch für 3 Bit/Sample. Hierbei resultiert die Reduktion der Bitrate für beide Codierverfahren zum Teil in einer deutlichen Verschiebung der ODG-Werte in einen Bereich, ab dem die Hörbarkeit von Unterschieden zum Originalsignal selbst für ungeübte Hörer wahrscheinlich wird. Das in dieser Arbeit entwickelte Codierverfahren erreicht dabei im Mittel eine mit einem Wert von -0,62 um 0,25 bessere PEAQ-Bewertung. Für die breitbandige Codierung gemäß [SHZ14] ergeben sich für die kritischen SQAM-Signale „27 Kastagnetten“ und „32 Triangel“ sowie eines der synthetischen Signale deutliche Ausreißer, die in Bereichen des ODG-Wertes landen, in denen die entstehenden Artefakte für die in dieser Arbeit beschriebenen Zielanwendungen nicht mehr als tolerabel gewertet werden können.

Um den Abgleich mit dem Stand der Technik abzurunden, ist in Abbildung 5.3 der Vergleich von Ergebnissen des Subband-Codecs zu einer Evaluation des, bereits in der Einleitung dieser Arbeit diskutierten, Opus-Codecs dargestellt. Dieser wurde aus den Quelldateien in der Version 1.1.2 übersetzt und über das Softwarepaket „Opus-tools“ in der Version 0.1.9 für das Encodieren und Decodieren benutzt. Hierbei wurde der Opus-Codec mit der minimal einstellbaren Blockgröße von 2,5 ms,

einer erzwungenen konstanten Bitrate, keiner erlaubten Containerverzögerung sowie der maximal möglichen Komplexität bei den Berechnungen betrieben.

In den, in Abbildung 5.3 dargestellten, Ergebnissen zeigt sich der mit höherer Latenz und durch Transformationscodierung erzielbare Gewinn in der PEAQ-ODG Bewertung, wobei sich für die Mehrzahl der Signale sogar Werte größer Null ergeben, was für eine nahezu perfekte Rekonstruktion am Ausgang des Decoders spricht.

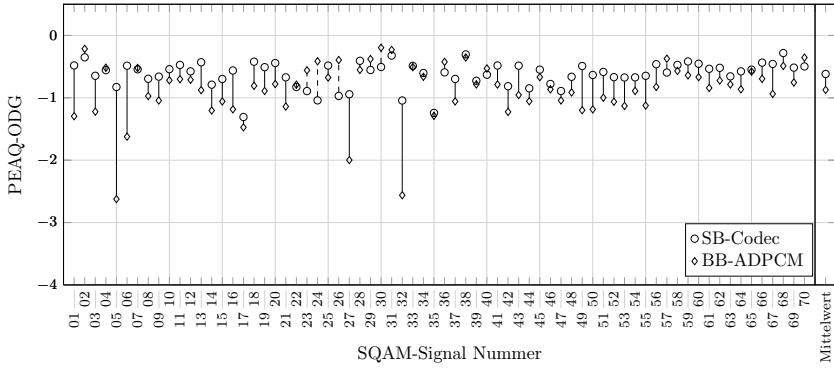


Abbildung 5.2: Ergebnisse der PEAQ-Evaluation des entwickelten Subband-Codecs (SB) für das SQAM-Testset und 3 Bit/Sample (Netto) im Vergleich zu einer Breitband-ADPCM identischer Bitrate gemäß [SHZ14].

Allerdings weisen die Ergebnisse des Opus-Codecs für die natürlichen Signale auch zwei auffällige Ausreißer auf, die in ODG-Bereiche fallen, bei denen deutliche Artefakte der Codierung hörbar werden. Die sich für die synthetischen Signale beim Opus-Codec ergebenden schlechten Bewertungen sind nicht etwa Ergebnis einer Übersteuerung der internen Signalverarbeitung durch das Eingangssignal. Sie sind vielmehr, soweit nachvollziehbar, Resultat der Synthese sinusartiger Signale im Decoder als eine Art Schmalbandrauschen bzw. dabei entstehender nichtlinearer Verzerrungen. Die Ergebnisse der synthetischen Signale spielen aber, wie auch schon für den Breitband-Codec, nur eine untergeordnete Rolle bei der Gesamtbewertung des Verfahrens, zumal die PEAQ-Bewertung solcher Signale unter Umständen nicht linear mit der tatsächlich empfundenen Audioqualität einhergeht.

Insgesamt fällt das Ergebnis des Subband-Codecs im Vergleich zum Opus-Codec mit seiner um fast 2 ms höheren Latenz in Bezug auf die maximal erzielbare Audioqualität für natürliche Signale daher etwas schlechter aus. Allerdings erreicht ersterer durch die globale Optimierung unter besonderer Berücksichtigung deutlich schlechter bewerteter Signale für alle PEAQ-ODGs ein Ergebnis, das auf eine Audioqualität hindeutet, die immer noch als nahezu transparent zu bewerten ist. Eine weitere leichte Erhöhung der Datenrate des in dieser Arbeit entwickelten Verfahrens lässt zudem das Verbessern des mittleren ODG-Wertes auf -0.2 zu, wobei eine Überschreitung der „Null Punkte Linie“ jedoch nur in wenigen Fällen stattfindet.

Ein wesentlicher und bereits diskutierter Vorteil der Subband- gegenüber der Breitband-Codierung ergibt sich in der feiner einstellbaren Bitrate. Die in Abbildung 5.4 dargestellten Ergebnisse verdeutlichen, wie durch schrittweise Erhöhung der Datenrate eine stetige Verbesserung des mittleren ODG-Wertes zu erreichen ist. Für Anwendungen, in denen nur eine leichte Erhöhung der Datenrate tolerabel, aber eine bessere Audioqualität als die für 3 Bit/Sample benötigt wird, ist es somit besser möglich, einen optimalen Arbeitspunkt zu finden.

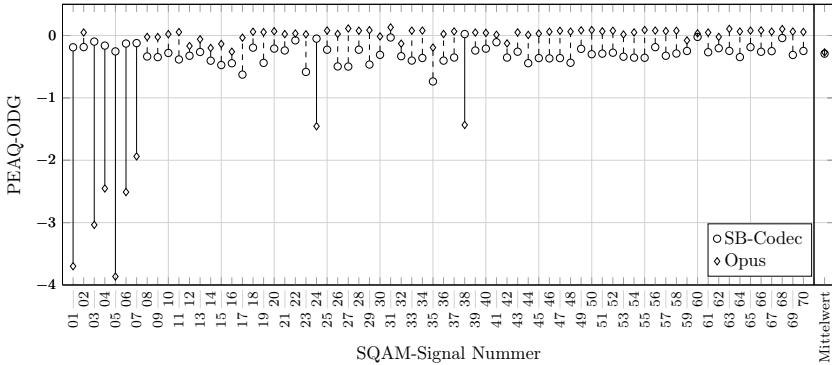


Abbildung 5.3: Ergebnisse der PEAQ-Evaluation des entwickelten Subband-Codecs (SB) für das SQAM-Testset und 4 Bit/Sample (Netto) im Vergleich zum Opus-Codec bei entsprechender Datenrate.

Zusammenfassend lässt sich somit festhalten, dass der in dieser Arbeit entwickelte Codec im PEAQ-basierten Vergleich zu anderen Verfahren durchaus konkurrenzfähig ist. Er erlaubt bei der moderaten Bitrate von 192 kbit/s eine PEAQ-Bewertung, die auf äußert geringe Unterschiede des codierten zum originalen Signal hinweist. Darauf hinaus stellt er selbst für die geringe Übertragungsrate von 144 kbit/s für viele Signale eine als immer noch gut zu wertende Audioqualität bereit. Durch die globale Parameteroptimierung werden dabei für das gesamte SQAM-Testset gleichmäßig gute Ergebnisse ohne starke Ausreißer erreicht. Allerdings bedürfen diese PEAQ-basierten Abschätzungen einer zusätzlichen Bestätigung durch subjektive Tests.

5.2 Ergebnisse subjektiver Tests

Neben der Evaluierung des in dieser Arbeit entwickelten Codecs mittels des PEAQ-Algorithmus wurden im Rahmen dieser Arbeit auch Probandenhörversuche nach einer standardisierten Methode zur Erfassung der subjektiv wahrgenommenen Audioqualität durchgeführt. Dieses ist notwendig und sinnvoll, um die Ergebnisse aus

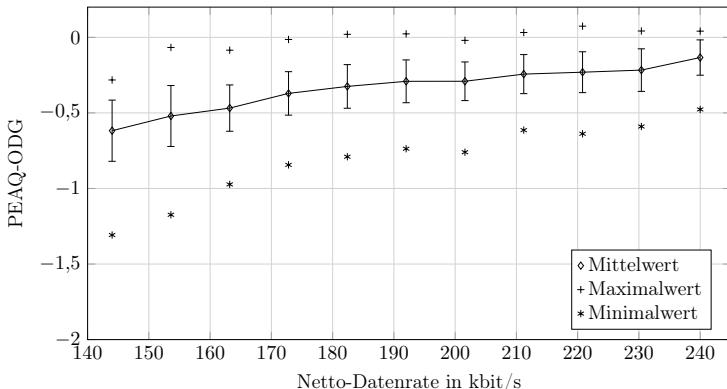


Abbildung 5.4: Ergebnisse der PEAQ-Evaluation des entwickelten Subband-Codecs als Mittelwert (\diamond), Standardabweichung (-) sowie Maximal- und Minimalwert (+,*) für das SQAM-Testset und verschiedene optimierte Bitraten.

dem vorherigen Kapitel zu bestätigen sowie weitere Aufschlüsse über die Leistungsfähigkeit des entwickelten Codierverfahrens zu erlangen.

Der folgende Abschnitt dient dem Beschreiben des dabei genutzten Testaufbaus sowie des Vorgehens bei der Durchführung und stellt die erzielten Ergebnisse und Erkenntnisse vor.

5.2.1 Testaufbau und Durchführung

Zur Durchführung der subjektiven Tests wurde ein Vorgehen gemäß des, von der ITU in dem Standard BS.1534-1 beschriebenen, Verfahrens *Multi-Stimulus Test with Hidden Reference and Anchor* (MUSHRA) [ITU03] gewählt. Dieses sieht entsprechend der Namensgebung einen Hörversuch mit mehreren Stimuli, einer versteckten Referenz und einem oder mehreren so genannten Ankersignalen vor.

Abbildung 5.5 (a) zeigt ein Bildschirmfoto der dabei verwendeten graphischen Benutzeroberfläche. Diese ermöglicht das direkte Umschalten zwischen der Referenz und den zufällig angeordneten, zu bewertenden, Signalen sowie die Zuweisung einer Note gemäß der Qualitätsskala. Für die visuelle Orientierung und das Erstellen von Wiederholungsschleifen ist zusätzlich die Wellenform des Referenzsignals dargestellt. Die Probanden haben die Möglichkeit, innerhalb der Versuchsdurchführung zu alten Ergebnissen zurückzukehren und Änderungen an den Ergebnissen vorzunehmen. Der Wechsel zum nächsten Signal ist allerdings nur möglich, wenn einem der Signale 100 Punkte zugeordnet wurden, da eines der Signale die versteckte Referenz ist, die es nach Möglichkeit gilt, zu identifizieren.

Der Hörversuch wurde in einem mit Stellwänden und Stoffbahnen zusätzlich akustisch bedämpften Büraum (vgl. Abb. 5.5 (b)) unter Verwendung eines Kopfhö-

rerverstärkers STAX SRM-1/MK-2 Professional mit dem Kopfhörer SR Lambda Professional sowie einem RME Babyface als DA-Wandler durchgeführt. Das Probandenfeld setzte sich aus insgesamt 20 lokalen Musikern, Musikproduzenten und Masteringenieuren sowie Studenten der Hochschule für Musik, Theater und Medien Hannover aus dem Studiengang für Populärmusik zusammen. Damit handelt es sich um einen Personenkreis, dem Erfahrungen im kritischen Hören unterstellt werden kann.

Der Test gliederte sich in eine Trainingsphase, die es den Probanden ermöglichte, sich mit der Benutzeroberfläche und den Umgebungsbedingungen vertraut zu machen sowie in die eigentliche Evaluationsphase. Die Anordnung der Signale über den Versuchsverlauf war zur Mittelung von Lern- und Ermüdungseffekten für jeden Probanden verschieden und zufällig. Den Probanden wurde es innerhalb vorgegebener Grenzen erlaubt, die Abhörlautstärke anzupassen.

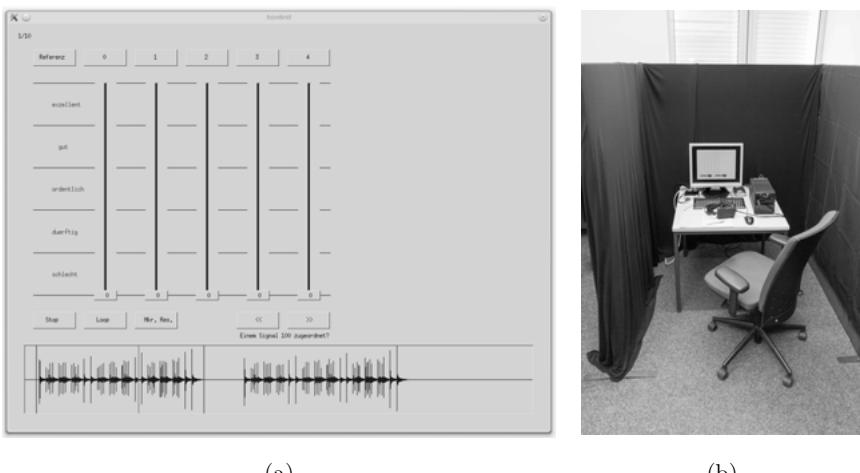


Abbildung 5.5: Details zu den subjektiven Tests: (a) Verwendete graphische Benutzeroberfläche zur Durchführung der MUSHRA-Hörversuche (b) Versuchsaufbau.

Die subjektive Bewertung erfolgte für die bereits im vorherigen Kapitel mit dem PEAQ-Algorithmus evaluierten Codecversionen mit 3 und 4 Bit/Sample. Die Auswahl der Signale geschah gemäß der in Kapitel 1 beschriebenen Anwendungsfelder für eine Audiocodierung mit geringer Latenz und auf Basis der Tendenzen der PEAQ-Evaluation. So wurden anhand der Zielanwendung die Sprachsignale weiblich Deutsch (53) und männlich Deutsch (54), die Gesangssignale Sopran (44) und Bass (47) sowie die Trompete (21) für den Test verwendet. Zusätzlich wurden die als am kritischsten bekannten Signale Kastagnetten (27), Triangel (32) und Glockenspiel (35) hinzugenommen. Aufgrund ihres schlechten Ergebnisses bei der PEAQ-Evaluation wurde das Testset für den Hörversuch zusätzlich um die Signale Bass

Klarinette (17) und Horn (23) ergänzt. Damit umfasste der Versuch 10 zu bewertende Signalanordnungen, was nach Erfahrung des Autors im Bereich der maximalen Anzahl bei zwei Codecs mit verhältnismäßig hoher Audioqualität und zwei Ankern liegt. Um zudem die Zeit für die Versuchsdurchführung für jeden Probanden im Rahmen zu halten und das vollständige Hören der einzelnen Signale sicherzustellen, wurden die längeren Signale 21, 23, 32 und 35 mit ihren sich wiederholenden identischen „Phrasen“ auf die inneren beiden sich unterscheidenden Anteile gekürzt. Eine Evaluation der geschnittenen Signale mit dem PEAQ-Algorithmus ergab einen maximalen Unterschied der ODG-Werte von 0,1 zu den Ergebnissen der ungeschnittenen. Die Ankersignale wurden gemäß der Vorschläge und Vorgaben des Standards als mit 3,5 kHz und 7 kHz gefilterte Signale aus der Referenz erzeugt.

Als Ziel des Hörversuches wurde die Bestätigung der PEAQ-Ergebnisse für den entwickelten Subband-Codec formuliert. Die Evaluation fand daher auch nur für diesen Codec statt, um Aussagen für möglichst viele Signale zu erhalten. Die von den Probanden über die Vergabe der Noten zu beantwortende Fragestellung zielte auf die von den „Codierverfahren“ erreichte Audio- bzw. Rekonstruktionsqualität im relativen Vergleich zur Referenz ab.

5.2.2 Auswertung und Ergebnisse

Für die Analyse und Darstellung der erhobenen Daten von Probandenstudien wie dem MUSHRA-Hörversuch kann eine Vielzahl an statistischen Auswertungsmethoden wie z.B. *Boxplots* und *Signifikanztests* herangezogen werden. Der MUSHRA-Standard sieht allerdings konkret nur das Berechnen von Mittelwerten und Konfidenzintervallen vor. Da der durchgeführte Hörversuch lediglich der Bestätigung von PEAQ-Tendenzen dient, wurde diese einfache Analysetechnik übernommen. Dieses Vorgehen wurde auch gewählt, da nicht der Anspruch erhoben werden kann und soll, dass der Versuch alle an komplexere Berechnungsmethoden gestellten Voraussetzungen erfüllt und um dadurch eine Überinterpretation der Ergebnisse zu vermeiden.

Abbildung 5.6 zeigt demnach die Mittelwerte und 95 % Konfidenzintervalle der MUSHRA-Bewertungen² für die Referenz, die beiden Anker sowie die Codec Versionen mit 3 Bit/Sample und 4 Bit/Sample für die 10 im vorherigen Abschnitt beschriebenen Signale. Es ist zu erkennen, dass sich die Tendenzen der PEAQ-Evaluation in groben Zügen bestätigen. Demnach ergeben sich für den mit 4 Bit/Sample betriebenen Codec für alle Signale ein Mittelwert größer 80 und für die meisten Signale sogar eine mittlere Bewertung über 90, wobei dabei relativ geringe Konfidenzintervalle vorliegen. Die, bei vielen Probanden aufgetretene, mehrfache Verwechselung der Referenz mit einem durch eine der Codec-Versionen erzeugten Signal spricht hierbei nach Erfahrung des Autors weniger für ein eingeschränktes Urteilsvermögen des Probanden, sondern vielmehr für den Codec und seine zum Teil nahezu transparente Audioqualität. Ob die zum Teil auftretenden Unzulänglichkeiten der Codierung ein Ausschlusskriterium für die Verwendung in manchen der in Kapitel

²Die einzelnen Wertungsergebnisse der Probanden befinden sich im Anhang A.2 dieser Arbeit.

1 skizzierten Anwendungsszenarien bedeuten könnten, war allerdings nicht direkter Gegenstand der präsentierten Evaluationen und bedarf weiterer Untersuchungen. Dennoch wird aufgrund der Ergebnisse davon ausgegangen, dass die Audioqualität des Subband-Codecs für 4 Bit/Sample auch erhöhten Ansprüchen gerecht wird.

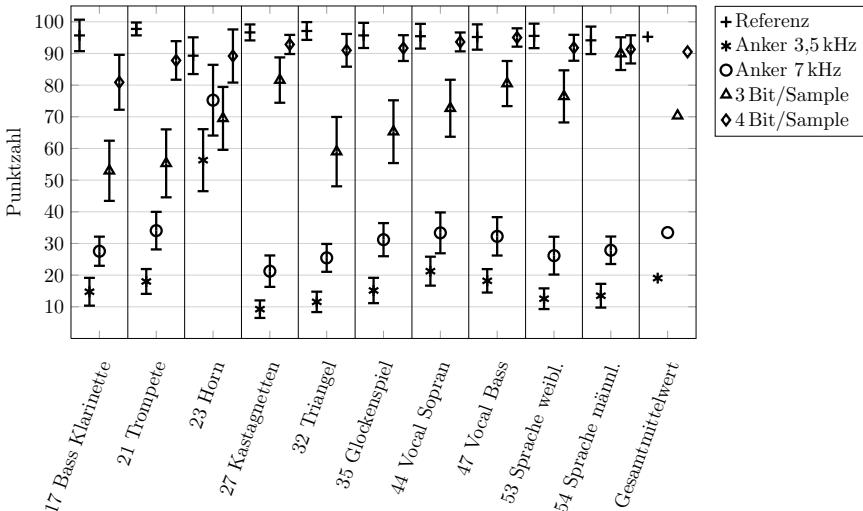


Abbildung 5.6: Ergebnisse der subjektiven Evaluation des entwickelten Codierverfahrens nach dem MUSHRA-Standard als Mittelwerte und 95 % Konfidenzintervalle.

Bei der Codec-Version mit 3 Bit/Sample zeigt sich allerdings, wie bereits in den PEAQ-Ergebnissen deutlich wurde, dass außer bei den Gesangs- und Sprachsignalen durch die Codierung in den meisten Fällen Artefakte entstehen, die deutlich hörbar und zum Teil leicht störend werden. So erreicht der Codec in dieser Konfiguration für die Signale, die kritisch einzuordnen sind und bei denen die PEAQ-Bewertung größtenteils auch nur Werte kleiner –1 ergab, nur eine MUSHRA-Bewertung im Mittelfeld der Notenskala. Hierzu ist anzumerken, dass für den Hörversuch allerdings bewusst ein besonders kritischer Probandenkreis gewählt wurde, was die Tendenz erklären könnte, dass bereits bei etwas leichter identifizierbaren Unterschieden zur Referenz die Signale als lediglich „ordentlich“ gewertet wurden. Die im Vergleich zur 4 Bit/Sample Version größeren Konfidenzintervalle der MUSHRA-Punktzahlen verdeutlichen des Weiteren die Tatsache, dass die Bewertung der auftretenden Artefakte zwischen den Probanden sehr stark variiert, da deren Wahrnehmung und Wertung deutlich stärker von der individuellen Hörphysiologie und persönlichen Präferenzen abhängig ist.

Eine letzte interessante Beobachtung lässt sich bei den Ergebnissen für das Signal Horn (23) machen. Demnach wird der Einfluss der Tiefpassfilterungen für dieses Signal mit seinem ohnehin schon auf natürliche Weise bandbegrenztem Spektrum als

weniger oder gleich störend bewertet, als die durch die Codierung für 3 Bit/Sample auftretenden Artefakte. Auch das Ergebnis für das Referenzsignal fällt gegenüber den anderen ab, da häufiger eine Verwechslung mit dem Signal der 4 Bit/Sample Codec-Version auftritt.

Abbildung 5.7 zeigt zusätzlich einen Vergleich der Hörversuchsergebnisse zu einer PEAQ-Evaluation. Hierzu wurden die PEAQ-ODGs für die Hörversuchssignale über eine lineare Umrechnung mittels einer Multiplikation mit 25 und einer Addition des Wertes 100 in den MUSHRA-Wertebereich abgebildet. Ein derartiges Vorgehen ist zwar in der Auslegung des PEAQ- und MUSHRA-Verfahrens nicht vorgesehen und entzieht sich einer statistischen Grundlage, es soll an dieser Stelle aber auch lediglich dem Abgleich von Tendenzen dienen.

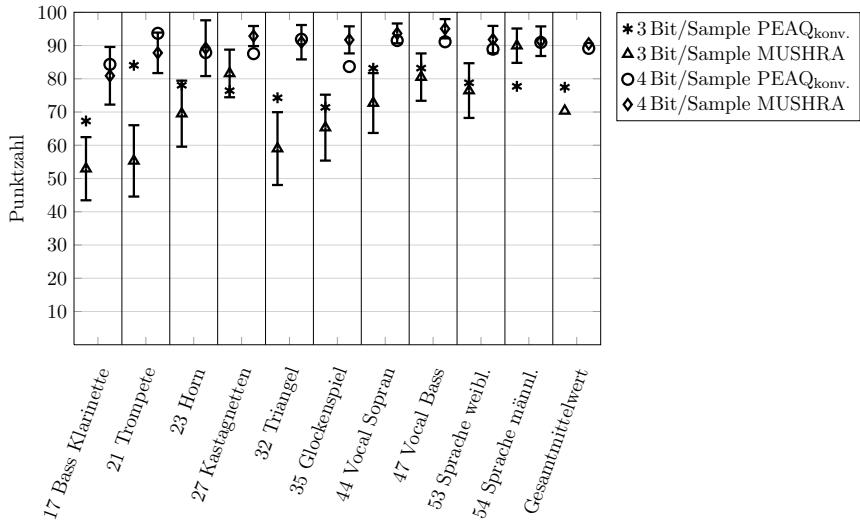


Abbildung 5.7: Vergleich der MUSHRA-Ergebnisse mit, in den Wertebereich von 0 bis 100 konvertierten, PEAQ-ODGs für die Signale des Hörversuches.

Beim Vergleich der Ergebnisse ist zu erkennen, dass sich für die Version des Codecs mit 4 Bit/Sample eine relativ hohe Übereinstimmung in den umgerechneten PEAQ-ODGs und den MUSHRA-Punktzahlen ergibt und die PEAQ-Werte zum größten Teil sogar in die Konfidenzintervalle der MUSHRA-Ergebnisse fallen. Für die Decoderversion mit 3 Bit/Sample entsteht jedoch ein leicht anderes Bild. Demnach fällt die Hörversuchsbewertung tendenziell schlechter aus als die umgerechneten PEAQ-Werte und sie weist für das Signal Trompete (21) sogar einen starken Ausreißer auf. Somit bestätigt sich im Rahmen dieser Auswertung die bereits geäußerte Vermutung, dass die Probanden Signale mit leichter identifizierbaren Unterschieden zum Referenzsignal insgesamt kritischer bewerten, als es die mittels des PEAQ-Algorithmus ermittelten Werte vorhersagen.

5.3 Einfluss von Kanalübertragungsfehlern

In diesem Abschnitt erfolgt die grobe Abschätzung des Einflusses von Kanalübertragungsfehlern auf die wahrgenommene Audioqualität des in dieser Arbeit entwickelten Audiocodierungsverfahrens. Hierbei kam wiederum der PEAQ-Algorithmus zur Anwendung. Dieser wurde in seiner Konzeption und dem Training auf Hörversuchsergebnisse ursprünglich zwar nicht für diese Art der Bewertung ausgelegt, kann aber gemäß eigener Erfahrungen und Veröffentlichungen wie [SHZ14] für das Erfassen von Tendenzen des Bitfehlerverhaltens sowie dem relativen Vergleich mit anderen Codierverfahren Anwendung finden. Die PEAQ-Evaluation hat den großen Vorteil gegenüber subjektiven Tests, dass damit eine Vielzahl an Bitfehlermustern und Raten evaluiert werden und somit ein umfassender Überblick über das Verhalten eines Codierverfahrens bei Übertragungsfehlern erstellt werden kann.

Bei den hier durchgeführten Abschätzungen wurde auf die Verwendung eines komplexeren Modells zum Erzeugen von Bitfehlermustern gemäß einer Kanalsimulation verzichtet und lediglich Einzelbitfehler definierter Fehlerraten in den Nutzdatenstrom eingeführt. Von einer Simulation des Störens der nötigen Synchronisations- und Zusatzinformationen wurde abgesehen, da an dieser Stelle hauptsächlich der Einfluss einkoppelnder Fehler auf die Adaptionsalgorithmen der Prädiktion und adaptiven Quantisierung von Interesse war. Das Bewerten des Codec-Verhaltens nach auftretenden Bündelfehlern oder dem, aufgrund eines als fehlerhaft erkannten Rahmenkopfes, eventuell nötigen Verwerfens ganzer Datenrahmen ist dabei ebenso kein Gegenstand dieser Untersuchung.

Um den Vergleich mit Ergebnissen aus [SHZ14] sowie eine angemessene Simulationszeit zu ermöglichen, wurden die verwendeten Signale des SQAM-Testsets ebenfalls auf die dort verwendeten 10 s langen Sequenzen gekürzt. Hierbei wurde für die Evaluation des Subband-Codecs jedes SQAM-Signal 100 Mal aus einem mit einer entsprechenden Rate gestörten Datenstrom decodiert, eine PEAQ-Evaluation durchgeführt und der Mittelwert über alle Signale berechnet. Eine Analyse von Verläufen der PEAQ-Werte verschiedener Fehlerraten bei steigender Evaluationszahl zeigte, dass deren weitere Erhöhung kaum noch Einfluss auf die Ergebnisse hat.

Die Evaluation fand für die beiden bereits untersuchten Codec-Versionen mit ihren aus 3 Bit/Sample und 4 Bit/Sample resultierenden Datenraten sowie jeweils mit und ohne Übertragung des quantisierten Prädiktionsfehlernormalisierungsfaktors statt. Der Vergleich mit den Ergebnissen aus [SHZ14] erfolgt dabei für eine Periode der Prädiktionsfehlernormalisierungsfaktübertragung von $r_v = 40$, da diese in Bezug auf den resultierenden Signalisierungsaufwand am dichtesten an dem sich für den Subband-Codec ergebenden liegt.

In Abbildung 5.8 sind die Ergebnisse der PEAQ-Evaluation für den Subband-Codec mit 4 Bit/Sample dargestellt. Es zeigt sich, dass seine PEAQ-Bewertung für Fehlerraten bis 10^{-4} in den Bereich der Ergebnisse aus [SHZ14] ohne Normalisierungsfaktübertragung fällt. Der durch das Übertragen eines quantisierten Prädiktionsfehlernormalisierungsfaktors erzielbare Gewinn in der PEAQ-Bewertung fällt allerdings deutlich geringer aus als bei den Verläufen aus [SHZ14]. Dies begründet sich vermut-

lich durch die Tatsache, dass der Prädiktionsfehlernormalisierungsfaktor trotz der Auswahl des geringsten Wertes aus [SHZ14] für den Subband-Codec weniger häufig und nur für das Tiefpass-Band übertragen wird, so dass dessen Effekt deutlich geringer ausfällt. Der leichte Unterschied des PEAQ-Mittelwertes im fehlerlosen Fall beim Vergleich zu den Ergebnissen aus Unterabschnitt 5.1.2 liegt in der Verwendung der gekürzten SQAM-Signale begründet.

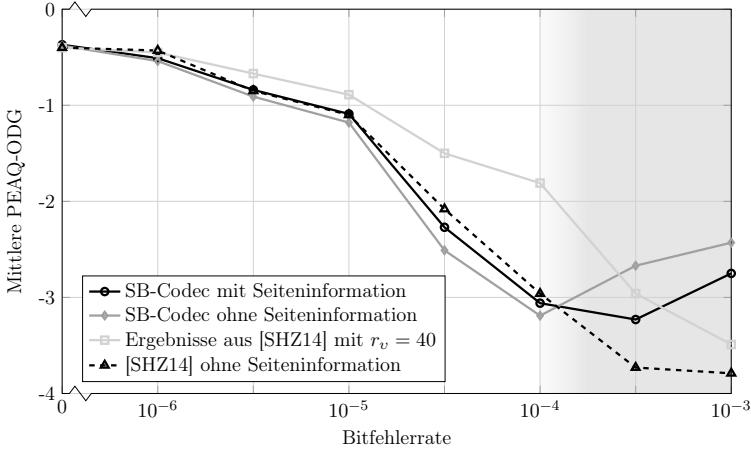


Abbildung 5.8: PEAQ-Ergebnisse des Subband-Codecs (SB) bei 4 Bit/Sample als Mittelwert für jeweils 100 Realisierungen verschiedener Bitfehlerraten und die Signale des SQAM-Testsets im Vergleich zu entsprechenden Ergebnissen aus [SHZ14].

Auffallend ist der, ab einer Fehlerrate von 10^{-4} , wieder erfolgende leichte Anstieg im PEAQ-Gesamtmittelwert. Dieser wird, soweit nachvollziehbar, dadurch verursacht, dass bei höheren Fehlerraten die im Prädiktor und der adaptiven Pegelschätzung des Decoders befindlichen Fehler noch nicht wieder vollständig abgeklungen sind, bevor ein neuer Fehler einkoppelt. Der Codec gerät dadurch für manche Signale in einen Zustand, der durch die dann zum Teil auftretenden starken Abweichungen der decodierten und originalen Wellenform über das PEAQ-Verfahren nicht mehr erwartungsgemäß bewertet wird. Der Anstieg in den mittleren ODG für die Codec-Version ohne übertragende Zusatzinformation fällt dabei etwas stärker aus, da in diesem Fall, durch die länger in die Prädiktoradaption einkoppelnden Fehler im Normalisierungssignal der adaptiven Quantisierung, der beschriebene Zustand früher erreicht wird. Allerdings korrelieren diese PEAQ-Ergebnisse nicht mit dem subjektiven Eindruck, der durch informelle Tests gewonnen wurde, weshalb die Ergebnisse nicht für einen Vergleich herangezogen werden sollten und daher grau hinterlegt sind. Evaluierungen mit anderen Parametrisierungen, die in ihrer Audioqualität im fehlerlosen Fall etwas gegenüber dem hier untersuchten Parametersatz abfallen, führen jedoch zu Verläufen mit zum Teil monoton fallenden Kurven, die zwischen den hier gezeigten Kurven

aus [SHZ14] liegen. Es kann daher davon ausgegangen werden, dass durch weitere Untersuchungen zur Optimierung des Fehlerverhaltens des Subband-Codecs eine vergleichbare PEAQ-Bewertung im Fall höherer Bitfehlerraten erzielt werden kann. Abbildung 5.9 enthält die Ergebnisse der Fehlerevaluation für die 3 Bit/Sample-Version des Codecs. Hierfür ergibt sich, wie auch schon für den fehlerlosen Fall, ein etwas anderes Bild. Der Subband-Codec erreicht für geringe Fehlerraten eine leicht bessere oder gleiche PEAQ-Bewertung wie der Ansatz aus [SHZ14], fällt jedoch mit steigender Fehlerrate deutlich mehr gegenüber diesem ab. Auch in diesem Fall ergibt sich ein leichter Anstieg in den PEAQ-Ergebnissen für Fehlerraten größer 10^{-4} mit der diskutierten Ursache und dem Hinweis, dass dieser nicht als Indikator für ein besseres Fehlerverhalten hin zu höheren Fehlerraten zu werten ist.

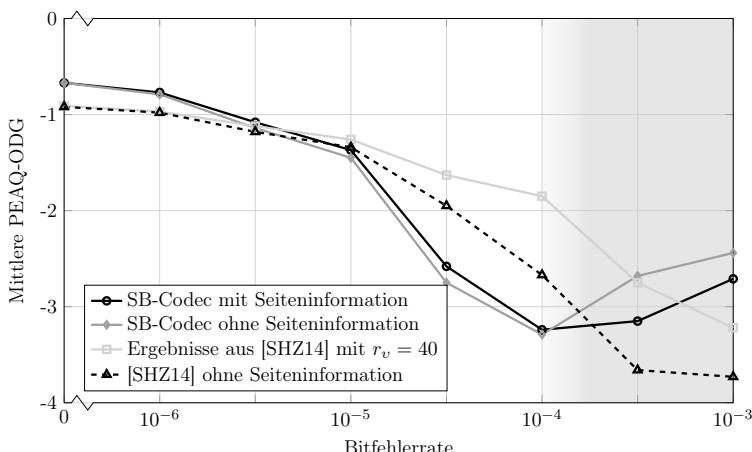


Abbildung 5.9: PEAQ-Ergebnisse des Subband-Codecs (SB) bei 3 Bit/Sample als Mittelwert für jeweils 100 Realisierungen verschiedener Bitfehlerraten und die Signale des SQAM-Testsets im Vergleich zu entsprechenden Ergebnissen aus [SHZ14].

Grundsätzlich ist für die präsentierten Ergebnisse abschließend anzumerken, dass es sich, wie bereits angedeutet, lediglich um eine grobe Abschätzung handeln kann und der Vergleich mit den Ergebnissen aus [SHZ14] nur eingeschränkt aussagekräftig und fair ist. Dies wird unter anderem auch durch die Tatsache bedingt, dass die Evaluation des Subband-Codecs mit einer höheren Abtastrate durchgeführt wurde als bei den Simulationen in [SHZ14], was bei einer gleichen Zahl an Bits pro Sample in einer höheren Gesamtzahl an Fehlern resultiert. Fraglich ist zudem, ob die in Anlehnung an [SHZ14] verwendeten Fehlerraten eine praktische Relevanz haben und in einem echten Übertragungssystem ein Codec wirklich über eine längere Zeit mit einer Rate von 10^{-3} betrieben oder nicht auf Basis von Informationen aus der Kanalcodierung eingegriffen und die Ausgabe des decodierten Signales verhindert würde. Des Wei-

teren ist der Einfluss der Normalisierungsfaktorübertragung auf die Audioqualität im Einzelfall gemäß informeller Tests stärker als er über die gemittelten PEAQ-Ergebnisse zum Ausdruck kommt. Inwiefern die Ergebnisse einer PEAQ-Bewertung von, durch Übertragungsfehler beeinflussten, Audiosignalen sowie die zu erzielenden Verbesserungen durch Übertragung von Seiteninformationen mit den Ergebnissen subjektiver Tests korrelieren, ist daher in weiteren Untersuchungen zu klären.

Zusammenfassend kann dennoch festgehalten werden, dass sich mit dem in dieser Arbeit entwickelten Codierverfahren für geringe Bitfehlerraten eine bei Ermittlung über das PEAQ-Verfahren zu einem Breitband-Ansatz vergleichbare Audioqualität bei auftretenden Kanalübertragungsfehlern erreichen lässt. Mit steigender Fehlerrate und sinkender Bitrate ist jedoch mit einem etwas schlechteren Abschneiden bei der PEAQ-Bewertung zu rechnen. Die Einführung der Übertragung eines Normalisierungsfaktorsignales innerhalb der adaptiven Quantisierung erlaubt dabei eine leichte Verbesserung in den Ergebnissen mittlerer Fehlerraten.

Kapitel

6

Zusammenfassung und Ausblick

Fehlerrobuste Audiocodierung mit geringer Latenz hat eine Vielzahl von Anwendungsfeldern. So kann diese durch Einsatz in drahtlosen Mikrofon- und Kopfhörersystemen, Audiostreaming-Applikationen und modernen Hörhilfen für effizientere Übertragungsstrecken, erweiterte Funktionsumfänge und kostengünstigere Produkte sorgen. Allerdings eignen sich bestehende standardisierte, quelloffene oder in der Literatur zu findende Verfahren aufgrund zu hoher Latenzen, zu geringer Audioqualität oder nicht vorliegender Fehlerrobustheit nur eingeschränkt für die Verwendung in solchen Systemen. Aus diesem Grund wird bei vielen Anwendungen auf proprietäre Lösungen zurückgegriffen.

In der vorliegenden Arbeit wurde daher das Ziel der Entwicklung eines Verfahrens zur fehlerrobusten Audiocodierung mit geringer Latenz verfolgt, das diese Nachteile überwindet. Hierzu wurden erfolgversprechende Ansätze um Maßnahmen zum Erreichen einer algorithmischen Fehlerrobustheit ergänzt, die Codierstrukturen hinsichtlich einer Anpassung an wechselnde Signaleigenschaften erweitert und diese bezüglich ihrer Parameter automatisiert anhand einer psychoakustischen Bewertungsmethode global optimiert. Damit soll es möglich sein, bei moderaten Bitraten bis zu 200 kbit/s eine nahezu transparente Audioqualität zu erreichen.

Nach einer Einleitung in die Grundlagen sowie Anwendungsszenarien von Audiocodierung mit geringer Latenz und in den Stand der Technik wurde in einem ersten Kapitel dieser Arbeit der darin untersuchte Ansatz einer prädiktiven Subband-Codierung motiviert und vorgestellt. Dieser sieht eine Verarbeitung des Eingangssignales in Teilbändern, die durch eine Filterbank mit geringer Latenz erzeugt werden, gemäß des Grundprinzips einer im Folgenden mit ADPCM abgekürzten *Adaptive Differential Pulse Code Modulation* vor.

In dem anschließenden Kapitel 2 wurden Verfahren zum Entwurf von Filterbänken mit geringer Latenz gemäß den Ansätzen von Nayebi, Keiler/Nguyen und Schuller präsentiert. Diese wurden auf ihre Eignung für die Optimierung von Filterbankstrukturen für den vorliegenden Anwendungsfall untersucht und die mit ihnen erstellten Entwürfe über geeignete Messgrößen miteinander verglichen. In Folge dieser Auswertung fiel die Wahl auf einen 5-Kanal Entwurf, der mit dem Entwurfsverfahren

nach Nayebi erstellt wurde, da dieser bei der Abwägung der Latenz, der Rekonstruktionsqualität und den Frequenzverläufen der Subband-Filter am geeignetsten erschien. Als Resultat weiterer Untersuchungen wurde im Rahmen dieser Arbeit für das Verfahren nach Nayebi eine neuartige Methode zur globalen Optimierung von Entwurfsparametern eingeführt. Mit dieser ist es möglich, die Entwürfe hinsichtlich der maximal zu erreichenden Sperrdämpfung um bis zu 10 dB weiter zu verbessern. Die eigentliche Codierung der über die Filterbank erzeugten Subband-Signale erfolgt mit der in Kapitel 3 vorgestellten Algorithmik zur adaptiven Prädiktion und Quantisierung. Es wurden Methoden zur fehlerrobusten Prädiktoradaption mittels des GAL-Algorithmus erörtert und neuartige Ansätze zur blockbasierten Koeffizientenberechnung mit dem Burg-Algorithmus und der Levinson-Durbin Rekursion auf Basis des Prädiktionsfehlersignales der ADPCM vorgeschlagen. Die Verfahren wurden hierbei hinsichtlich ihrer Robustheit gegenüber Übertragungsfehlern beim Einsatz in einer ADPCM analysiert, was in der Erkenntnis resultierte, dass sich lediglich mit der blockbasierten Prädiktoradaption auf das Fehlersignal eine algorithmisch sichergestellte Codecrobustheit erreichen lässt. Die Algorithmen wurden außerdem bezüglich ihres Berechnungsaufwandes miteinander verglichen, wonach sich durch die blockbasierten Verfahren die Prädiktoradaption deutlich effizienter realisieren lässt. Des Weiteren wurde ein Ansatz zur adaptiven Quantisierung über eine Prädiktionsfehlernormalisierung mittels Hüllkurvenschätzung um Maßnahmen zum Erreichen einer Fehlerrobustheit erweitert sowie das Vorgehen bei dem Entwurf von geeigneten Quantisiererkennlinien beschrieben. Die eingeführte verbesserte dynamische Bit-Allokation inklusive einer Begrenzung der Menge erlaubter Bitkombinationen lässt es zu, eine signalangepasste Verteilung der zur Verfügung stehenden Datenrate auf die Subbänder vorzunehmen und dabei den nötigen Signalisierungsaufwand vernachlässigbar gering zu halten.

Die in Kapitel 4 beschriebene Methode zur globalen Audiocodec-Optimierung gestattet es, eine geeignete Parametrisierung für den Codec zu finden, die mit hoher Wahrscheinlichkeit nah an der optimal erreichbaren in Bezug auf die resultierende Audioqualität ist. Es wurden Details zu der dabei innerhalb der Kostenfunktion verwendeten und mit PEAQ abgekürzten Bewertungsmethode der *Perceptual Evaluation of Audio Quality* sowie dem Vorgehen bei der globalen Optimierung präsentiert. Letztere erfolgte unter Nutzung des *Genetic Algorithms*, da sich dieser im Rahmen durchgeföhrter Untersuchungen als vorteilhaft gegenüber anderen Verfahren für das Lösen des vorliegenden Optimierungsproblems erwiesen hat. Anhand exemplarisch dargestellter Ergebnisse für durchgeföhrte Optimierungen wurden deren Resultate und das Konvergenzverhalten des Optimierungsverfahrens dokumentiert. Dieses ermöglicht demnach eine vollständig automatisierte Einstellung aller relevanten Codecparameter ohne signifikante Ausreißer bei der über das PEAQ-Verfahren ermittelten Audioqualität, was auf manuellem Wege undenkbar wäre. Der Nachweis der Leistungsfähigkeit des entwickelten Audiocodierungsverfahrens und der Vergleich zum Stand der Technik erfolgte in dem letzten Kapitel 5 unter anderem über eine Evaluation mit dem PEAQ-Algorithmus. Im Rahmen dieser Untersuchung erreicht das entwickelte Verfahren für eine Nettodatenrate von

192 kbit/s und bei einer Auswertung des vollständigen SQAM-Testsets einen Mittelwert der PEAQ-ODG von -0,3. Dieser fällt im Vergleich zu den Referenzverfahren des OPUS-Codecs mit seiner deutlich höheren Latenz und einer Breitband-ADPCM gemäß [SHZ14], deren Robustheit gegenüber Kanalübertragungsfehlern nicht vollständig algorithmisch sichergestellt ist, nur geringfügig schlechter aus. Er liegt dabei aber immer noch in einem Bereich, bei dem eine Unterscheidung von originalem und codiertem Signal selbst für geübte Hörer schwierig wird. Bei einer Bitrate von 144 kbit/s werden für kritische Signale jedoch zum Teil deutliche Unterschiede hörbar. Dabei erreicht der Subband-Codec mit einem Mittelwert der PEAQ-ODG von -0,62 jedoch immer noch eine um 0,25 bessere Bewertung als die Breitband-ADPCM und weist nicht deren zum Teil starke negative Ausreißer der PEAQ-ODG für kritische Signale auf. Eine Evaluierung der Audioqualität des Subband-Codecs für andere Bitraten im angestrebten Betriebsbereich zeigte, dass es mit seiner im Vergleich zu einer Breitband-Codierung relativ frei einstellbaren Bitrate zusätzlich möglich ist, eine bessere Abwägung zwischen Audioqualität und Datenrate zu treffen.

Auf Basis eines MUSHRA-Hörversuches mit im kritischen Hören geübten Probanden wurde zudem eine subjektive Bewertung der Audioqualität vorgenommen. Die dabei erhaltenen Ergebnisse bestätigen die Tendenzen der PEAQ-Evaluation. Demnach erreicht der Codec für 192 kbit/s eine Hörversuchsbewertung, die den Schluss zulässt, dass eine Unterscheidung zwischen codiertem und originalem Signal selbst für geübte Hörer und kritische Signale schwierig ist. Die Ergebnisse für die Codec-Version mit 144 kbit/s deuten jedoch darauf hin, dass diese genau wie die Breitband-ADPCM für Anwendungen mit höheren Ansprüchen an die Audioqualität kritischer Signale in der vorliegenden Version noch nicht zum Einsatz kommen sollte.

Eine Abschätzung der Audioqualität des entwickelten Subband-Codecs bei auftretenden Kanalübertragungsfehlern und ein Vergleich mit Ergebnissen aus der Literatur zeigen, dass die PEAQ-Bewertung für 192 kbit/s und praxisrelevante Fehlerraten vergleichbar zu einer fehlerrobusten Breitband-Codierung ohne übertragene Nebeninformationen ist. Für eine Bitrate von 144 kbit/s fällt die mittlere PEAQ-ODG allerdings bei höheren Fehlerraten gegenüber dem Vergleichsverfahren leicht ab. Die eingeführte Übertragung von Seiteninformation zur schnelleren Synchronisation nach Übertragungsfehlern erlaubt eine leichte Verbesserung der Ergebnisse, die jedoch unter anderem aufgrund größerer Aktualisierungsintervalle bei der Subband-Codierung geringer ausfällt als beim Referenzverfahren aus der Literatur.

Zusammenfassend lässt sich festhalten, dass es mit dem in dieser Arbeit vorgestellten Codierverfahren möglich ist, bei Nettobitraten unterhalb 200 kbit/s eine Audioqualität zu erreichen, die die Anforderungen vieler der eingangs beschriebenen Zielanwendungen erfüllt. Diese Ergebnisse wurden bei einer algorithmisch sichergestellten Codecrobustheit erzielt und durch subjektive Tests bestätigt. Hierbei liegt trotz des Einsatzes einer Filterbank die algorithmische Latenz des Codecs bei nur 0,5 ms. Die Audioqualität des entwickelten Verfahrens bei auftretenden Kanalübertragungsfehlern ist bei geringen Fehlerraten vergleichbar zu der einer Breitband-Codierung und kann durch Übertragung zusätzlicher Seiteninformation weiter verbessert werden.

Ausblickend lässt sich für den in dieser Arbeit präsentierten Codieransatz eine Vielzahl möglicher Anknüpfungspunkte für weitere Arbeiten identifizieren. Mittels der in Abschnitt 3.4.1 dieser Arbeit als Alternative zur dynamischen Bit-Allokation erstmals untersuchten adaptiven Vor- und Nachfilterung für die Subband-Codierung ist es, wie gezeigt werden konnte, ebenfalls möglich, eine Verbesserung der Audioqualität jedoch ohne zusätzlichen Signalisierungsaufwand zu erreichen. Da der Fokus bei den durchgeführten Experimenten jedoch auf dem Einfluss auf die Audioqualität des Codecs im fehlerlosen Fall lag, sollten weiterführende Arbeiten die Analyse und Verbesserung des Einflusses einkoppelnder Übertragungsfehler zum Ziel haben. Damit könnte das Verfahren durch Kombination mit der dynamischen Bit-Allokation zur Verbesserung der Audioqualität des Subband-Codecs bei niedrigen Bitraten verwendet werden, wenn der erhöhte Rechenaufwand akzeptabel ist.

Mit den in Abschnitt 3.3.4 präsentierten Ergebnissen einer vielversprechenden Untersuchung eines Konzeptes zu einer kombinierten fortschrittlichen ADPCM und Vektorquantisierung für eine Breitband-Codierung konnte darüber hinaus demonstriert werden, dass damit weitere Gewinne in der Audioqualität der Breitband-ADPCM erzielt werden können. Zukünftige Untersuchungen sollten auf die Anwendung dieser Techniken innerhalb der Subband-Codierung abzielen.

Die in dieser Arbeit durchgeführten Untersuchungen zeigen, dass die Audioqualität des entwickelten Codecs bei reiner Wiedergabe über Kopfhörer im Probandenversuch als sehr gut bewertet wird. Für Anwendungen, wie das in Kapitel 1 beschriebene Anwendungsszenario einer live-Bühnenperformance mit In-Ear-Monitoring, ist es allerdings nötig, weitere Untersuchungen mit einer Rückübertragung von durch den Probanden selbst gesungenen oder gespielten Signalen anzustellen.

Über die Analysen zum Einfluss von Kanalübertragungsfehlern auf die Audioqualität des entwickelten Coderverfahrens wurde im Rahmen dieser Arbeit lediglich die aktuelle Leistungsfähigkeit des Verfahrens, bei einer Optimierung, die auf maximale Audioqualität im fehlerlosen Fall abzielte, erfasst. Eine weitere systematische Untersuchung der auftretenden Effekte bei fehlerhafter Übertragung und deren Auswirkung auf die Audioqualität könnte neue Aufschlüsse über mögliche Ansätze zur Verbesserung liefern.

Weiterhin könnten aktive Maßnahmen zur Detektion und Behandlung auftretender Kanalübertragungsfehler eine maßgebliche Verbesserung der Audioqualität im Fehlerfall ermöglichen (vgl. [SHZ14]). Eine zusätzliche Kombination mit Verfahren zur Fehlerverschleierung (vgl. [Stö12, PSO12]) sollte zudem für das Behandeln von Bündelfehlern in Betracht gezogen werden. Der Freiheitsgrad bei der Einstellung der Bitrate des Codecs könnte in Szenarien, bei denen am Encoder eine Information über den Kanalzustand vorliegt, außerdem genutzt werden, um adaptiv das Verhältnis zwischen Nutzdaten und Fehlerschutz anzupassen. Es könnte zudem überprüft werden, ob sich das Grundprinzip der so genannten *Softbitdecodierung* aus [Fin98] auf die ADPCM-basierte Subband-Codierung übertragen lässt.

Anhang



Tabellarische Ergebnisse

A.1 Ergebnisse der objektiven Evaluierung

Tabelle A.1: ODG-Werte der untersuchten AudioCodec Versionen aus Kapitel 5.

SQAM Signal	Subband-Codec 4 Bit/Sample	Subband-Codec 3 Bit/Sample	Breitband-Codec 4 Bit/Sample	Breitband-Codec 3 Bit/Sample	Opus Codec \approx 4 Bit/Sample
01 Sine 1kHz -20, -10, -0 dB	-0,189	-0,481	-0,109	-1,295	-3,699
02 Band-limited pink noise	-0,186	-0,351	0,022	-0,215	0,048
03 Electronic gong 100 Hz	-0,098	-0,647	-0,140	-1,222	-3,036
04 Electr. gong 400 Hz	-0,162	-0,556	-0,058	-0,518	-2,453
05 Electr. gong 5 kHz	-0,255	-0,827	-0,198	-2,624	-3,865
06 Electr. gong 500 Hz vibr,	-0,129	-0,485	-0,830	-1,623	-2,510
07 Electr. tune	-0,122	-0,543	-0,437	-0,522	-1,938
08 Violin	-0,336	-0,695	-0,176	-0,972	-0,023
09 Viola	-0,347	-0,660	-0,268	-1,043	-0,025
10 Violoncello	-0,280	-0,541	-0,088	-0,719	0,022
11 Double-bass	-0,384	-0,473	-0,214	-0,701	0,055
12 Piccolo	-0,325	-0,576	-0,119	-0,711	-0,168
13 Flute	-0,264	-0,430	-0,115	-0,877	-0,060
14 Oboe	-0,404	-0,789	-0,355	-1,203	-0,200
15 Cor anglais	-0,474	-0,698	-0,392	-1,058	-0,135
16 Clarinet	-0,445	-0,562	-0,260	-1,185	-0,260
17 Bass-clarinet	-0,627	-1,308	-0,371	-1,473	-0,035
18 Bassoon	-0,195	-0,421	-0,426	-0,808	0,059
19 Contra-bassoon	-0,439	-0,509	-0,422	-0,890	0,054
20 Saxophone	-0,210	-0,442	-0,278	-0,779	0,068

wird fortgesetzt...

... Fortsetzung

SQAM Signal	Subband-Codec 4 Bit/Sample	Subband-Codec 3 Bit/Sample	Breitband-Codec 4 Bit/Sample	Breitband-Codec 3 Bit/Sample	Opus Codec ≈4 Bit/Sample
21 Trumpet	-0,240	-0,672	-0,043	-1,140	0,022
22 Trombone	-0,078	-0,827	-0,027	-0,790	0,032
23 Horn	-0,584	-0,892	-0,077	-0,559	0,019
24 Tuba	-0,050	-1,041	-0,103	-0,414	-1,455
25 Harp	-0,228	-0,483	-0,181	-0,675	0,080
26 Claves	-0,496	-0,970	0,042	-0,395	0,025
27 Castanets	-0,498	-0,943	-0,321	-1,998	0,112
28 Side drum	-0,228	-0,407	-0,326	-0,552	0,077
29 Bass drum	-0,466	-0,554	0,082	-0,379	0,085
30 Kettle-drums	-0,310	-0,506	0,090	-0,197	-0,015
31 Cymbal	-0,032	-0,323	-0,012	-0,232	0,130
32 Triangles	-0,332	-1,043	-0,430	-2,562	-0,131
33 Gong	-0,402	-0,488	0,034	-0,508	0,078
34 Tubular bells	-0,361	-0,603	-0,114	-0,663	0,079
35 Glockenspiel	-0,737	-1,242	-0,657	-1,289	-0,194
36 Xylophone	-0,403	-0,592	-0,021	-0,424	0,024
37 Vibraphone	-0,353	-0,698	-0,332	-1,058	0,065
38 Marimba	0,023	-0,304	0,089	-0,354	-1,434
39 Grand piano	-0,241	-0,727	-0,020	-0,785	0,047
40 Harpsichord	-0,211	-0,630	-0,023	-0,529	0,041
41 Celesta	-0,108	-0,482	-0,115	-0,787	0,012
42 Accordion	-0,354	-0,814	-0,179	-1,225	-0,127
43 Organ	-0,260	-0,485	-0,175	-0,955	0,050
44 Soprano	-0,444	-0,848	-0,482	-1,055	0,009
45 Alto	-0,362	-0,548	-0,055	-0,668	0,031
46 Tenor	-0,367	-0,778	-0,317	-0,864	0,060
47 Bass	-0,363	-0,891	-0,186	-1,043	0,075
48 Quartet	-0,437	-0,663	-0,142	-0,914	0,062
49 Female speech (English)	-0,214	-0,490	-0,320	-1,199	0,083
50 Male speech (English)	-0,300	-0,634	-0,468	-1,186	0,087
51 Female speech (French)	-0,291	-0,585	-0,475	-0,998	0,067
52 Male speech (French)	-0,276	-0,669	-0,444	-1,063	0,076
53 Female speech (German)	-0,340	-0,674	-0,377	-1,130	0,016

wird fortgesetzt...

... Fortsetzung

SQAM Signal	Subband-Codec 4 Bit/ Sample	Subband-Codec 3 Bit/ Sample	Breitband-Codec 4 Bit/ Sample	Breitband-Codec 3 Bit/ Sample	Opus Codec ≈4 Bit/ Sample
54 Male speech (German)	-0,355	-0,672	-0,298	-0,890	0,050
55 Trumpet Haydn	-0,357	-0,647	-0,104	-1,124	0,088
56 Organ Handel	-0,187	-0,461	-0,216	-0,825	0,079
57 Organ Bach	-0,325	-0,595	0,002	-0,373	0,070
58 Guitar Sarasate	-0,289	-0,473	-0,045	-0,569	0,076
59 Violin Ravel	-0,248	-0,417	-0,159	-0,639	-0,079
60 Piano Schubert	-0,023	-0,454	0,004	-0,669	0,035
61 Soprano Mozart	-0,268	-0,535	-0,064	-0,842	0,048
62 Soprano Spiritual	-0,204	-0,519	-0,172	-0,723	-0,022
63 Soloists Verdi	-0,247	-0,656	-0,237	-0,786	0,106
64 Choir Orff	-0,346	-0,576	-0,280	-0,864	0,064
65 Orchestra Strauss	-0,187	-0,548	-0,065	-0,580	0,078
66 Wind ensemble Stravinsky	-0,261	-0,436	-0,292	-0,696	0,082
67 Wind ensemble Mozart	-0,251	-0,457	-0,270	-0,937	0,062
68 Orchestra Baird	-0,041	-0,282	-0,038	-0,489	0,103
69 ABBA	-0,312	-0,516	-0,266	-0,755	0,063
70 Eddie Rabbitt	-0,248	-0,497	-0,013	-0,355	0,057

A.2 Ergebnisse der subjektiven Evaluierung

Tabelle A.2: Detaillierte Ergebnisse der subjektiven Tests gemäß des MUSHRA-Standards für die Subband-Codec Versionen mit 3 (Cod3) und 4 Bit/Sample (Cod4) im Vergleich zur Referenz (Ref) und den Ankersignalen (Ank3, Ank7).

Proband /Codec	SQAM-Signal									
	Signal 17	Signal 21	Signal 23	Signal 27	Signal 32	Signal 35	Signal 44	Signal 47	Signal 53	Signal 54
A/Ref	100	100	100	99	100	100	100	100	97	98
A/Ank3	4	9	31	1	0	7	29	18	8	15
A/Ank7	17	22	43	9	8	25	29	33	27	35
A/Cod3	30	28	82	84	62	41	98	93	59	99
A/Cod4	86	86	98	100	98	92	99	95	100	100
B/Ref	100	100	82	100	100	71	100	100	75	100
B/Ank3	8	14	24	9	4	8	13	23	9	11
B/Ank7	25	27	37	19	10	19	26	35	24	23
B/Cod3	14	22	18	72	29	14	45	83	76	91
B/Cod4	77	69	100	90	82	100	92	86	100	80
C/Ref	95	100	95	95	100	95	100	100	100	100
C/Ank3	30	20	70	10	10	30	30	30	20	30
C/Ank7	50	40	80	30	30	50	60	60	45	40
C/Cod3	75	55	60	80	95	100	80	95	100	100
C/Cod4	100	95	100	100	100	100	95	95	95	90
D/Ref	100	96	92	96	100	100	77	95	100	98
D/Ank3	7	5	68	1	3	7	5	9	7	6
D/Ank7	11	15	93	3	11	24	17	24	12	21
D/Cod3	25	21	96	69	18	75	72	70	75	86
D/Cod4	31	100	100	100	76	84	100	100	92	100
E/Ref	100	100	95	100	100	100	100	100	100	100
E/Ank3	25	25	98	10	15	10	30	15	10	10
E/Ank7	35	40	98	20	30	30	40	30	25	25
E/Cod3	35	50	75	80	45	45	88	95	95	95
E/Cod4	60	90	100	97	98	90	90	98	95	95
F/Ref	87	100	90	91	94	100	85	100	100	89
F/Ank3	12	18	65	8	14	20	18	18	17	22
F/Ank7	31	46	72	15	27	37	27	26	33	38
F/Cod3	43	38	77	95	86	65	95	94	86	93
F/Cod4	100	92	100	100	100	82	100	100	94	100
G/Ref	100	100	100	100	100	100	100	100	80	90

wird fortgesetzt...

... Fortsetzung

Proband /Codec	SQAM-Signal									
	Signal 17	Signal 21	Signal 23	Signal 27	Signal 32	Signal 35	Signal 44	Signal 47	Signal 53	Signal 54
G/Ank3	20	20	30	10	15	20	20	30	20	25
G/Ank7	30	20	70	30	20	25	20	30	20	30
G/Cod3	40	40	30	51	40	50	40	60	50	60
G/Cod4	76	80	80	88	69	80	85	86	100	100
H/Ref	95	96	93	100	100	93	100	100	100	93
H/Ank3	23	21	60	21	18	22	22	29	23	15
H/Ank7	34	37	76	34	37	44	35	53	55	37
H/Cod3	54	59	100	70	48	65	71	69	79	100
H/Cod4	100	100	80	93	93	100	97	95	89	83
I/Ref	100	95	100	100	100	100	95	100	95	90
I/Ank3	10	15	30	10	15	10	10	10	10	10
I/Ank7	30	30	64	20	25	20	25	25	25	20
I/Cod3	60	65	55	85	65	65	80	85	87	100
I/Cod4	95	100	95	95	100	95	100	90	100	95
J/Ref	100	100	92	100	84	90	88	71	77	70
J/Ank3	12	27	65	11	20	13	36	19	20	25
J/Ank7	24	39	73	38	34	54	59	51	48	42
J/Cod3	73	73	80	57	88	69	93	73	100	75
J/Cod4	59	60	100	81	100	100	100	100	70	100
K/Ref	100	100	100	100	100	100	100	100	100	100
K/Ank3	10	23	40	5	15	5	15	15	10	10
K/Ank7	27	50	50	15	35	15	40	20	20	34
K/Cod3	70	97	60	90	96	80	90	94	98	96
K/Cod4	65	97	99	98	99	99	96	98	99	98
L/Ref	100	95	90	81	80	90	100	90	100	100
L/Ank3	0	0	50	0	0	0	20	20	0	0
L/Ank7	30	20	95	20	20	40	34	40	20	30
L/Cod3	80	80	100	100	70	61	75	95	80	80
L/Cod4	90	100	93	90	100	100	80	100	90	90
M/Ref	100	100	60	100	100	75	100	100	100	74
M/Ank3	20	22	50	20	17	24	24	25	24	21
M/Ank7	26	31	100	31	26	37	33	36	37	33
M/Cod3	61	57	71	61	66	66	60	72	62	100
M/Cod4	89	68	76	82	86	100	83	88	76	69
N/Ref	56	100	73	88	88	100	73	76	100	100
N/Ank3	32	20	100	11	17	15	31	19	17	14

wird fortgesetzt...

... Fortsetzung

Proband /Codec	SQAM-Signal									
	Signal 17	Signal 21	Signal 23	Signal 27	Signal 32	Signal 35	Signal 44	Signal 47	Signal 53	Signal 54
N/Ank7	38	57	91	30	31	31	35	35	30	33
N/Cod3	89	82	50	100	69	84	85	93	74	93
N/Cod4	100	71	32	84	100	89	100	100	90	85
O/Ref	81	84	74	100	100	100	100	95	100	81
O/Ank3	29	34	82	20	21	22	40	30	20	20
O/Ank7	40	54	100	41	34	37	60	42	32	34
O/Cod3	51	69	60	74	52	80	70	90	87	100
O/Cod4	100	100	54	86	66	71	90	100	78	95
P/Ref	100	100	85	100	100	100	100	84	90	100
P/Ank3	6	11	55	9	14	13	10	9	8	7
P/Ank7	14	30	100	14	31	27	26	18	13	17
P/Cod3	44	49	64	74	40	91	60	83	49	87
P/Cod4	86	77	88	89	88	96	86	100	100	70
Q/Ref	100	100	100	90	100	100	100	100	100	100
Q/Ank3	10	10	50	10	10	10	20	10	10	10
Q/Ank7	20	20	70	20	20	20	30	20	20	20
Q/Cod3	40	40	80	100	30	40	75	60	60	70
Q/Cod4	70	80	90	95	80	80	90	80	90	90
R/Ref	100	100	100	93	100	100	91	100	100	100
R/Ank3	10	22	57	4	16	14	10	11	5	5
R/Ank7	20	32	89	9	42	15	21	27	12	18
R/Cod3	43	26	72	100	31	45	40	39	42	93
R/Cod4	62	96	100	100	90	85	100	90	85	93
S/Ref	100	100	65	100	100	100	100	95	97	100
S/Ank3	20	30	50	10	0	20	30	20	10	10
S/Ank7	35	50	90	20	20	40	40	35	20	20
S/Cod3	80	60	90	95	80	90	95	81	80	90
S/Cod4	80	95	100	90	95	95	95	100	100	95
T/Ref	100	89	100	100	96	100	100	98	100	100
T/Ank3	7	14	51	5	7	33	12	4	3	4
T/Ank7	14	21	14	7	18	34	10	5	5	7
T/Cod3	52	95	70	95	70	80	42	86	90	91
T/Cod4	92	100	99	99	100	96	95	100	93	98

Literaturverzeichnis

- [Ale16] S.M. Alessio. Digital Signal Processing and Spectral Analysis for Scientists. Signals and Communication Technology. Springer International Publishing, 2016.
- [Blu12] Bluetooth Audio Video Working Group. Advanced Audio Distribution Profile Specification, Adopted Version 13, Juli 2012.
- [Bur75] J.P. Burg. Maximum Entropy Spectral Analysis. Diss., Stanford University, 1975.
- [CM95] C.D. Creusere and S.K. Mitra. A Simple Method for Designing High-Quality Prototype Filters for M-band Pseudo QMF Banks. IEEE Transactions on Signal Processing, 43(4):1005–1007, April 1995.
- [CP88] A. Charbonnier and J.P. Petit. Sub-band ADPCM Coding for High Quality Audio Signals. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), April 1988.
- [dBGO95] F. de Bont, M. Groenewegen, and W. Oomen. A High-Quality Audio Coding System at 128 kb/s. 98th AES Convention, Februar 1995.
- [Die83] M. Dietrich. Codierung von Sprachsignalen mit 64 kbit/s für Fernsprechen mit erhöhter Tonqualität. Diss., Universität Hannover, 1983.
- [Dur60] J. Durbin. The Fitting of Time Series Models. Review of the International Statistical Institute, 28:233–243, 1960.
- [EBU88] EBU Tech. 3253-E. Sound Quality Assessment Material. Recordings for Subjective Tests, April 1988.
- [EBU08] EBU SQAM CD - Sound Quality Assessment Material Recordings for Subjective Tests. <https://tech.ebu.ch/publications/sqamcd>, 2008. Online - letzter Zugriff 14.03.2016.
- [ES00] B. Edler and G. Schuller. Audio Coding Using a Psychoacoustic Pre- and Post-Filter. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Juni 2000.
- [FHM⁺15] G. Fuchs, C.R. Helmrich, G. Markovic, M. Neusinger, E. Ravelli, and T. Moriya. Low Delay LPC and MDCT-based Audio Coding in the EVS Codec. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), April 2015.
- [Fin98] T. Fingscheidt. Softbit-Sprachdecodierung in digitalen Mobilfunksystemen. Diss., RWTH Aachen, Juli 1998.

- [FLA15] Xiph.Org Foundation. FLAC - Free Lossless Audio Codec. <https://xiph.org/flac/index.html>, 2015. Online - letzter Zugriff 26.11.2015.
- [GCM⁺13] I. Geisel, K. Cordes, J. Mahnke, S. Jöllenbeck, J. Ostermann, J. Arlt, W. Ertmer, and C. Klempt. Evolutionary optimization of an experimental apparatus. *Applied Physics Letters (APL)*, 102(21):214105 – 214105–4, May 2013.
- [GG10] I. Glover and P.M. Grant. *Digital Communications*. Prentice Hall, 2010.
- [Gol89] D.E. Goldberg. *Genetic Algorithms in Search, Optimization & Machine Learning*. Addison-Wesley, 1989.
- [Got03] G. Gottardi. Perceptual Evaluation of Audio Quality beta. <https://sourceforge.net/projects/peaqb/>, 2003. Online - letzter Zugriff 15.03.2016.
- [HD07] J. Herre and S. Disch. New Concepts in Parametric Coding of Spatial Audio: From SAC to SAOC. *IEEE International Conference on Multimedia and Expo*, Juli 2007.
- [HH09] C. Hoene and M. Hyder. Considering Bluetooth's Subband Codec (SBC) for Wideband Speech and Audio on the Internet. *Technischer Bericht 2009-3*, Wilhelm-Schickard-Institut, Universität Tübingen, Oktober 2009.
- [HJ61] R. Hooke and T.A. Jeeves. Direct Search - Solution of Numerical and Statistical Problems. *Journal of the Association for Computing Machinery (ACM)*, 8(2):212–229, 1961.
- [Hol09] M. Holters. Verzögerungsfreie Audiocodierung. Diss., Helmut-Schmidt-Universität - Universität der Bundeswehr Hamburg, 2009.
- [HZ09] M. Holters and U. Zölzer. Automatic Parameter Optimization for a Perceptual Audio Codec. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2009.
- [HZ15] M. Holters and U. Zölzer. GstPEAQ - An Open Source Implementation of the PEAQ Algorithm. *18th International Conference on Digital Audio Effects (DAFx-15)*, Dezember 2015.
- [IET12] IETF RFC 6716. Definition of the Opus Audio Codec, September 2012.
- [ISO93] ISO/IEC 11172-3. Information technology - Coding of Moving Pictures and associated Audio at up to about 1,5Mbit/s - Part 3: Audio, 1993.
- [ISO97] ISO/IEC 13818-7. Information Technology - Generic Coding of Moving Pictures and associated Audio Information - Part 7: Advanced Audio Coding (AAC), 1997.
- [ITU92] ITU-T G.728. Coding of Speech at 16 kbit/s using Low-delay Code Excited Linear Prediction, September 1992.
- [ITU93] ITU-T G.711. Pulse Code Modulation (PCM) of Voice Frequencies, 1993.

- [ITU01] ITU-R BS.1387-1. Method for Objective Measurements of Perceived Audio Quality, November 2001.
- [ITU03] ITU-R BS.1534-1. Method for the Subjective Assessment of Intermediate Quality Level of Coding Systems, Januar 2003.
- [ITU09] ITU. Low Delay Audio Coding for Broadcasting Applications. Technischer Bericht ITU-R BS.2161, November 2009.
- [Kab03] P. Kabal. An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality. Department of Electrical & Computer Engineering, McCill University, Montreal, Dezember 2003.
- [Kei06] F. Keiler. Beiträge zur Audiocodierung mit kurzer Latenzzeit. Diss., Helmut-Schmidt-Universität - Universität der Bundeswehr Hamburg, 2006.
- [KGJV83] S. Kirkpatrick, C.D. Gelatt Jr, and M.P. Vecchi. Optimization by Simulated Annealing. *Science*, 220(4598):671–680, 1983.
- [Kra12] T. Krause. Untersuchung verschiedener Verfahren für den Entwurf von Filterbankstrukturen mit geringer Latenz. Studienarbeit, Leibniz Universität Hannover, 2012.
- [Krü10] H.U. Krüger. Low Delay Audio Coding Based on Logarithmic Spherical Vector Quantization. Diss., Rheinisch-Westfälische Technische Hochschule Aachen, 2010.
- [KSWH04] U. Krämer, G. Schuller, S. Klier Wabnik, and J. Hirschfeld. Ultra Low Delay Audio Coding with Constant Bit Rate. 117th AES Convention, San Francisco, Oktober 2004.
- [KW08] Z. Kurtisi and L. Wolf. Using WAVPACK for Real-Time Audio Coding in Interactive Applications. IEEE International Conference on Multimedia and Expo (ICME), April 2008.
- [Lam11] The LAME Project. <http://lame.sourceforge.net/>, 2011. Online - letzter Zugriff 02.02.2016.
- [Lam15] T. Lamprecht. Vektorquantisierung für die Audiocodierung mit geringer Latenz. Bachelorarbeit, Leibniz Universität Hannover, 2015.
- [LB07] M. Lester and J. Boley. The Effects of Latency on Live Sound Monitoring. 123rd AES Convention, Oktober 2007.
- [Lev47] N. Levinson. The Wiener RMS Error Criterion in Filter Design and Prediction. *Journal of Mathematical Physics*, 25:261–278, 1947.
- [Llo82] S.P. Lloyd. Least Squares Quantization in PCM. *IEEE Transactions on Information Theory*, 28:129–137, März 1982.
- [Luo05] G.Y. Luo. Real-time Ultra-low Delay Audio Coding for Wireless Transmission. The 2nd IEE/EURASIP Conference on DSP enabled Radio, September 2005.
- [Mak77] J. Makhoul. Stable and Efficient Lattice Methods for Linear Predic-

- tion. Acoustics, Speech and Signal Processing, IEEE Transactions on, 25(5):423–428, 1977.
- [Mat08] B. Matschkal. Spherical Logarithmic Quantization. Diss., Friedrich-Alexander Universität Erlangen-Nürnberg, 2008.
- [MH10] B. Matschkal and J.B. Huber. Spherical Logarithmic Quantization. IEEE Transactions on Audio, Speech and Language Processing, 18(1):126–140, Januar 2010.
- [MLRR99] D. Martinez, F. Lopez, M. Rosa, and N. Ruiz. A Nearly Transparent Low Delay Audio Coder. Recent Advances in Signal Processing and Communications (WSES Press), 1999.
- [MRLR00] D. Martinez, M. Rosa, F. Lopez, and N. Ruiz. Low Delay Audio Coder using Adaptive Vector Quantization. 10th European Signal Processing Conference, September 2000.
- [MW16] MathWorks - Genetic Algorithm: Find Global Minima for Highly Non-linear Problems. <http://de.mathworks.com/discovery/genetic-algorithm.html>, 2016. Online - letzter Zugriff 08.04.2016.
- [NBS92] K. Nayebi, T.B. Barnwell, and M.J.T. Smith. Time-Domain Filter Bank Analysis: A New Design Theory. IEEE Transactions on Signal Processing, 40:1412–1429, Juni 1992.
- [NBS94] K. Nayebi, T.B. Barnwell, and M.J.T. Smith. Low Delay FIR Filter Banks: Design and Evaluation. IEEE Transactions on Signal Processing, 42:24–34, Januar 1994.
- [Ngu95] T.Q. Nguyen. Digital Filter Bank Design Quadratic-constrained Formulation. IEEE Transactions on Signal Processing, 43(9):2103–2108, September 1995.
- [NH96] T.Q. Nguyen and P. Heller. Biorthogonal Cosine-Modulated Filter Bank. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Mai 1996.
- [Pau15] M. Pauli. HiRes-Audio um Längen besser als mp3 - Die Rückkehr des Hi-Fi im Digitalen. http://www.deutschlandradiokultur.de/hires-audio-um-laengen-besser-als-mp3.1067.de.html?dram:article_id=282630, 2015. Online - letzter Zugriff 27.11.2015.
- [Pho12] Phonac AG. Binaural VoiceStream Technology - Intelligente binaurale Algorithmen zur Verbesserung des Sprachverständens, 2012.
- [PKO12] S. Preihs, T. Krause, and J. Ostermann. Comparison of Filter Bank Design Algorithms for use in Low Delay Audio Coding. 133rd AES Convention, Oktober 2012.
- [PKVA95] S.M. Phoong, C.W. Kim, P.P. Vaidyanathan, and R. Ansari. A New Class of Two-channel Biorthogonal Filter Banks and Wavelet Bases. IEEE Transactions on Signal Processing, 43(3):649–665, März 1995.

- [PLO16] S. Preihs, T. Lamprecht, and J. Ostermann. Error Robust Low Delay Audio Coding using Spherical Logarithmic Quantization. European Signal Processing Conference (EUSIPCO), August 2016.
- [PO11] S. Preihs and J. Ostermann. Error Robust Low Delay Audio Coding based on Subband ADPCM. 131st AES Convention, Oktober 2011.
- [PO15a] S. Preihs and J. Ostermann. Globally Optimized Dynamic Bit-Allocation Strategy for Subband ADPCM-Based Low Delay Audio Coding. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), April 2015.
- [PO15b] S. Preihs and J. Ostermann. Using Cascaded Global Optimization for Filter Bank Design of a Low Delay Audio Codec. 139th AES Convention, Oktober 2015.
- [PSO12] S. Preihs, F.-R. Stöter, and J. Ostermann. Low Delay Error Concealment for Audio Signals. 46th AES Conference on Audio Forensics, Juni 2012.
- [PWO15] S. Preihs, C. Wacker, and J. Ostermann. Adaptive Pre- and Post-Filtering for a Subband ADPCM-based Low Delay Audio Codec. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), Oktober 2015.
- [RCR07] A.B. Renaud, A. Carôt, and P. Rebelo. Networked Music Performance: State of the Art. 30th AES International Conference, März 2007.
- [Ros60] H.H. Rosenbrock. An Automatic Method for Finding the Greatest or Least Value of a Function. *The Computer Journal*, 3:175–184, 1960.
- [Sch97] G. Schuller. Zeitvariante Filterbänke mit niedriger Systemverzögerung und perfekter Rekonstruktion. Diss., Universität Hannover, Juli 1997.
- [SHZ13a] G. Simkus, M. Holters, and U. Zölzer. Error Robust Delay-Free Lossy Audio Coding Based on ADPCM. 16th Int. Conference on Digital Audio Effects (DAFx), September 2013.
- [SHZ13b] G. Simkus, M. Holters, and U. Zölzer. Ultra-low Latency Audio Coding based on DPCM and Block Companding. 14th International Workshop on Image and Audio Analysis for Multimedia Interactive Services (WIAMIS), Juli 2013.
- [SHZ14] G. Simkus, M. Holters, and U. Zölzer. Error Resilience Enhancement for A Robust ADPCM Audio Coding Scheme. IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP), Mai 2014.
- [SS91] M. Smyth and S. Smyth. APT-X100: A Low-Delay, Low Bit-Rate, Sub-Band ADPCM Audio Coder for Broadcasting. 10th AES International Conference: Images of Audio, September 1991.
- [SS95] G. Schuller and M.J.T. Smith. A New Algorithm for Efficient Low Delay Filter Bank Design. IEEE International Conference on Acoustics,

- Speech and Signal Processing (ICASSP), Mai 1995.
- [SS96] G. Schuller and M.J.T. Smith. New Framework for Modulated Perfect Reconstruction Filter Banks. *IEEE Transactions on Signal Processing*, 44:1941–1954, August 1996.
- [SSJ⁺08] M. Schnell, M. Schmidt, M. Jander, T. Albert, R. Geiger, V. Ruoppila, P. Ekstrand, M. Lutzky, and B. Grill. MPEG-4 Enhanced Low Delay AAC - A New Standard for High Quality Communication. 125th AES Convention, Oktober 2008.
- [Stö12] F.-R. Stöter. Verzögerungssarme Fehlerverschleierung für Audiosignale. Diplomarbeit, Leibniz Universität Hannover, 2012.
- [SYHE02] G. Schuller, Bin Yu, Dawei Huang, and B. Edler. Perceptual Audio Coding using Adaptive Pre- and Post-filters and Lossless Compression. *IEEE Transactions on Speech and Audio Processing*, 10:379–390, September 2002.
- [THS13] B.S. Tiwari, M. Harish, and D. Sinha. A New Ultra Low Delay Audio Communication Coder. 135th AES Convention, Oktober 2013.
- [TTB⁺98] T. Thiede, W.C. Treurniet, R. Bitto, T. Sporer, K. Brandenburg, C. Schmidmer, M. Keyhl, J.G. Beerends, C. Colomes, G. Stoll, and B. Feiten. PEAQ - der künftige ITU-Standard zur objektiven Messung der wahrgenommenen Audioqualität. 20. Tonmeistertagung, 1998.
- [Vai93] P.P Vaidyanathan. Multirate Systems and Filter Banks. Prentice Hall, Englewood Cliffs, 1993.
- [VMTV13] J.-M. Valin, G. Maxwell, T.B. Terriberry, and K. Vos. High-Quality, Low-Delay Music Coding in the Opus Codec. 135th AES Convention, Oktober 2013.
- [VTMM10] J.-M. Valin, T.B. Terriberry, C. Montgomery, and G. Maxwell. A High-Quality Speech and Audio Codec With Less Than 10-ms Delay. *IEE Transactions on Audio, Speech, and Language Processing*, 18:58–67, 2010.
- [Wav98] WavPack - Hybrid Lossless Audio Compression. <http://www.wavpack.com/index.html>, 1998. Online - letzter Zugriff 27.11.2015.
- [Wer07] M. Werner. Digitale Signalverarbeitung mit MATLAB-Praktikum: Zustandsraumdarstellung, Lattice-Strukturen, Prädiktion und adaptive Filter. Studium Technik. Vieweg+Teubner Verlag, 2007.
- [Zwi61] E. Zwicker. Subdivision of the Audible Frequency Range into Critical Bands. *The Journal of the Acoustical Society of America*, 33(2), Februar 1961.

Lebenslauf

Stephan Preihs

geboren am 02.09.1983 in Hannover

Beruf

seit 01/2017

Postdoktorand am *Institut für Kommunikationstechnik (IKT) der Gottfried Wilhelm Leibniz Universität Hannover*

08/2010 - 12/2016

Wissenschaftlicher Mitarbeiter am *Institut für Informationsverarbeitung (TNT) der Gottfried Wilhelm Leibniz Universität Hannover*

Praktikum

10/2008 - 03/2009

Sennheiser Research Laboratory, Palo Alto, USA

Studium

10/2004 - 03/2010

Studium der Elektrotechnik mit Studienrichtung Nachrichtentechnik an der *Gottfried Wilhelm Leibniz Universität Hannover*
Abschluss mit Diplom (Dipl.-Ing.)

Zivildienst

07/2003 - 04/2004

Beim Paritätischen Wohlfahrtsverband Hannover

Schulbildung

08/1996 - 06/2003

Humboldtschule Hannover (Abschluss Abitur)

08/1994 - 06/1996

Orientierungsstufe Martensplatz, Hannover

08/1990 - 07/1994

Grundschule Wettbergen, Hannover

Online-Buchshop für Ingenieure

■■■ VDI nachrichten

Online-Shops



Fachliteratur und mehr -
jetzt bequem online recher-
chieren & bestellen unter:
www.vdi-nachrichten.com/
Der-Shop-im-Ueberblick



Täglich aktualisiert:
Neuerscheinungen
VDI-Schriftenreihen



BUCHSHOP

Im Buchshop von vdi-nachrichten.com finden Ingenieure und Techniker ein speziell auf sie zugeschnittenes, umfassendes Literaturangebot.

Mit der komfortablen Schnellsuche werden Sie in den VDI-Schriftenreihen und im Verzeichnis lieferbarer Bücher unter 1.000.000 Titeln garantiert fündig.

Im Buchshop stehen für Sie bereit:

VDI-Berichte und die Reihe **Kunststofftechnik**:

Berichte nationaler und internationaler technischer Fachtagungen der VDI-Fachgliederungen

Fortschritt-Berichte VDI:

Dissertationen, Habilitationen und Forschungsberichte aus sämtlichen ingenieurwissenschaftlichen Fachrichtungen

Newsletter „Neuerscheinungen“:

Kostenfreie Infos zu aktuellen Titeln der VDI-Schriftenreihen bequem per E-Mail

Autoren-Service:

Umfassende Betreuung bei der Veröffentlichung Ihrer Arbeit in der Reihe Fortschritt-Berichte VDI

Buch- und Medien-Service:

Beschaffung aller am Markt verfügbaren Zeitschriften, Zeitungen, Fortsetzungsreihen, Handbücher, Technische Regelwerke, elektronische Medien und vieles mehr – einzeln oder im Abo und mit weltweitem Lieferservice

VDI nachrichten

BUCHSHOP www.vdi-nachrichten.com/Der-Shop-im-Ueberblick

Die Reihen der Fortschritt-Berichte VDI:

- 1 Konstruktionstechnik/Maschinenelemente
- 2 Fertigungstechnik
- 3 Verfahrenstechnik
- 4 Bauingenieurwesen
- 5 Grund- und Werkstoffe/Kunststoffe
- 6 Energietechnik
- 7 Strömungstechnik
- 8 Mess-, Steuerungs- und Regelungstechnik
- 9 Elektronik/Mikro- und Nanotechnik
- 10 Informatik/Kommunikation
- 11 Schwingungstechnik
- 12 Verkehrstechnik/Fahrzeugtechnik
- 13 Fördertechnik/Logistik
- 14 Landtechnik/Lebensmitteltechnik
 - 15 Umwelttechnik
 - 16 Technik und Wirtschaft
 - 17 Biotechnik/Medizintechnik
 - 18 Mechanik/Bruchmechanik
 - 19 Wärmetechnik/Kältetechnik
- 20 Rechnerunterstützte Verfahren (CAD, CAM, CAE CAQ, CIM ...)
- 21 Elektrotechnik
- 22 Mensch-Maschine-Systeme
- 23 Technische Gebäudeausrüstung

ISBN 978-3-18-385410-3